Integrated Experimental and Software Methods for Non-Targeted Analyses Investigation of Vehicle-Derived Chemicals and Their Transformation Products

Ximin Hu

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2023

Reading Committee:

Edward P. Kolodziej, Chair

David. A. C. Beck

Michael C. Dodd

Katherine T. Peter

Program Authorized to Offer Degree:

Civil and Environmental Engineering

University of Washington

Abstract

Integrated Experimental and Software Methods for Non-Targeted Analyses Investigation of Vehicle-Derived Chemicals and Their Transformation Products

Ximin Hu

Chair of the Supervisory Committee:

Edward P. Kolodziej

Department of Civil and Environmental Engineering

Water pollution is a significant environmental issue that can yield detrimental impacts on human health and ecosystem. Compounding the basic environmental pollution problem is the fact that many chemicals can undergo various transformation reactions under environmental conditions to form structurally similar transformation products (TPs). Notably, some TPs may contribute significantly to the potential for adverse effects in environmental systems while their characteristics and transformations are not fully understood. For example, among various pollutant classes, vehicle-derived chemicals and their TPs are a growing concern, representing compounds that are concurrently abundant (by mass use), widespread,

and poorly characterized or identified. For instance, our group recently identified 6PPD (N-(1,3-dimethylbutyl)-N'-phenyl-p-phenylenediamine), a ubiquitously used tire rubber antioxidant, would be transformed under environmental conditions into 6PPDQ, a toxicant which is responsible for the coho salmon (*Oncorhynchus kisutch*) acute mortality observed for several decades in the Pacific Northwest. Therefore, there is a pressing need to chemically characterize and to investigate the environmental fate and transport of these classes of vehicle-derived chemicals and their TPs for comprehensive environmental risk assessment, remediation and policy making.

Problematically, identifying and tracking specific groups of pollutants or contaminant sources is often challenging due to the complex chemical matrices present in surface waters and the relatively narrow detection capabilities of traditional analytical methods (e.g., targeted analytical methods via low-resolution mass spectrometry) that focus on limited numbers of pre-defined analytes. Non-targeted analysis (NTA) can potentially address this challenge by coupling high-resolution mass spectrometry (HRMS) instrumentation with data science approaches for data analysis to better identify or quantify unknown pollutants or contamination sources in complex mixtures. However, the development of HRMS data processing workflows is still in a relatively early stage, is often labor intensive, and can lack well-established protocols and integrated data processing capabilities, especially for open-source software, for environmental data and systems.

To address the challenges above, this thesis communicates the development of improved integrated experimental and open-source software methods for NTA using HRMS

instrumentation and capabilities. Subsequently, these methodologies were deployed for the chemical characterization of vehicle-derived chemicals, focusing on the industrial antiozonant 6PPD widely used in tire rubbers and related transformation products. Chapter 1 presents an introduction to these systems. Chapters 2-4 communicate the analysis and characterization of 6PPD and related TPs. 6PPD TPs include 6PPD-quinone (6PPDQ), which is an emerging contaminant that was previously identified as the "primary causal toxicant" for acute mortality events in stormwater-exposed coho salmon. Given its extraordinarily high toxicity, interest in 6PPDQ properties and fate extends from local to global scales. Experiments were conducted to (a) investigate the physiochemical properties of 6PPDQ; (b) measure 6PPD ozonation dynamics and quantify known TPs (i.e., 6PPDQ); (c) prioritize other potential 6PPD TPs with HRMS-based NTA; and (d) evaluate the environmental fate of 6PPD and formation of TPs under varied environmental conditions (ozone exposure, aerobic and anaerobic conditions) while quantifying 6PPD and TPs with liquid chromatography-tandem mass spectrometry (LC-MS/MS). Chapters 5-6 focus on method development for HRMS NTA, including workflow development for NTA data processing and development and optimization of source identification and apportionment methodologies.

Chapter 1 provides an introduction and explains the research goals of the thesis. Chapter 2 discusses the investigation of the physiochemical properties of 6PPDQ, including logK_{ow}, solubility, leaching potentials, sorption potentials and aqueous stability. We focused on reporting chemical characteristics relevant to the fate and transport of the recently discovered environmental toxicant 6PPDQ. The aqueous solubility and octanol–water partitioning

coefficient ($\log K_{ow}$) for 6PPDQ were measured as 38 ± 10 μg L^{-1} and 4.30 ± 0.02, respectively. Within the context of analytical measurement and laboratory processing, sorption to various laboratory materials was evaluated, indicating that glass was largely inert but loss of 6PPDQ (including non-recoverable mass) to other material types was common, including materials commonly found in the laboratory. Aqueous leaching simulations from tire tread wear particles (TWPs) indicated short-term release of ~5.2 μg 6PPDQ per gram TWP over 6 h under flow-through conditions. Aqueous stability tests observed a slight-to-moderate loss of 6PPDQ over 47 days (26 ± 3% loss) for pH 5, 7, and 9. These measured physicochemical properties suggest that 6PPDQ is generally poorly soluble (almost surprisingly so) but fairly stable over short time periods (~5% loss over 3d) in simple aqueous systems. 6PPDQ can also leach readily from TWPs for subsequent environmental transport, posing high potential for adverse effects in local aquatic environments.

Chapter 3 evaluated the transformation kinetics of 6PPD degradation and 6PPDQ formation using heterogeneous 6PPD ozonation in the gas phase. We investigated TP formation occurring during heterogeneous reaction of gas-phase ozone with 6PPD; exposures included both pure 6PPD solids and TWP rubber particles that contained 6PPD. Oxidative transformation occurred during these ozonation conditions (inlet O_3 concentration ~360 ppbv), with up to 81% of 6PPD mass reacting over 6 h. Conversion of 6PPD to 6PPDQ was confirmed at a 9.7% molar yield for pure 6PPD solid and 0.95% molar yield for 6PPD present within TWPs, representing likely minima over these time scales and conditions. Leveraging HRMS suspect screening approaches, we identified 19 probable 6PPD-derived TPs in both ozonated

6PPD and TWP samples, underscoring formation of diverse TPs from this antioxidant compound. By screening environmental samples, nine 6PPD-derived TPs were subsequently detected within roadway runoff to communicate environmental relevance. The data confirms that when tire rubber antioxidants react with ozone, as intended, they form and release various TPs to surrounding environments with high potential to impair water quality.

Extending from Chapter 3, Chapter 4 evaluates 6PPD fate and transformation under additional model environmental conditions. Different reaction conditions were tested to comparatively investigate 6PPD transformation processes, including gas phase ozonation, gas phase and aqueous phase aerobic exposures. The gas phase ozonation experiments included conditions that were not evaluated in Chapter 3. During these experiments, 6PPD mass reacted progressively during ozone exposure under tested conditions (93±2% 6PPD degraded during 2.5 h; inlet [O_3] = 360 ppbv). The aqueous phase ozonation experiment determined the apparent reaction rate constant of 6PPD with molecular O_3 as $(1.18\pm0.16)\times10^6$. Additionally, 6PPD environmental transformation under aerobic conditions in gaseous and aqueous phases suggested 6PPD is prone to transformation in water (half-life: 10.0 h) and fairly stable under gas phase aerobic exposures if they lack or only contain trace molecular ozone (half-life: 69.3 days). The formation of 6PPD TPs was evaluated during all these exposure conditions. Among all the tested conditions, 6PPDQ formation was only observed from 6PPD in gaseous systems where ozone was clearly present. The results suggested that 6PPD-gaseous O_3 interaction is the primary formation pathway of 6PPDQ under ambient environmental conditions, while the

environmental fate, transport, and potential effects of other, still poorly understood 6PPD-TPs merit further investigation.

Chapter 5 describes the development of an open-source Python package *Mass-suite (MSS)* as a full, expandable open-source HRMS data analysis toolbox with multiple NTA data processing capabilities. *MSS* provides flexible, user-defined workflows for HRMS data processing and analysis, including both basic functions (e.g., feature extraction, data reduction, feature annotation, data visualization, and statistical analyses) and advanced exploratory data mining and predictive modeling capabilities that are not provided by currently available open-source software (e.g., unsupervised clustering analyses, a machine learning-based source tracking and apportionment tool). As a key advance, most core *MSS* functions are supported by machine learning algorithms (e.g., clustering algorithms and predictive modeling algorithms) to facilitate function accuracy and/or efficiency. *MSS* reliability was validated with mixed chemical standards of known composition, with 99.5% feature extraction accuracy and ~52% overlap of extracted features relative to other open-source software tools. Example user cases of laboratory data evaluation are provided to illustrate *MSS* functionalities and demonstrate platform reliability. *MSS* expands available HRMS data analysis workflows for water quality evaluation and environmental forensics and is readily integrated with existing database and screening platforms. Though fully described in Chapter 5, applications of some *MSS* functions are also described in Chapters 3 and 4 to facilitate data analysis of complex chemical mixtures.

Chapter 6 focused on the development and optimization of HRMS data processing workflows with machine learning methodologies. Different algorithms and methodologies

were trained with ~550 data points and optimized for source identification and apportionment predictions. Both qualitative and quantitative models were leveraged to facilitate the data analysis and enable insight into NTA data tools that do not rely on feature identification. Dot product-logistic regression (100% Acc.) and XGBoost classification model (93.3% Acc. on training data) were trained for source identification. For source quantification method development, various strategies were evaluated to mitigate the potential risks of overfitting derived from relatively scarce sample data points (n=~550) versus substantial detection numbers of unique chemical features (n=~3000). The methods tested include feature selection (XGBoost), feature extraction (similarity scores and principal component analysis) and regularization regressions (lasso, ridge and elastic net regressions). The results suggested that regularization models outperformed other models on source apportionment predictions (with the highest R^2 scores of predicted and actual value), while PCA (Principal Component Analysis) dimension reduction coupled with SVM (Support-Vector Machine) regression is a strong candidate data processing strategy that merits further training and optimization efforts for source apportionment applications.

These research findings provide valuable insights for researchers, policymakers, and regulatory agencies in understanding and mitigating the impact of emerging contaminants on human and environmental health. Leveraging the use of targeted analysis approaches, the transformation dynamics and the environmental fate and transport of 6PPD and 6PPD-derived TPs were further characterized by these studies. On the other hand, the data processing workflow described here provides a workflow and developmental template for future non-

targeted analysis investigations of complex chemical mixtures in aquatic environments, especially for source tracking studies. Future investigations of other related emerging contaminants (i.e., especially other PPD family chemicals) and their transformation products are essential for improving the overall risk assessment of these types of pollutants in aquatic environments. Further development and optimization of these analytical workflows, along with more sample collection and analysis efforts, is needed to fully optimize potential HRMS-NTA workflows for contamination source identification and quantification for different source and background matrices. In conclusion, this research provides critical information to develop management strategies for emerging contaminants derived from 6PPD used in vehicle tires, as well as develop novel HRMS methodologies for source tracking studies.

Acknowledgments

Grand journeys are always short as time flies. It still feels like I just started my Ph.D. not long ago while I'm wrapping up this thesis. There were so many things happening during my Ph.D., including attempting, learning, experiencing failure, growth, and achieving success. Also, numerous people supported me, worked with me, and guided me throughout my journey which I'd like to thank. All my story started with the (last minute) offer from Prof. Edward Kolodziej, my advisor and mentor. Thank you, Ed, for allowing me to start my research at UW and explore whatever I want to try. I truly enjoyed our talks and discussions, where I learned the art of critical thinking. I've also acquired the motto of 'curiosity motivates science' from you. Every conversation we had, regardless of the topic, brought me delight as I received your insightful questions. We also brainstormed and delved into new territories together from those questions, and I learned tons about what I still didn't know or hadn't considered before. I'll treasure and apply such curiosity and scientific integrity from the Kolodziej lab in my future life.

I'm grateful to my co-advisors (although not on the official documents) and committee members, Prof. Michael Dodd, Prof. David Beck, Dr. Katherine Peter, and Prof. František Tureček for their guidance throughout my graduate program. I'm so fortunate to have such much guidance and support from researchers with different expertise and deploy interdisciplinary knowledge to my research works. Thank you, Mike, as I started as a scientist knowing how to prepare, conduct and report a study 'in general', you showed me excellent examples of being a scientist conveying detailed and accurate results. I'm always amazed by

your knowledge and instinct (I know most of which also derived from knowledge) with organic chemistry problems that always navigate a clear path out for my research. I am also grateful for your understanding and support during my Ph.D. program, you are always helpful when I encounter any issues with administrative paperwork or manuscript writing. Thank you, Dave, for leading me into the computational science field. The lectures you gave and the coding sessions we had together prepared me as a (although still not 100% expert) programmer and gave me fascinating insights into machine learning and data science. The conversations we had always sparked my ideas about interdisciplinary applications of data science and mass spectrometry data; those inspirations were the main drivers of my last two chapters. Thank you, Kathy, for all your dedication and thoughtful ideas. You are a great colleague and patient mentor, and it feels so reliable and delightful to work and chat with you, as you are always helpful and provide me the detailed guidance. You are also my 'go-to' example of being a 'professional scientist', I will always think of you as setting a standard when approaching and solving a problem. Prof. Tureček, thank you for the Analytical Mass Spectrometry lectures in my first year and for being the graduate school representative in my defense; it is my honor to have you as my committee member.

I am also grateful to the brilliant researchers from Kolodziej team at Center for Urban Waters: Dr. Zhenyu Tian and Dr. Haoqi (Nina) Zhao. Zhenyu, thank you for always providing me with informative, helpful, or creative advice (either for experiments, writing, or personal life), for teaching me professional skills and for all the fun talks in the lab. Nina, thank you for discussing the research ideas, setting up and conducting experiments with me (I

learned to be cautious about ozonation experiments!) and for your immense support to me for manuscript writing, presentation practice and so much more things professionally or personally. Most of the chapters in the thesis would not have come to fruition without your unwavering support. I'm also fortunate to have tons of assistance at CUW, including Melissa Gonzalez (thanks for all the work you've done), Craig Rideout (for your heavy lifting), David Wark (for taking care of QTOF), Marlyn Del Cid (for your foods and helps with paper works), Prof. Joel Baker (for providing us such good research lab) and Prof. Andy James (for all the discussions we had about HRMS data processing). Additionally, UW CEE professors and researchers also offered me innumerous support, especially Prof. Gregory Korshin, Prof. Jess Ray, Prof. Mike Brett and Songlin Wang. I would also like to thank my peers for their professional and personal support, including Dr. Huan He, Dr. Sin-Yi Liou, Alanna Hildebrandt and Jennifer Hooper.

Lastly, I would like to thank my family and friends, especially my beloved wife, I will never have enough thanks for your support, care, understanding and love that brings me here.

Table of Contents

Chapter 1: Introduction

1.1 Emerging Contaminants and Transformation Products (TPs)

Water pollution is one of the most problematic environmental issues of the 21st century because developing and intensifying human societies are producing and dispersing large amounts of diverse organic contaminants into the environment, especially in aquatic systems (Hollender et al., 2017; René P. Schwarzenbach, 2010). The U.S. Environmental Protection Agency (EPA) reported that ~11,000 industrial or municipal wastewater facilities had illegally discharged pollution during FY 2018 (U. S. Government Accountability, 2022). Over 906,000 chemicals are listed on the CompTox Chemicals Dashboard (Williams et al., 2017) and 350,000 chemicals have been registered for production and use (Wang et al., 2020). These inventory estimates suggest a wide range of potential chemical contaminants subject to release and detection in surface waters, with substantial numbers of additional unknown contaminants (e.g. transformation products, technical mixtures, chemicals not registered in CAS) remaining wholly uncharacterized. Increasing human population and activities keep increasing the numbers and masses of discharged contaminants, underscoring urgent need for monitoring and management of anthropogenic pollution. (Lin et al., 2022; René P. Schwarzenbach, 2010). Environmental scientists and engineers have worked to address these problems for decades with extensive research to detect contaminants and manage pollution sources. Such efforts range from traditional nutrient pollutants such as nitrogen and phosphorous (Kroiss et al., 2011; Puckett, 1995) to the “emerging organic contaminant” pollution of pharmaceuticals, personal

care products (PPCPs), pesticides, industrial and commercial chemicals present within consumer products or used in their manufacture (**Table 1.1**(Freyria et al., 2018)) (Ebele et al., 2017; Yang et al., 2017). Compared to the contaminants mentioned above, only scarce information exists for many emerging contaminants. Therefore, there is an especially pressing research need to characterize these groups of chemicals with respect to their environmental occurrence, fate, transport, and effects.

Table 1.1. Example PPCPs and emerging contaminants (ECs) commonly detected in water bodies. Published by Freyria et al., (2018).

Main Categories	Families	Examples
Drugs and Pharmaceuticals	antibiotics/antibacterials	amoxicillin, metronidazole, ofloxacin
	steroids	estrone, 17β-estradiol, testosterone
	β-blocker	propranolol, salbutamol, atenolol
	nonsteroidal anti-inflammatory drugs (NSAID)	aspirin, ibuprofen, naproxen, ketoprofen
	antiepileptic/anticonvulsants	gabapentin, carbamazepine
	antidepressant/hypnotic	diazepam, venlafaxine, amitriptyline, dosulepin, meprobanate
	analgesic	morphine, propoxyphene, paracetamol
	hypertension	valsartan
	lipid regulation	bezafibrate, simvastatin, clofibric acid
	erectile dysfunction	sildenafil
Stimulant and generally illegal drugs	hallucinogen	3,4-Methylenedioxymethamphetamine (MDMA), 3,4-Methylenedioxy-*N*-ethylamphetamine (MDEA), 3,4-Methylenedioxyamphetamine (MDA)
	stimulant	amphetamine, cocaine, benzylpiperazine
	human indicator	caffeine, nicotine
Personal care products	preservative	methylparaben, propylparaben
	sunscreen agent	1-benzophenone, homosalate
	disinfectants/antiseptic	chloramines, chlorine, chlorine dioxide, chlorhexidine digluconate
	fragrances	musk xylol, tonalide
Pesticides/Herbicides	organohalogenated compounds	Dichlorodiphenyltrichloroethane (DDT), lindane, vinclozolin, clopyralid
	nitrogen containing	simazine, phenylurea

Over the past decade, microplastic pollution is one of the most popular and public examples of topical problems associated with emerging contaminants discharge as increasing evidence

documents the wide distribution and potential of microplastics to induce harmful health effects on humans and other organisms (**Figure 1.1**) (D'Avignon et al., 2022).

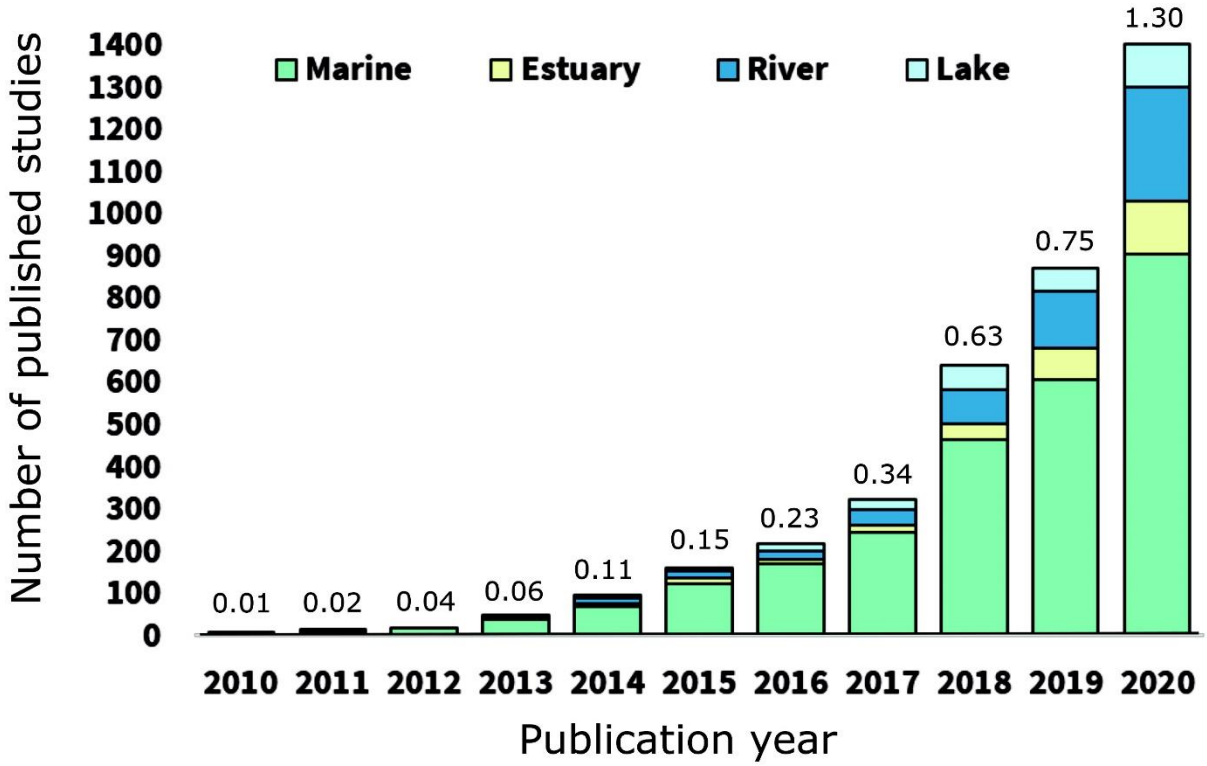


Figure 1.1. Number of peer-reviewed studies on microplastic pollution (particles ⩽ 5 mm) for various waterbody types, published between 2010 and 2020. Values above the bars are the percentages (%) of studies on microplastic pollution based on research efforts in the 25 reviewed journals. Published by D'Avignon et al (2022).

A wide range of consumer products contributes to concerns over microplastic release, from food packaging to disposable equipment, making microplastic pollution ubiquitous in the environment (Campanale et al., 2020; Pironti et al., 2021). In freshwater and marine systems, tire road wear particles (TRWP) that contain tire wear particles (TWP) are considered the largest overall contributor to environmental microplastic discharge. For example, Kole et al (Kole et al., 2017) estimated 1,252,000-1,797,000 metric tons of TRWP were annually released in the US. Because TWP consists of various chemicals and materials, many of which are water

soluble, those components are also leached from TWP when wetted, suggesting that TWP represents a substantial and pervasive source of chemical pollution to aquatic environments. The most well documented materials released by TWP include carbon black, polyaromatic hydrocarbons, zinc, and synthetic or natural rubber phases. In addition, there exist many other tire rubber additive chemicals, including accelerators, antioxidants and vulcanizing agents, that are poorly or not described by tire manufacturers and are environmentally dispersed during tire production, use, and recycling/disposal (Huntink, 2003; Sheridan, 2010). Many of these chemicals reflect proprietary ingredients, technical mixtures, or uncharacterized products whose chemical compositions are not available to the public or are substantially undefined with respect to their use and pollution potential.

All of these complex chemical mixtures released from TWP are themselves potential sources of additional novel or uncharacterized transformation product chemicals formed from industrial and environmental reaction processes, complicating our understanding of chemical composition and identity when environmental samples are analyzed. These complex mixtures are transported into the environment by TWP dispersal, including airborne pathways, and could potentially induce adverse environmental effects or human health effects upon exposure to sensitive populations. Unfortunately, aside from a few chemicals (such as HMMM family chemicals and 1,3-diphenylurea) (Alhelou et al., 2019; Johannessen et al., 2021; Zhang et al., 2023), most of the parent compound-environmental TPs relationships are not yet evaluated and remain largely unknown. For example, the significant conceptual and informational barriers posed by “confidential and proprietary business information” protections of the chemical

ingredients used to make tires severely limit assessment of environmental contamination and associated health risks to humans and other organisms that are exposed to these materials in some fashion. Absent regulatory and policy changes that mandate detailed release of these types of information to researchers and the general public, focused environmental studies employing advanced analytical instrumentation are needed to overcome these current limitations.

Because they have potential bioactive characteristics, analysis and characterization of TPs is a critical need for environmental research in general. TPs are "secondary" organic compounds that are derived from the primary "parent" contaminants by environmental processes such as oxidation, biodegradation, photolysis and hydrolysis (Chen et al., 2012; Jin et al., 2012; Zhang et al., 2016). While many TPs are likely benign and not bioavailable, some TPs could induce problematic adverse environmental effects if they are: a) concentrated in the environment; b) resistant to transformation, mobile with surface water or bioavailable; or c) have bioactive or toxic characteristics, including those distinct from their parent compounds (Escher and Fenner, 2011; Yin et al., 2017; Picó and Barceló, 2015). Although some parent compounds (which are then chemically modified into TPs by environmental and industrial processes) and TPs are well-studied (Hu et al., 2022; Schollée et al., 2018; Weatherly and Gosse, 2017), many to most environmentally relevant emerging contaminants remain unknown and lack detailed investigation. Furthermore, absent a compelling external reason to explore their identities, most TPs have received much less research attention and structural

characterization relative to their parent compounds (Ross et al., 2018; Tian et al., 2021b; Zhao et al., 2021), creating a substantial data gap.

To investigate these groups of TWP-derived emerging contaminants and their TPs, recent research studies have focused on understanding their sources, environmental fate, and transport to evaluate the potential for adverse effects of TWP-derived chemical exposures. Examples of tire-derived chemicals (except metals like zinc) that are found to have adverse environmental and human health effects include N-(1,3-dimethylbutyl)-N'-phenyl-p-phenylenediamine (6PPD), and different polycyclic aromatic hydrocarbons (PAHs) (Chen et al., 2023; Patel et al., 2020; Varshney et al., 2022; Zhang et al., 2023). Specifically, a representative example that was ultimately linked to TWP was the longstanding unexplained acute mortality of coho salmon (*Oncorhynchus kisutch*) in the Pacific Northwest region (Chow et al., 2019; McIntyre et al., 2018; Scholz et al., 2011; Spromberg and Scholz, 2011). Peter et al. utilized an non-targeted analysis (NTA) workflow to prioritize candidate chemical features that were linked to mortality events in coho salmon; these data suggested TWP was the most likely source of detected chemical features also linked to mortality events (Peter et al., 2018). Subsequently, Tian et al. (Tian et al., 2021b) reported the discovery of 6PPDQ (2-((4-Methylpentan-2-yl)amino)-5-(phenylamino)cyclohexa-2,5-diene-1,4-dione), a highly toxic TP (LC_{50}=95 ng/L; (Tian et al., 2021a)) of the common tire antioxidant and anti-ozonant compound 6PPD that was identified as the “primary causal toxicant” for observations of coho mortality. Reported to be among the most toxic chemicals known for aquatic organisms, 6PPDQ was also found to be toxic to other salmonids (Brinkmann et al., 2022). Since initial

publication, 6PPDQ has been reported globally in roadway runoff, surface water, snowmelt, roadside soils, house dust, river sediments, estuary sediments, deep sea sediments, and atmospheric particulate matter (Cao et al., 2022; Challis et al., 2021; Deng et al., 2022; Hiki et al., 2021; Johannessen et al., 2021; Klöckner et al., 2021a; Rauert et al., 2022; Seiwert et al., 2022; Tian et al., 2021a; Zeng et al., 2023). The parent chemical of 6PPDQ, 6PPD is ubiquitously used in tire rubber products (0.4-2% by mass; (Sheridan, 2010); it is then transformed into 6PPDQ via oxidative reactions (see **Figure 1.2**), especially gas phase reactions with ozone, during the tire service life and then deposited into the environment via TWP (Councell et al., 2004; Klöckner et al., 2021b; OSPAR Commission, 2006c; Seiwert et al., 2022; Tian et al., 2021b; Wagner et al., 2018). TWP transports 6PPDQ and many other tire derived contaminants throughout the environment as it is flushed into roadway runoff during storm events or aerosolized by traffic and weather events (Cao et al., 2022; Chen et al., 2023; Rødland et al., 2022; Seiwert et al., 2022; Zhang et al., 2021). These processes subsequently release 6PPDQ throughout the environment where exposure to sensitive species like coho salmon occurs.

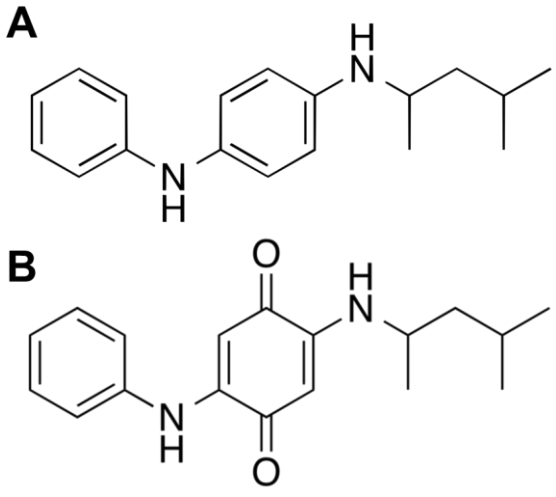


Figure 1.2. Chemical structure of (A) 6PPD and (B) 6PPDQ. 6PPDQ is formed as the rubber antioxidant 6PPD reacts with ground level atmospheric ozone on tire rubber surfaces.

Relative to the environmental prevalence and very high toxicity of 6PPDQ and the drastic adverse environmental effects induced upon exposure, there exists only fragmented information about 6PPD-derived TPs, including 6PPDQ, and other tire rubber derived environmental contaminants. Substantial data gaps exist regarding their chemical properties, formation, environmental fate, transport, and management (Klöckner et al., 2021a; Tian et al., 2021b, 2021a). The transformation of 6PPD under oxidative conditions, believed to be the primary source of 6PPDQ, has received limited attention until recently (Hu et al., 2022; Huntink, 2003; Lattimer et al., 1983; Seiwert et al., 2022). To enable more effective management and control of tire-rubber derived environmental risks arising from chemical discharges, especially for those related to 6PPDQ, it is critical to assess 6PPD transformation while also investigating the fate and transport of 6PPD and related TPs (Drzyzga, 2003; Huntink, 2003; OSPAR Commission, 2006c). In this thesis, we focused our efforts on investigation of the common tire antioxidant 6PPD and key TPs, especially 6PPDQ, to address critical data gaps and improve our understanding of 6PPD fate (**Chapters 2-4**).

1.2 Characterization of Organic Pollutants Using Mass Spectrometry (MS) Analytical Methods

To better manage emerging contaminants, efficient and accurate detection and chemical characterization, which is critical for risk assessment and toxicity evaluation, is necessary.

Traditionally, environmental analytical methods only covered analysis of a few compounds concurrently, which are usually chemically related in some way and often prioritized analytically and for management by evidence of toxicity or exposure risk or production and discharge characteristics. These "targeted" methods were developed by optimizing instrument methods and performance starting from available chemical standards; they provided excellent sensitivity and selectivity for compound detection and quantification. However, the heavy initial reliance on commercial standards consequently limits characterization efforts to a relatively small group of chemicals where commercial standards are available, often failing to capture novel contaminants or TPs where commercial standards are usually unavailable. For example, EPA drinking water regulations only list 53 chemicals for specific management (US EPA, 2015), leaving many thousands, or more, of emerging contaminants unregulated and lacking systemic characterization. The discrepancy between the time-consuming and laborious detection method development versus fast growing and increasing recognition of emerging contaminant pollution generates a substantial capability gap for contaminant monitoring and management efforts. To address this analytical coverage issue, various analytical techniques have been implemented in environmental studies. Novel detection approaches emerged with increased processing and data analysis capabilities to address the rising concern around emerging contaminants. Available methods usually include either spectrophotometry methods (UV and fluorescence) or mass spectrometry (MS) detectors coupled with gas or liquid chromatography (GC and LC) (**Figure 1.3**) (Huang, 2017; Kumar et al., 2022; Noguera-Oviedo and Aga, 2016).

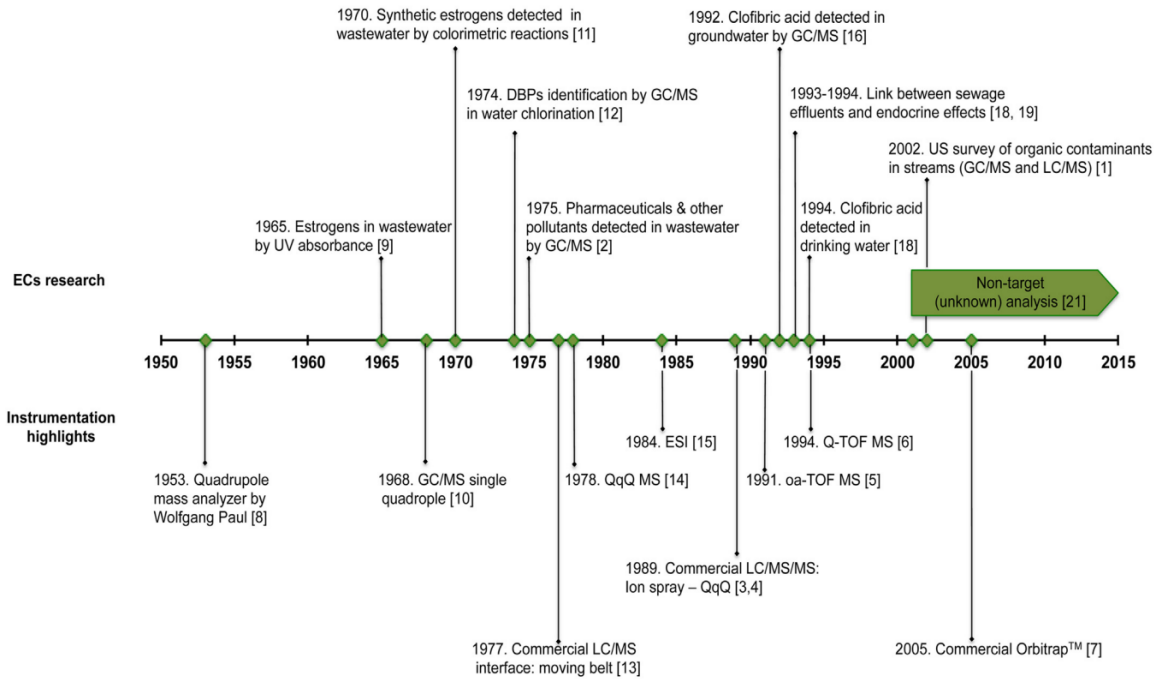


Figure 1.3. Timeline highlighting major events in emerging contaminants research and related instrumentation development, published by Noguera-Oviedo and Aga (2016).

Among available techniques, liquid chromatography – high-resolution mass spectrometry (LC-HRMS) analyses provide especially comprehensive and open-ended screening capabilities to define chemical compositions of complex samples; such methods have opened new frontiers in environmental chemistry, health, and engineering research over the past 25 years (Acena et al., 2015; Aydogan, 2020). Relying on effective LC separation and high signal resolution (compared to other methods) HRMS provides high sensitivity of compound detection and coverage of a broad spectrum of contaminants with reasonably confident identification capabilities. Hundreds to thousands of chemical contaminants (often reported as "features" that reflect an exact mass-retention time pair) are typically concurrently detected in each environmental sample analyzed. HRMS data coupled with well-established database searches based on exact mass, chemical formula, isotopic pattern, or MS/MS fragment data that can link to chemical identity also allow researchers to conduct 'suspect screening' or

'NTA' analyses of the emerging contaminants, in sharp contrast to the targeted analysis of small groups of chemicals that requires standard or *a priori* knowledge of the samples (Hollender et al., 2017; Hollender et al., 2019). The suspect screening method utilizes pre-determined lists, external databases or reaction pathway predictions to annotate chemicals that could potentially be present in the samples. Subsequently, instrument detection capabilities are biased to those contaminants "suspected" or potentially present to establish identification or quantification. However, this approach could still be limited because it requires some experimental information or initial record of the chemicals in databases (it is essentially an educated guess as to identity or presence) and therefore can fail at true chemical discovery tasks.

To overcome the challenge of discovery and compositional characterization absent any initial information, "non-target analysis" (NTA) workflows were developed to exploit the massive data collection capacity of HRMS instrument and enable open ended contaminant identification. In a typical NTA workflow, detected MS features – where a feature is a single presumptive detection of a chemical or its adducts/isotopologues and represented as an exact mass (m/z value) - retention time (RT) pair, usually involving extensive parameter tuning and quality assessment of the extracted peaks – would be extracted, aligned and characterized according to the user needs. Instead of screening against pre-determined lists or external databases, in NTA studies, detected features are filtered and analyzed non-selectively to complement user defined parameters, expert knowledge or external information associated with the sampling sites (e.g., near highways, effluent of treatment plants, presence of

known/related chemicals) (Hollender et al., 2017). Therefore, NTA is amenable to discovering and characterizing unknown chemicals (e.g., not CAS listed, uncharacterized TPs) that are potentially hazardous but remain unrecognized such as those anthropogenic or natural contaminants discharged as primary contaminants. NTA is also capable of screening contaminant TPs or other structurally related compounds or compound classes which could be potentially more prevalent or harmful than the parent compounds and therefore, merit more comprehensive investigations. Often, all three analytical approaches (targeted analysis, suspect screening and NTA) are co-leveraged to assist researchers in better characterizing contaminants and evaluating environmental pollution and health risks. For example, a typical integrated workflow for marine toxin screening is summarized in **Figure 1.4** by Dom et al (Dom et al., 2018). Similarly, these approaches were employed in these studies communicated here for the investigation of roadway and tire rubber-derived emerging contaminants (i.e., 6PPD) and their TPs (**Chapters 2-4**). Concurrently, a software package is developed and optimized to assist the NTA data processing, especially source tracking studies (**Chapters 5 and 6**).

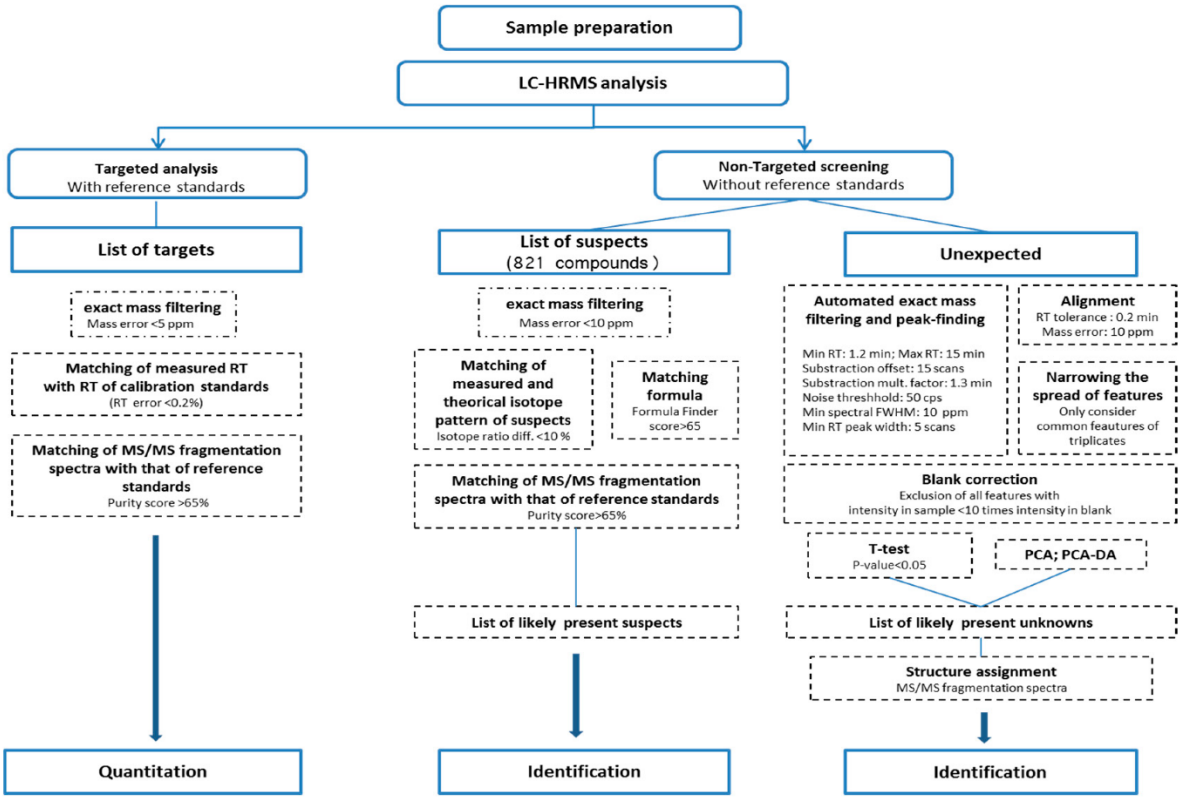


Figure 1.4. Processing workflow for contaminant identification and quantification of marine toxins, published by Dom et al (2018).

1.2.1 Targeted Analysis

As described above, targeted analysis enables high sensitivity and accuracy regarding the quantification of concentrations to method selectivity. Therefore, this approach is usually very reliable and accurate, suitable for monitoring environmental contaminants or known TPs at trace level concentrations. This analytical approach is commonly conducted on low-resolution tandem mass spectrometry (e.g., triple quadrupole (QqQ)) instruments that have high sensitivity for targeted transition monitoring and allows accurate qualitative and quantitative analysis of the selected chemicals when chemical standards are available. For example, Xu and Knoerr et al. developed a water analysis method based on LC-MS/MS for

dimethylsilanediol quantification and extended the method to soil, sediment and biosolids samples for trace contaminant quantification down to 2 μg/L (direct water analysis) and 10 to 100 ng/g d.w. (solid analysis) (Xu and Knoerr, 2020) concentrations. Evgenidou et al. reviewed the detections of PPCP TPs and illicit drugs in municipal waters, where most of the reported studies leveraged targeted analysis with low resolution tandem mass spectrometry (Evgenidou et al., 2015). Similarly, Hou et al. developed a LC-MS/MS method that covers 39 analytes; reporting the widespread occurrence of stormwater-derived chemicals and quantifying them in urban waters (Hou et al., 2019).

In addition to the conventional primary contaminants (i.e., parent compounds), TPs could also be characterized by targeted methods when standards are available. For example, Kolpin et al. reported the occurrence of TPs of cyanazine, which are more prevalent than the parent compound, in groundwater utilizing low resolution tandem mass spectrometry (Kolpin et al., 2001; Scribner et al., 2000). As reported by Tian et al. (Tian et al., 2021a, 2021b), 6PPDQ, a TP formed from tire rubber antioxidant 6PPD, is capable of inducing acute mortality of exposed coho salmon at environmentally relevant concentrations (LC_{50}=95 ng/L). Via chemical formula match of $C_{18}H_{22}N_2O_2$, 6PPDQ was firstly reported as a 6PPD ozonation TP by Lattimer et al. (Lattimer et al., 1983), but was incorrectly annotated as a "dinitrone product" in the original report. Building on previous studies (Cataldo et al., 2015; Huntink, 2003; Lattimer et al., 1983) to better understand the formation of 6PPDQ and transformation of its precursor chemical 6PPD, we conducted a series of oxidative exposure experiments of 6PPD and measured 6PPDQ yield during 6PPD transformation process (**Chapter 3 and 4**).

1.2.2 Suspect Screening

While targeted analysis works very well when employed with 'known chemicals' (i.e., chemicals known to be present in the samples with available reference standards), it is usually limited to narrower (by number) chemical screening studies due to the lack of reference standards for so many possible environmental contaminants. To broaden the scope of screening in complex mixtures, suspect screening methods screen large lists of pre-prioritized chemicals of interest by their m/z ratio, retention time, chemical formula or MS^2 information (Hollender et al., 2019). For example, Gerona et al. (Gerona et al., 2018) screened serum samples using LC-QTOF against a suspect list and developed a workflow to effectively monitor environmental organic acids, providing a comprehensive evaluation of human exposures to these chemicals (**Figure 1.5**).

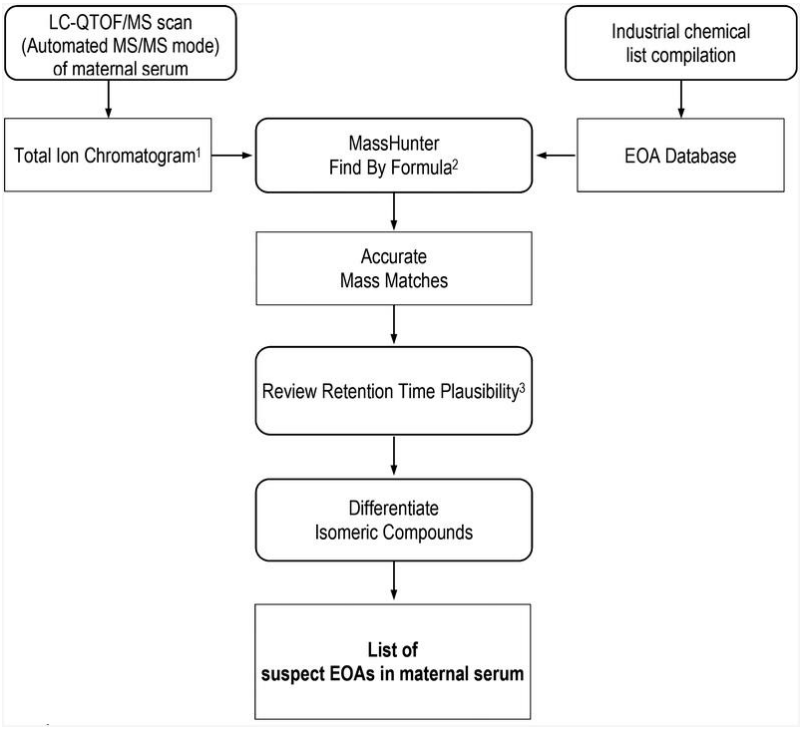


Figure 1.5. Example suspect screening workflow to identify suspect environmental organic acids in pregnant women serum utilizing LC-QTOF/MS scan with pre-determined list and external database, published by Gerona et al (2018).

Compared to targeted analysis, suspect screening doesn't require commercial standards to optimize chemical detection and therefore can cover a much wider range of chemicals based on their expected exact mass. It is therefore well-suited to screen potential TPs of known or suspected parent compound structures based on reaction pathway predictions, as structural similarity itself is a prioritization strategy. For example, previous assessments of 6PPD fate have primarily focused on its industrial chemistry within rubber formulations and involving processes simulating the tire rubber working conditions such as thermal and oxidative transformation (Cataldo et al., 2015; Huang et al., 2001; Kalkornsurapranee et al., 2013; Lattimer et al., 1983). Generally, such data have limited connectivity to environmental systems and often used less comprehensive or sophisticated analytical approaches until recently (Klöckner et al., 2021a; Seiwert et al., 2022; Tian et al., 2021b, 2021a). Given the presence of 6PPDQ and its yield (~10%; determined in **Chapter 3**) during 6PPD ozonation process, we hypothesize that other major TPs form during the 6PPD transformation process. In this thesis, suspect screening approaches were deployed to investigate reaction kinetics of the 6PPD derived TPs under different conditions (**Chapters 3 and 4**).

1.2.3 Non-targeted Analysis

Although targeted and suspect screening methods are powerful for the identification and quantification of a pre-selected range of chemicals, broad-spectrum analytical methods that

provide effective characterization and identification within complex mixtures containing many unknown or unanticipated compounds also are necessary. NTA workflows derived from HRMS data acquisition can effectively address this need, providing comprehensive characterization of chemical composition, identification of previously unknown contaminants, evaluation of compositional change across samples, and tracking of contaminant sources through the environment (Gonzalez de Vega et al., 2021; Peter et al., 2019; Schollée et al., 2018; Tian et al., 2021b; Wang et al., 2021). For instance, NTA studies have revealed the widespread contamination of PFAS chemicals (Banzhaf et al., 2017; Sun et al., 2016) in drinking waters that led to a series of remediation efforts and regulation policies (Ross et al., 2018; US EPA, 2016). A typical NTA workflow is summarized by Hollender et al (**Figure 1.6**) (Hollender et al., 2019).

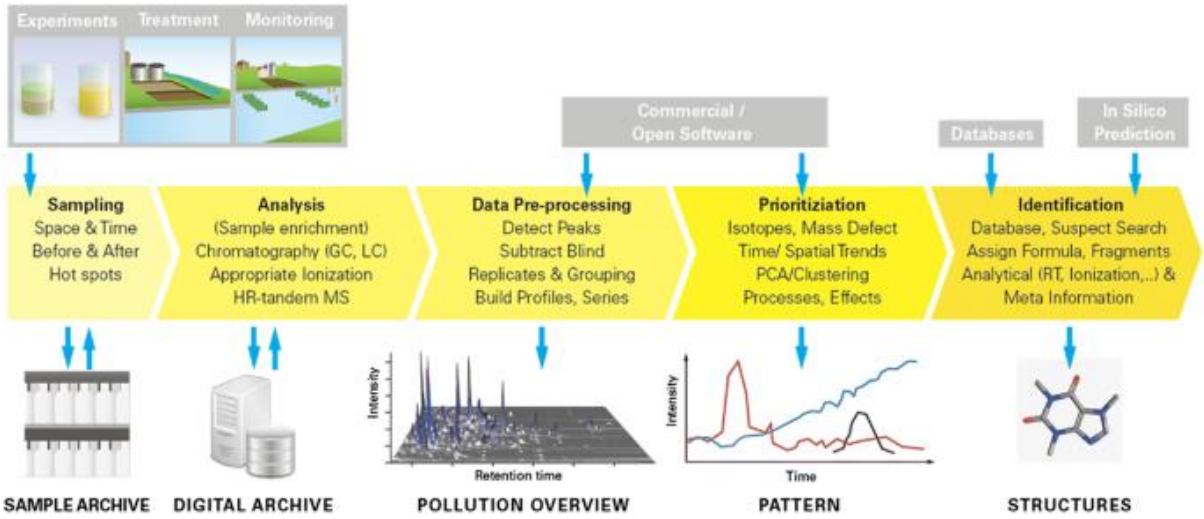


Figure 1.6. Typical workflow for NTA of environmental samples, from Hollender et al. (2019).

Broad-spectrum chemical detection capabilities present significant challenges in data processing because of the complex and massive HRMS datasets. An efficient data processing workflow coupled with substantial computational resources is needed to reduce and align the

data, exclude noisy detections, and extract meaningful information from the NTA datasets. Ideally, this process requires a fully or semi-automated workflow to reduce the tedious laborious work while also ensuring a confident outcome. Machine learning algorithms, referring to a broad range of algorithms that are trained on a given dataset to deliver intelligent predictions about complementary datasets (Nichols et al., 2018), seem to fit this problem especially well. Machine learning is commonly employed for large datasets where manual operation is impractical. Common applications of machine learning include classification, regression, image recognition, or natural language processing, with different machine learning algorithms (supervised learning, unsupervised learning, reinforcement learning, etc.) involved (Zhang, 2010). Machine learning algorithms have also been applied to LC-MS data processing. Melnikov et al. developed a deep learning-based algorithm to assist LC-MS chromatogram peak picking process with high accuracy (Melnikov et al., 2020). Leveraging message-passing neural networks, Osipenko et al. developed a workflow for LC-MS feature retention time prediction (Osipenko et al., 2022). In addition, these algorithms are also popular for metabolomics applications to support clinical decisions, guide metabolic engineering and facilitate biological studies (Liebal et al., 2020).

As above, machine learning algorithms could facilitate a NTA workflow. In NTA workflows that are typically screening datasets containing many thousands of chemical features, feature prioritization is one of the most critical data processing steps. This step identifies potential features of concern, effectively reduces the size of the dataset, and facilitates further analysis (Blum et al., 2017; Du et al., 2017; Schollée et al., 2018; Tian et al., 2021b). Various

machine learning based clustering analysis algorithms provide the capability of efficient feature grouping and prioritization to aid data analysis and find system critical features. One of the most commonly used unsupervised clustering algorithms is density-based spatial clustering of applications with noise (DBSCAN) (Ester et al., n.d.). Using this approach, features with similar behavior patterns (e.g., continuous trends or regular fluctuations) are automatically clustered with user-selected parameters, while outliers that diverge from recognized trends are effectively excluded. Consequently, features with specific behavior patterns (e.g., generated during chemical reactions or resistant to water treatment process) would be effectively prioritized and available for further investigations.

Another promising application of machine learning algorithms to NTA data is source apportionment, reflecting the quantitative prediction of chemical contaminant contributions from various possible sources to environmental systems (Peter et al., 2019). For example, selected HRMS features are used to populate predictive models to estimate user desired parameters or outcomes, such as contamination source concentration, sample matrix effect classification based on the training data that are inputs to user selected models. Exploiting the fast-growing research interest around machine learning and data science, the combination of such novel computational methodology with analytical sciences can facilitate environmental NTA studies. Because the HRMS data is usually parsed as metadata (datasets consist of m/z & RT pair, detection abundance and sample labels), assuming the detected chemical features would be intrinsically relevant (e.g., the relative RT/detection abundance difference between features might be consistent in the same sample matrix) or sample dependent (e.g., the detection

abundances correlate with contaminant source concentrations, or features from the same source have similar abundance change across a series of samples), they could be processed using 'data science' techniques – for example, using statistical methods or modeling approaches employing all the detected features rather than small subsets of a single or a few chemicals to understand the system.

Data processing software, including commercial and open-source software, is usually employed for user defined data analysis workflow. However, because NTA HRMS data processing workflow development still requires considerable effort and lacks well-established protocols, currently available software packages can sometimes have technical or analytical issues with data sharing, function flexibility or processing speed. Also, most software packages do not easily support machine learning algorithms or customized workflows (e.g., source tracking workflow). To address the absence of customizable NTA workflow for environmental sample analysis, open-source tools and methodologies for handling HRMS data and implementing NTA workflows are urgently needed. In this thesis, an open-source Python package was developed that could implement user-defined NTA outcomes (**Chapter 5**) and applied the workflow in **Chapter 3** to prioritize potential environmentally relevant TPs from 6PPD. We also proposed a source identification and apportionment workflow in **Chapter 6** to facilitate data processing for HRMS datasets for source tracking studies.

1.3 Thesis Structure

To address all the key research needs and address existing data gaps around 6PPD fate, and TP formation and identification, HRMS analytical and software methodologies, we focused on investigation of the emerging tire rubber derived contaminant 6PPD and its TPs, especially 6PPDQ, while developing NTA workflows to better improve HRMS data screening capabilities and strategies. The thesis communicates these study efforts and is organized into six chapters. **Chapter 1** provides an introduction and explains the research goals of the thesis. **Chapter 2** describes the investigation of the physiochemical properties, including logK_{ow}, solubility, leaching potentials, sorption potentials, and aqueous stability of 6PPDQ. Notably, we simulated environmental leaching conditions with a lab-scale flow through leaching system and derived an empirical model to describe 6PPDQ leaching kinetics from TWP with targeted approaches (e.g., LC-MS/MS). In **Chapters 3 and 4**, we developed and leveraged the LC-MS/MS-based targeted analysis method and LC-QTOF-based suspect screening approach to detect and measure 6PPD and 6PPD TPs via ozonation reactions under various conditions. **Chapter 3** describes the transformation kinetics of 6PPD degradation and 6PPDQ formation using heterogeneous 6PPD ozonation in the gas phase. Aside from 6PPDQ, other potential TPs (N=19) were prioritized by applying a machine learning clustering algorithm. In addition, by comparatively screening environmental samples, 9 TPs were detected in surface runoff, suggesting that 6PPDQ and other related TPs could form and release into the environment during transformation of 6PPD in tire rubbers and environmental systems. Extending from **Chapter 3**, **Chapter 4** describes 6PPD environmental fate and transformation under additional

conditions. Different reaction systems were built to simulate 6PPD transformation processes, including gas phase oxidation and aqueous oxidative exposures. Concurrently with the above efforts, a customized data processing workflow was developed to assist the NTA, especially focusing on the TP prioritization process. **Chapter 5** describes the development of an open-source Python package *Mass-suite (MSS)* as a full, expandable open-source HRMS data analysis toolbox with multiple NTA data processing capabilities. The *MSS* package provides various functions and covers basic data processing (e.g., feature extraction and data visualization), and advanced data analysis (e.g., unsupervised clustering analysis, source tracking modeling). *MSS* is driven by machine learning algorithms that are not currently available within other NTA workflows or tools. Applications of *MSS* are described in **Chapter 3 and 4** to facilitate data analysis of complex chemical mixtures. **Chapter 6** evaluates the performance of different models for source tracking studies, including the accuracy, flexibility and generalizability of different models dealing with samples with different source concentration and background chemical matrices. Subsequently, we propose a data processing workflow of contamination source identification and quantification based on the machine learning modeling algorithms extended from **Chapter 5**.

References

Aceña, J., Stampachiacchiere, S., Pérez, S., Barceló, D. 2015. Advances in liquid chromatography-high-resolution mass spectrometry for quantitative and qualitative environmental analysis. Anal Bioanal Chem. 407(21):6289-99. doi: 10.1007/s00216-015-8852-6

Alhelou, R., Seiwert, B., Reemtsma, T., 2019. Hexamethoxymethylmelamine – A precursor of persistent and mobile contaminants in municipal wastewater and the water cycle. Water Research 165, 114973. https://doi.org/10.1016/j.watres.2019.114973

Aydoğan, C. 2020. Recent advances and applications in LC-HRMS for food and plant natural products: a critical review. Anal Bioanal Chem. 412(9):1973-1991. doi: 10.1007/s00216-019-02328-6. Epub 2020 Jan 7. PMID: 31912180.

Banzhaf, S., Filipovic, M., Lewis, J., Sparrenbom, C.J., Barthel, R., 2017. A review of contamination of surface-, ground-, and drinking water in Sweden by perfluoroalkyl and polyfluoroalkyl substances (PFASs). Ambio 46, 335–346. https://doi.org/10.1007/s13280-016-0848-8

Blum, K.M., Andersson, P.L., Renman, G., Ahrens, L., Gros, M., Wiberg, K., Haglund, P., 2017. Non-target screening and prioritization of potentially persistent, bioaccumulating and toxic domestic wastewater contaminants and their removal in on-site and large-scale sewage treatment plants. Sci. Total Environ. 575, 265–275. https://doi.org/10.1016/j.scitotenv.2016.09.135

Brinkmann, M., Montgomery, D., Selinger, S., Miller, J.G.P., Stock, E., Alcaraz, A.J., Challis, J.K., Weber, L., Janz, D., Hecker, M., Wiseman, S. 2022. Acute Toxicity of the Tire Rubber-Derived Chemical 6PPD-quinone to Four Fishes of Commercial, Cultural, and Ecological Importance. Environ. Sci. Technol. Lett. 2022, 9, 4, 333-338, https://doi.org/10.1021/acs.estlett.2c00050

Campanale, C., Massarelli, C., Savino, I., Locaputo, V., Uricchio, V.F., 2020. A Detailed Review Study on Potential Effects of Microplastics and Additives of Concern on Human Health. Int J Environ Res Public Health 17, 1212. https://doi.org/10.3390/ijerph17041212

Cao, G., Wang, W., Zhang, J., Wu, P., Zhao, X., Yang, Z., Hu, D., Cai, Z., 2022. New Evidence of Rubber-Derived Quinones in Water, Air, and Soil. Environ. Sci. Technol. 56, 4142–4150. https://doi.org/10.1021/acs.est.1c07376

Cataldo, F., Faucette, B., Huang, S., Ebenezer, W., 2015. On the early reaction stages of ozone with N,N′-substituted p-phenylenediamines (6PPD, 77PD) and N,N′,N″-substituted-1,3,5-triazine "Durazone®": An electron spin resonance (ESR) and electronic absorption spectroscopy study. Polymer Degradation and Stability 111, 223–231. https://doi.org/10.1016/j.polymdegradstab.2014.11.011

Chen, S., Liu, C., Peng, C., Liu, H., Hu, M., Zhong, G., 2012. Biodegradation of Chlorpyrifos and Its Hydrolysis Product 3,5,6-Trichloro-2-Pyridinol by a New Fungal Strain Cladosporium cladosporioides Hu-01. PLoS One 7. https://doi.org/10.1371/journal.pone.0047205

Chen, X., He, T., Yang, X., Gan, Y., Qing, X., Wang, J., Huang, Y., 2023. Analysis, environmental occurrence, fate and potential toxicity of tire wear compounds 6PPD and

6PPD-quinone. J Hazard Mater 452, 131245. https://doi.org/10.1016/j.jhazmat.2023.131245

Councell, T.B., Duckenfield, K.U., Landa, E.R., Callender, E., 2004. Tire-Wear Particles as a Source of Zinc to the Environment. Environ. Sci. Technol. 38, 4206–4214. https://doi.org/10.1021/es034631f

D'Avignon, G., Gregory-Eaves, I., Ricciardi, A., 2022. Microplastics in lakes and rivers: an issue of emerging significance to limnology. Environ. Rev. 30, 228–244. https://doi.org/10.1139/er-2021-0048

Dom, I., Biré, R., Hort, V., Lavison-Bompard, G., Nicolas, M., Guérin, T., 2018. Extended Targeted and Non-Targeted Strategies for the Analysis of Marine Toxins in Mussels and Oysters by (LC-HRMS). Toxins 10, 375. https://doi.org/10.3390/toxins10090375

Drzyzga, O., 2003. Diphenylamine and derivatives in the environment: a review. Chemosphere 53, 809–818. https://doi.org/10.1016/S0045-6535(03)00613-1

Du, B., Lofton, J.M., Peter, K.T., Gipe, A.D., James, C.A., McIntyre, J.K., Scholz, N.L., Baker, J.E., Kolodziej, E.P., 2017. Development of suspect and non-target screening methods for detection of organic contaminants in highway runoff and fish tissue with high-resolution time-of-flight mass spectrometry. Environ. Sci.: Processes Impacts 19, 1185–1196. https://doi.org/10.1039/C7EM00243B

Ebele, A.J., Abou-Elwafa Abdallah, M., Harrad, S., 2017. Pharmaceuticals and personal care products (PPCPs) in the freshwater aquatic environment. Emerging Contaminants 3, 1–16. https://doi.org/10.1016/j.emcon.2016.12.004

Escher, B.I., Fenner, K., 2011. Recent Advances in Environmental Risk Assessment of Transformation Products. Environ. Sci. Technol. 45, 3835–3847. https://doi.org/10.1021/es1030799

Ester, M., Kriegel, H.-P., Xu, X., n.d. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise 6.

Freyria, F.S., Geobaldo, F., Bonelli, B., 2018. Nanomaterials for the Abatement of Pharmaceuticals and Personal Care Products from Wastewater. Applied Sciences 8, 170. https://doi.org/10.3390/app8020170

Gerona, R.R., Schwartz, J.M., Pan, J., Friesen, M.M., Lin, T., Woodruff, T.J., 2018. Suspect Screening of Maternal Serum to Identify New Environmental Chemical Biomonitoring Targets using Liquid Chromatography-Quadrupole Time-Of-Flight Mass Spectrometry. J Expo Sci Environ Epidemiol 28, 101–108. https://doi.org/10.1038/jes.2017.28

Gonzalez de Vega, R., Cameron, A., Clases, D., Dodgen, T.M., Doble, P.A., Bishop, D.P., 2021. Simultaneous targeted and non-targeted analysis of per- and polyfluoroalkyl substances in environmental samples by liquid chromatography-ion mobility-quadrupole time of flight-mass spectrometry and mass defect analysis. J. Chromatogr. A 1653, 462423. https://doi.org/10.1016/j.chroma.2021.462423

Hollender, J., Schymanski, E.L., Singer, H.P., Ferguson, P.L., 2017. Nontarget Screening with High Resolution Mass Spectrometry in the Environment: Ready to Go? Environ. Sci. Technol. 51, 11505–11512. https://doi.org/10.1021/acs.est.7b02184

Hollender, J., van Bavel, B., Dulio, V., Farmen, E., Furtmann, K., Koschorreck, J., Kunkel, U., Krauss, M., Munthe, J., Schlabach, M., Slobodnik, J., Stroomberg, G., Ternes, T., Thomaidis, N.S., Togola, A., Tornero, V., 2019. High resolution mass spectrometry-

based non-target screening can support regulatory environmental monitoring and chemicals management. Environmental Sciences Europe 31, 42. https://doi.org/10.1186/s12302-019-0225-x

Hou, F., Tian, Z., Peter, K.T., Wu, C., Gipe, A.D., Zhao, H., Alegria, E.A., Liu, F., Kolodziej, E.P., 2019. Quantification of organic contaminants in urban stormwater by isotope dilution and liquid chromatography-tandem mass spectrometry. Anal Bioanal Chem 411, 7791–7806. https://doi.org/10.1007/s00216-019-02177-3

Hu, X., Zhao, H.N., Tian, Z., Peter, K.T., Dodd, M.C., Kolodziej, E.P., 2022. Transformation Product Formation upon Heterogeneous Ozonation of the Tire Rubber Antioxidant 6PPD (N-(1,3-dimethylbutyl)-N′-phenyl-p-phenylenediamine). Environ. Sci. Technol. Lett. 9, 413–419. https://doi.org/10.1021/acs.estlett.2c00187

Huang, D., LaCount, B.J., Castro, J.M., Ignatz-Hoover, F., 2001. Development of a service-simulating, accelerated aging test method for exterior tire rubber compounds I. Cyclic aging. Polymer Degradation and Stability 74, 353–362. https://doi.org/10.1016/S0141-3910(01)00185-9

Huang, P., 2017. In situ detection of water quality contamination events based on signal complexity analysis using online ultraviolet-visible spectral sensor [WWW Document]. URL https://opg.optica.org/ao/abstract.cfm?uri=ao-56-22-6317 (accessed 8.8.22).

Hübner, U., von Gunten, U., Jekel, M., 2015. Evaluation of the persistence of transformation products from ozonation of trace organic compounds – A critical review. Water Research 68, 150–170. https://doi.org/10.1016/j.watres.2014.09.051

Huntink, N.M., 2003. Durability of rubber products: Development of new antidegradants for long-term protection.

Jin, X., Peldszus, S., Huck, P.M., 2012. Reaction kinetics of selected micropollutants in ozonation and advanced oxidation processes. Water Research 46, 6519–6530. https://doi.org/10.1016/j.watres.2012.09.026

Johannessen, C., Helm, P., Metcalfe, C.D., 2021. Runoff of the Tire-Wear Compound, Hexamethoxymethyl-Melamine into Urban Watersheds. Arch Environ Contam Toxicol. https://doi.org/10.1007/s00244-021-00815-5

Kalkornsurapranee, E., Nakason, C., Kummerlöwe, C., Vennemann, N., 2013. Effect of Antioxidant on Properties of Thermoplastic Natural Rubber Based on ENR/TPU Blends. Advanced Materials Research 626, 229–232. https://doi.org/10.4028/www.scientific.net/AMR.626.229

Klöckner, P., Seiwert, B., Wagner, S., Reemtsma, T., 2021a. Organic Markers of Tire and Road Wear Particles in Sediments and Soils: Transformation Products of Major Antiozonants as Promising Candidates. Environ. Sci. Technol. https://doi.org/10.1021/acs.est.1c02723

Klöckner, P., Seiwert, B., Weyrauch, S., Escher, B.I., Reemtsma, T., Wagner, S., 2021b. Comprehensive characterization of tire and road wear particles in highway tunnel road dust by use of size and density fractionation. Chemosphere 279, 130530. https://doi.org/10.1016/j.chemosphere.2021.130530

Kole, P.J., Löhr, A.J., Van Belleghem, F.G.A.J., Ragas, A.M.J., 2017. Wear and Tear of Tyres: A Stealthy Source of Microplastics in the Environment. Int J Environ Res Public Health 14. https://doi.org/10.3390/ijerph14101265

Kroiss, H., Rechberger, H., Egle, L., 2011. Phosphorus in Water Quality and Waste Management, Integrated Waste Management - Volume II. IntechOpen. https://doi.org/10.5772/18482

Kumar, R., Vuppaladadiyam, A.K., Antunes, E., Whelan, A., Fearon, R., Sheehan, M., Reeves, L., 2022. Emerging contaminants in biosolids: Presence, fate and analytical techniques. Emerging Contaminants 8, 162–194. https://doi.org/10.1016/j.emcon.2022.03.004

Lattimer, R.P., Hooser, E.R., Layer, R.W., Rhee, C.K., 1983. Mechanisms of Ozonation of N-(1,3-Dimethylbutyl)-N′-Phenyl-p-Phenylenediamine. Rubber Chemistry and Technology 56, 431–439. https://doi.org/10.5254/1.3538136

Liebal, U.W., Phan, A.N.T., Sudhakar, M., Raman, K., Blank, L.M., 2020. Machine Learning Applications for Mass Spectrometry-Based Metabolomics. Metabolites 10, 243. https://doi.org/10.3390/metabo10060243

Lin, L., Yang, H., Xu, X., 2022. Effects of Water Pollution on Human Health and Disease Heterogeneity: A Review. Frontiers in Environmental Science 10.

Melnikov, A.D., Tsentalovich, Y.P., Yanshole, V.V., 2020. Deep Learning for the Precise Peak Detection in High-Resolution LC–MS Data. Anal. Chem. 92, 588–592. https://doi.org/10.1021/acs.analchem.9b04811

Nichols, J.A., Herbert Chan, H.W., Baker, M.A.B., 2018. Machine learning: applications of artificial intelligence to imaging and diagnosis. Biophys Rev 11, 111–118. https://doi.org/10.1007/s12551-018-0449-9

Noguera-Oviedo, K., Aga, D.S., 2016. Lessons learned from more than two decades of research on emerging contaminants in the environment. Journal of Hazardous Materials 316, 242–251. https://doi.org/10.1016/j.jhazmat.2016.04.058

Osipenko, S., Nikolaev, E., Kostyukevich, Y. 2022. Retention Time Prediction with Message-Passing Neural Networks. Separations 2022, 9, 291. https://doi.org/10.3390/separations9100291

OSPAR Commission, 2006c. OSPAR background document on 4-(dimethylbutylamino)diphenylamine (6PPD).

Patel, A.B., Shaikh, S., Jain, K.R., Desai, C., Madamwar, D., 2020. Polycyclic Aromatic Hydrocarbons: Sources, Toxicity, and Remediation Approaches. Front Microbiol 11, 562813. https://doi.org/10.3389/fmicb.2020.562813

Peter, K.T., Tian, Z., Wu, C., Lin, P., White, S., Du, B., McIntyre, J.K., Scholz, N.L., Kolodziej, E.P., 2018. Using High-Resolution Mass Spectrometry to Identify Organic Contaminants Linked to Urban Stormwater Mortality Syndrome in Coho Salmon. Environ. Sci. Technol. 52, 10317–10327. https://doi.org/10.1021/acs.est.8b03287

Peter, K.T., Wu, C., Tian, Z., Kolodziej, E.P., 2019. Application of Nontarget High Resolution Mass Spectrometry Data to Quantitative Source Apportionment. Environ. Sci. Technol. 53, 12257–12268. https://doi.org/10.1021/acs.est.9b04481

Picó, Y., Barceló, D., 2015. Transformation products of emerging contaminants in the environment and high-resolution mass spectrometry: a new horizon. Anal Bioanal Chem 407, 6257–6273. https://doi.org/10.1007/s00216-015-8739-6

Pironti, C., Ricciardi, M., Motta, O., Miele, Y., Proto, A., Montano, L., 2021. Microplastics in the Environment: Intake through the Food Web, Human Exposure and Toxicological Effects. Toxics 9, 224. https://doi.org/10.3390/toxics9090224

Puckett, L.J., 1995. Identifying the major sources of nutrient water pollution 7.

René P. Schwarzenbach, 2010. Global Water Pollution and Human Health | Annual Review of Environment and Resources [WWW Document]. URL https://www.annualreviews.org/doi/full/10.1146/annurev-environ-100809-125342 (accessed 8.8.22).

Rødland, E.S., Lind, O.C., Reid, M.J., Heier, L.S., Okoffo, E.D., Rauert, C., Thomas, K.V., Meland, S., 2022. Occurrence of tire and road wear particles in urban and peri-urban snowbanks, and their potential environmental implications. Science of The Total Environment 824, 153785. https://doi.org/10.1016/j.scitotenv.2022.153785

Ross, I., McDonough, J., Miles, J., Storch, P., Kochunarayanan, P.T., Kalve, E., Hurst, J., Dasgupta, S.S., Burdick, J., 2018. A review of emerging technologies for remediation of PFASs. Remediation Journal 28, 101–126. https://doi.org/10.1002/rem.21553

Schollée, J.E., Bourgin, M., von Gunten, U., McArdell, C.S., Hollender, J., 2018. Non-target screening to trace ozonation transformation products in a wastewater treatment train including different post-treatments. Water Res. 142, 267–278. https://doi.org/10.1016/j.watres.2018.05.045

Seiwert, B., Nihemaiti, M., Troussier, M., Weyrauch, S., Reemtsma, T., 2022. Abiotic oxidative transformation of 6-PPD and 6-PPD quinone from tires and occurrence of their products in snow from urban roads and in municipal wastewater. Water Research 212, 118122. https://doi.org/10.1016/j.watres.2022.118122

Sheridan, M., 2010. The Vanderbilt Rubber Handbook, 14th edition. R.T. Vanderbilt Company, Inc., Norwalk, CT.

Sun, M., Arevalo, E., Strynar, M., Lindstrom, A., Richardson, M., Kearns, B., Pickett, A., Smith, C., Knappe, D.R.U., 2016. Legacy and Emerging Perfluoroalkyl Substances Are Important Drinking Water Contaminants in the Cape Fear River Watershed of North Carolina. Environ. Sci. Technol. Lett. 3, 415–419. https://doi.org/10.1021/acs.estlett.6b00398

Tian, Z., Gonzalez, M., Rideout, C., Zhao, H., Hu, X., Wetzel, J., Mudrock, E., James, C.A., McIntyre, J.K., Kolodziej, E.P., 2021a. 6PPD-Quinone: Revised Toxicity Assessment and Quantification Method Development with a Commercial Standard. Environmental Science & Technology Letters.

Tian, Z., Zhao, H., Peter, K.T., Gonzalez, M., Wetzel, J., Wu, C., Hu, X., Prat, J., Mudrock, E., Hettinger, R., Cortina, A.E., Biswas, R.G., Kock, F.V.C., Soong, R., Jenne, A., Du, B., Hou, F., He, H., Lundeen, R., Gilbreath, A., Sutton, R., Scholz, N.L., Davis, J.W., Dodd, M.C., Simpson, A., McIntyre, J.K., Kolodziej, E.P., 2021b. A ubiquitous tire rubber–derived chemical induces acute mortality in coho salmon. Science 371, 185–189. https://doi.org/10.1126/science.abd6951

U. S. Government Accountability, 2022. Thousands of Discharges Keep Pollution Flowing: How can EPA better protect Our Nation's Waters? [WWW Document]. URL https://www.gao.gov/blog/thousands-discharges-keep-pollution-flowing-how-can-epa-better-protect-our-nations-waters (accessed 8.8.22).

US EPA, O., 2016. Per- and Polyfluoroalkyl Substances (PFAS) [WWW Document]. URL https://www.epa.gov/pfas (accessed 8.10.22).

US EPA, O., 2015. National Primary Drinking Water Regulations [WWW Document]. URL https://www.epa.gov/ground-water-and-drinking-water/national-primary-drinking-water-regulations (accessed 8.8.22).

Varshney, S., Gora, A.H., Siriyappagouder, P., Kiron, V., Olsvik, P.A., 2022. Toxicological effects of 6PPD and 6PPD quinone in zebrafish larvae. Journal of Hazardous Materials 424, 127623. https://doi.org/10.1016/j.jhazmat.2021.127623

Wagner, S., Hüffer, T., Klöckner, P., Wehrhahn, M., Hofmann, T., Reemtsma, T., 2018. Tire wear particles in the aquatic environment - A review on generation, analysis, occurrence, fate and effects. Water Res 139, 83–100. https://doi.org/10.1016/j.watres.2018.03.051

Wang, T., Duedahl-Olesen, L., Lauritz Frandsen, H., 2021. Targeted and non-targeted unexpected food contaminants analysis by LC/HRMS: Feasibility study on rice. Food Chemistry 338, 127957. https://doi.org/10.1016/j.foodchem.2020.127957

Wang, Z., Walker, G.W., Muir, C.G., Nagatani-Yoshida, K., 2020 Toward a Global Understanding of Chemical Pollution: A First Comprehensive Analysis of National and Regional Chemical Inventories. Environ. Sci. Technol 2020 54(5), 2575-2584. DOI: 10.1021/acs.est.9b06379

Weatherly, L.M., Gosse, J.A., 2017. Triclosan Exposure, Transformation, and Human Health Effects. J Toxicol Environ Health B Crit Rev 20, 447–469. https://doi.org/10.1080/10937404.2017.1399306

Williams, A.J., Grulke, C.M., Edwards, J., McEachran, A.D., Mansouri, K., Baker, N.C., Patlewicz, G., Shah, I., Wambaugh, J.F., Judson, R.S., Richard, A.M., 2017. The CompTox Chemistry Dashboard: a community data resource for environmental chemistry. J Cheminform 9, 61. https://doi.org/10.1186/s13321-017-0247-6

Yang, Y., Ok, Y.S., Kim, K.-H., Kwon, E.E., Tsang, Y.F., 2017. Occurrences and removal of pharmaceuticals and personal care products (PPCPs) in drinking water and water/sewage treatment plants: A review. Sci Total Environ 596–597, 303–320. https://doi.org/10.1016/j.scitotenv.2017.04.102

Zeng, L., Li, Y., Sun, Y., Liu, L.Y., Shen, M., Du, B. 2023. Widespread Occurrence and Transport of p-Phenylenediamines and Their Quinones in Sediments across Urban Rivers, Estuaries, Coasts, and Deep-Sea Regions. Environ Sci Technol. 2023 57(6):2393-2403. doi: 10.1021/acs.est.2c07652.

Zhang, H.-Y., Huang, Z., Liu, Y.-H., Hu, L.-X., He, L.-Y., Liu, Y.-S., Zhao, J.-L., Ying, G.-G., 2023. Occurrence and risks of 23 tire additives and their transformation products in an urban water system. Environment International 171, 107715. https://doi.org/10.1016/j.envint.2022.107715

Zhang, L.-H., Cheng, J.-H., You, X., Liang, X.-Y., Hu, Y.-Y., 2016. Photochemical defluorination of aqueous perfluorooctanoic acid (PFOA) by Fe(0)/GAC micro-electrolysis and VUV-Fenton photolysis. Environ Sci Pollut Res Int 23, 13531–13542. https://doi.org/10.1007/s11356-016-6539-y

Zhang, Y., 2010. New Advances in Machine Learning. BoD – Books on Demand.

Zhang, Y., Xu, C., Zhang, W., Qi, Z., Song, Y., Zhu, L., Dong, C., Chen, J., Cai, Z., 2021. p-Phenylenediamine Antioxidants in PM2.5: The Underestimated Urban Air Pollutants. Environ. Sci. Technol. https://doi.org/10.1021/acs.est.1c04500

Zhao, H.N., Tian, Z., Kim, K.E., Wang, R., Lam, K., Kolodziej, E.P., 2021. Biotransformation of Current-Use Progestin Dienogest and Drospirenone in Laboratory-Scale Activated Sludge Systems Forms High-Yield Products with Altered Endocrine Activity. Environ. Sci. Technol. 55, 13869–13880. https://doi.org/10.1021/acs.est.1c03805

Chapter 2: Chemical Characteristics, Leaching, and Stability of the Ubiquitous Tire Rubber-Derived Toxicant 6PPD-Quinone

2.0 Publication and contribution statement

One publication was incorporated for this chapter as below, and XH led most of the research and development effort. For Hu et al. 2023, XH designed and conducted the experiment, processed the samples, analyzed experimental data, and wrote the whole manuscript draft.

1. X. Hu, H. Zhao, Z. Tian, K. T. Peter, M. C. Dodd, and E. P. Kolodziej. "Chemical characteristics, leaching, and stability of the ubiquitous tire rubber-derived toxicant 6PPD-quinone". *Environmental Science: Processes & Impacts*, 2023, Advance Article. https://doi.org/10.1039/D3EM00047H

2.1 Introduction

Complementing decades of investigation of conventional water pollutants such as nutrients, salts, and trace heavy metals, (Ahamad et al., 2020; Braukmann and Böhme, 2011; Corsi et al., 2010; Kroiss et al., 2011; Puckett, 1995; Tchounwou et al., 2012; Zhou et al., 2020) various synthetic organic contaminants are now receiving more research attention in roadway systems (Alhelou et al., 2019; Hou et al., 2019; Klöckner et al., 2021b; Kumata et al., 2002; Peter et al., 2020, 2018; Unice et al., 2015; Wagner et al., 2018). Recently, 6PPD-quinone (2-((4-methylpentan-2-yl)amino)-5-(phenylamino)cyclohexa-2,5-diene-1,4-dione) or "6PPDQ", a tire rubber derived compound that can induce acute toxicity to coho salmon (*Oncorhynchus kisutch*), was reported globally in atmospheric particulate matter, dusts, soils, runoff, surface

waters and even human urine.(Du et al., 2022; Huang et al., 2021; Klöckner et al., 2021a; Seiwert et al., 2022; Tian et al., 2021b, 2021a; Wang et al., 2022, p. 5; Zhang et al., 2021, p. 5) 6PPDQ is a transformation product of the common antioxidant 6PPD (*N*-(1,3-dimethylbutyl)-*N'*-phenyl-*p*-phenylenediamine) that is ubiquitously used in tire rubber products (0.4-2% by mass).(Sheridan, 2010) The parent compound 6PPD converts into 6PPDQ via oxidative reactions, especially those involving reactions with atmospheric ozone, during the tire service life; 6PPDQ is therefore introduced into roadway environments by tire rubber use and wear.(Councell et al., 2004; Klöckner et al., 2021b; OSPAR Commission, 2006c; Seiwert et al., 2022; Tian et al., 2021b; Wagner et al., 2018) In addition to atmospheric pathways, both dissolved 6PPDQ and tire tread wear particles (TWP) containing 6PPDQ are transported in runoff from roadways during storm events where they can induce adverse ecological effects in sensitive aquatic species upon release.

Currently, there exists limited information concerning the chemical properties, environmental fate, and transport of 6PPDQ.(Klöckner et al., 2021a; Tian et al., 2021b, 2021a) To the best of our knowledge, the chemical properties and characteristics needed to inform the environmental risks of 6PPDQ have not been fully reported.(Drzyzga, 2003; Hiki et al., 2021; Huntink, 2003; OSPAR Commission, 2006c) To address this data gap, this study measured the aqueous solubility, logK_{OW}, sorption potential, leaching, and stability of 6PPDQ in environments typical of the laboratory and simple aqueous systems. Specifically, we measured or estimated: (1) basic physicochemical properties (logK_{OW} and solubility); (2) sorption potentials to different conventional materials; (3) leaching dynamics and timescales from representative TWP; (4) stability within simple aqueous solutions; and (5) potential for

environmental occurrence via mass balance calculations. These data will inform procedures for laboratory handling and analysis while guiding environmental fate and management strategies for 6PPDQ.

2.2 Materials and Methods

2.2.1 Chemicals and Reagents

6PPDQ (98.8%) and 6PPDQ-d_5 (98%) were purchased from HPC Standards Inc. (Atlanta, GA, USA). Sodium phosphate dibasic anhydrous, sodium phosphate monobasic anhydrous and sodium hydroxide solution (10N) were purchased from Fisher Scientific (Waltham, MA, USA). Methanol, acetone, formic acid, (all Optima LC/MS grade, ≥99%), absolute ethanol (200 Proof, ~99.5%), 1-octanol (ACS spectrophotometric grade, ≥99%) were purchased from Fisher Scientific (Waltham, MA, USA) and Sigma-Aldrich (St. Louis, MO, USA). Ultrapure deionized (DI) water was obtained using a water purification system (Thermo Barnstead Nanopure Diamond UV, Dubuque, IA, USA) or Milli-Q® Advantage A10 purification system (Millipore, Burlington, MA, USA). Glass Pasteur pipets were used for all the subsampling process for aqueous samples contains 6PPDQ to reduce the possible sorption loss of 6PPDQ. Alternative measurements including using PTFE lined tubing (e.g., during SPE process) or reduced contact time (e.g., <10s while pipetting 6PPDQ solution) were taken to handle 6PPDQ water solution when glass material is not available.

2.2.2 Aqueous Solubility and Octanol–Water Partition Coefficient

Solubility assessment of 6PPDQ followed the centrifugation method described in the OECD protocol for aquatic toxicity testing with modification.(OECD, 2019) Aqueous 6PPDQ solutions were prepared at different nominal concentrations (**Table 2.1**; 5-300 μg/L nominal) prior to experiments as follows.

Table 2.1. Aqueous 6PPDQ solution preparation conditions

Nominal Concentration (μg/L)	Spike Volume (μL)	Spiked Concentration (mg/L)	Final Volume (mL)	Actual Concentration (μg/L)*	Observed / Nominal Concentration (%)	Replicates
5	50	10	100	4.3 ± 0.3	85 ± 6	3
10	50	20	100	7.3 ± 0.2	73 ± 2	3
25	50	50	100	15.7 ± 0.3	63 ± 1	3
50	50	100	100	26.3 ± 1.9	53 ± 4	3
100	40	100	40	38.5 ± 2.1	39 ± 2	3
200	80	100	40	44.8 ± 0.8	22 ± 0	3
300	120	100	40	71.5 ± 16.5	24 ± 6	9

*The actual concentration of 6PPDQ stock solution is the same as the pre-centrifuge concentration of 6PPDQ in the **Figure 1** for each concentration group.

Pure aqueous phase 6PPDQ solutions were prepared for solubility, sorption assessment and stability experiments. Different volumes of ethanolic 6PPDQ stock solutions were transferred into amber glass bottles with Hamilton syringes, the solutions were applied to the bottom and side of the bottles as even as possible and dried thoroughly with high purity nitrogen (>99 %) at ~5 L/min flow rate for 150 s and placed within a fume hood for an additional 30 s to let excess organic solvent vapor evacuate. After drying, different volumes of ultrapure water were added into the bottles to reconstitute the aqueous 6PPDQ solution, the bottles were sonicated for 60 minutes and transferred onto a shaker (70 rpm/min) for mixing overnight (>12hr) before use. The mass recovery for the aqueous 6PPDQ solution at 5 μg/L concentration (100 μL of 100 mg/L) was measured in the preliminary studies at 85 ± 6 %.

The solutions were mixed on a shaker for >12 hours before experiments. Experiments were conducted in triplicate for each concentration. Given the tendency of 6PPDQ sorption to different materials, during all experiments the aqueous 6PPDQ solutions were stored in glass bottles and transferred with glass Pasteur pipets.

After equilibration, 10 mL aliquots from 6PPDQ aqueous stocks at different nominal concentrations (**Table 2.1**) were transferred into 15 mL high strength glass centrifuge tubes with rubber sleeves (DWK Life Sciences, Vineland, NJ, USA) and adapters (Thermo Scientific, Waltham, MA, USA). 500 μL aliquots were sampled from each tube before centrifugation. Then, centrifugation separations were performed with a Sorvall X1 centrifuge (equipped with a Fiberlite F15-8x50cy Fixed Angle Rotor) at 10,200 g for 30 min, with temperature control at 20 ℃. After centrifugation, 500 μL of supernatant was sampled from each tube. Samples were diluted with 450 μL methanol and spiked with 50 μL 6PPDQ-d_5 ISTD (5μg/L) for instrumental analysis **(**described in **Section 2.2.6).**

Octanol-water partition coefficient (logK_{OW}) measurements followed EPA guidelines with minor modification;(USEPA, 1979) experiments were conducted in triplicate. A 6PPDQ stock solution in ethanol (10 mg/L; 0.5 mL) was dried under gentle nitrogen flow (<1 L/min) within 15 mL glass centrifuge tubes, reconstituted with 1 mL octanol (precleaned with NaOH and ultrapure water to remove impurities) and 5 mL of ultrapure water, then sonicated (15 min). These mixtures were centrifuged (30 min, 2,500 rpm, 20 ℃) until complete separation of aqueous and organic phases. Octanol aliquots were withdrawn by glass pipet while remaining octanol phase was removed and discarded prior to aqueous phase sampling. Octanol

subsamples were diluted with methanol 1000-fold while aqueous phases were analyzed directly, given their low 6PPDQ concentration (~2 µg/L; well below the aqueous solubility limit). For instrumental analysis, 950 µL aliquots (both diluted octanol and aqueous phases) were spiked with 50 µL 6PPDQ-d5 ISTD.

2.2.3 Sorption to materials

An aqueous 6PPDQ stock solution (at 5 µg/L nominal concentration) was prepared as for solubility studies. All materials or containers were 6PPDQ free and pre-cleaned prior to experiments. The selected materials (sampling tube, polyethylene housing [ST-PE]; sampling tubing, polytetrafluoroethylene liner [ST-PTFE]; silicone sampling tubing [ST-Sil]; polytetrafluoroethylene SPE tubing [SPE-PTFE]; parafilm [PF]; rubber stopper [RS]; and bike tire rubber [BTR]) were cut into flat pieces (0.5 g) and placed within amber glass bottles. Centrifuge tubes (15 mL polypropylene centrifuge tube [CT-PP], 10 mL fluorinated ethylene propylene centrifuge tube [CT-FEP], and 15 mL glass centrifuge tube [CT-G]) and lab equipment (1L polypropylene sampling scoop [SS-PP] and stainless-steel attachment [SS]) were tested directly without further treatment. All sorption tests were conducted in triplicate for each material. Aliquots of 6PPDQ solution (10 mL for common materials and centrifuge tubes; 30 mL for tested containers) were transferred via glass pipets into the amber glass bottles containing material samples or the original testing containers (see **Table 2.2**). Tested vessels were mixed on the shaker (70 rpm) and 500 µL aliquots sampled with glass Pasteur pipets after 5 min, 30 min, and 24 h. These subsamples were diluted with 450 µL methanol and spiked with 50 µL 6PPDQ-d5 ISTD for instrumental analysis. To measure the total recoverable adsorbed 6PPDQ mass, after 24 hours mixing, aqueous 6PPDQ solution was removed from all the tested

containers and 10 mL methanol was added into the tubes. These were shaken and sonicated (30 min), then 950 μL aliquots were sampled and spiked with 50 μL 6PPDQ-d5 for analysis.

Table S2. Summary of materials used for sorption potential measurements.

	Testing material	Material Shape	6PPDQ stock volume added
Common Materials	Green stopper (GS)	Fragment pieces	10 mL
	Bike tire rubber (BTR)	Fragment pieces	10 mL
	Sampling Tube PE housing (ST-PE)	Coupon	10 mL
	Sampling Tube PTFE liner (ST-PTFE)	Coupon	10 mL
	Sampling Tube Silicone (ST-Silic)	Coupon	10 mL
	PTFE tubing for SPE (SPE-PTFE)	Coupon	10 mL
	Parafilm (PF)	Coupon	10 mL
Centrifuge Tubes	PP centrifuge tube (CT-PP)	Original container	10 mL
	FEP centrifuge tube (CT-FEP)	Original container	10 mL
	Glass centrifuge tube (CT-G)	Original container	10 mL
Containers	Stainless Steel (SS)	Cylinder shaped attachment	30 mL
	PP Sampling Scoop (SS-PP)	Original container	30 mL

2.2.4 Leaching dynamic of 6PPDQ from TWP

Preparation of TWP (focused on passenger car and light truck tires) for leaching studies was described elsewhere(Tian et al., 2021b). Tire tread wear particles (TWP) were made by physical abrasion of the treads of nine tires (brand, vehicle type, and use history in **Table 2.3**) using an angle grinder with a steel carbide disk (Parkes MCM90; Miller Tire, Wauseon, OH, U.S.A.) and mixed in equal weights. The TWPs were stored in a zip-lock bag to be isolated from ambient air prior to experiments.

Table 2.3. Information about the nine tires used to produce tire tread wear particles (TWPs), also reported in detail elsewhere.(Tian et al., 2021b) The TWP mixture used for these studies was an equal-weight mixture of the nine tire types listed below.

Tire ID	Manufacturer	Season	New/Used	Vehicle Type	Tire code
1	A	All-season	Used	Car	205/55R16 91H
2	B	All-season	Used	Car	195/70R14 91H
3	C	All-season	Used	Car	P185/65R15 86T
4	D	Winter	Used	Car	175/65R14 82S
5	E	All-season	Used	Car	P185/85R15 86S
6	E	All-season	Used	Light Truck	LT235/85R16 120R
7	F	All-season	New	Car	P205/60R15 90T
8	G	All-season	New	Car	P205/60R16 92H
9	H	All-season	Used	Car	P225/60R17 98T

The particle size distribution and surface area of TWP were characterized with surface analyzer and BET surface analysis (described in **Section 2.2.6**).

For leaching studies, 0.25 g of TWP were dispersed into 75 mL of pre-cleaned glass beads (0.3 cm diameter) in a glass beaker (100 mL). The mixture was transferred into a glass chromatography column (300 mm × 20 mm ID, fused mesh base); any residual TWP in the mixing beaker was transferred into the column as a slurry in 5 mL DI water (with transfer time limited to <30s to minimize pre-leaching contact time between water and TWPs). The column was then topped with two layers of clean glass beads and a stainless-steel mesh (**Figure S2.1**), ultrapure DI water was pumped through the column by a peristaltic pump at 45 mL/min using a single pass (i.e., effluent sent to waste) upflow design. Leachates were collected at 0.25, 0.5, 1, 1.5, 2, 3, 4, 5, and 6 h. Preliminary tests (15, 45, and 90 mL/min) determined optimal flow

rates without apparent advective limitations. These tests were conducted in triplicate with identical TWP mass and leaching time. At each time point, 200 mL of TWP leachate was collected, stored at 4 °C and extracted within 24 h.(Hou et al., 2019; Tian et al., 2021b) Leachates were extracted using preconditioned (10 mL methanol, 25mL water) SPE cartridges (Oasis HLB) at 5-10 mL/min. SPE cartridges were then rinsed with 10 mL DI water, air-dried (30 min), then eluted with methanol (2 × 5 mL). Eluates were concentrated under nitrogen to 1 mL for instrumental analysis. At the time this experiment was conducted, 6PPDQ-d_5 isotopic internal standard (ISTD) was not commercially available, so ISTD was not spiked into these extracts for recovery correction. Here, the method spiking recovery (i.e., 6PPDQ spiked leachate at t = 0, prior to leaching experiment) for 6PPDQ was measured as 69±7%, largely reflecting potential 6PPDQ loss during sample processing (e.g., SPE and evaporation concentration);(Tian et al., 2021b) concentration data were not normalized to observed recoveries for later calculations.

2.2.5 Aqueous stability

A 6PPDQ aqueous solution (20 μg/L) for aqueous stability tests was made using protocols described in **Section 2.2.2**. Aqueous stability measurements followed the methods of Su et al. with slight modification.(Su et al., 2016) Phosphate buffer (mixture of monopotassium and dipotassium phosphate) solutions (10 mM) were prepared and adjusted to pH 5, 7 and 9 (100 mL for each pH) using 0.1M sodium hydroxide. Twenty mL of 6PPDQ aqueous solution was then diluted with 80 mL buffer and separated into 20 mL triplicate aliquots in 100 mL amber glass bottles for stability tests, the bottles were capped with PTFE lined caps during the test. The remaining 40 mL pH adjusted 6PPDQ solution was retained identically and used for

pH measurements at 0, 7, 14 and 47 days using a pH meter (Thermo Scientific, Waltham, MA, USA). All bottles were mixed on a shaker under room temperature (20-22℃). The pH was generally stable (±0.05 for pH 5 and 7 groups, decreased to 8.7 for pH 9 solutions due to absorption of atmospheric carbon dioxide) throughout the experimental period. At each time point (0, 0.25, 1, 3, 7, 14, and 47 d), 500 μL of the solution was sampled with Pasteur glass pipet, diluted with 450 μL of methanol, spiked with 50 μL 6PPDQ-d5 ISTD (100 ug/L) and stored at -20 ℃ until instrument analysis.

2.2.6 Instrumental analysis

Physical properties of TWP were measured by a particle analyzer and a BET analyzer. The sizes of TWPs were measured (N=27 replicates) by a particle analyzer (Horiba) using refractive index at 1.5249.(McPherson and Cummings, n.d.) Water was used as chamber fluid and 0.5 g of TWP sample was mixed and measured; detailed measurement results are shown in **Figure S2.2A**.

The specific surface areas of the TWPs were measured by an Automated 3-station BET analyzer. Approximately 0.1 g samples were loaded into the pre-weighed BET cell, and the cell was weighed after the analysis with samples loaded to measure exact sample masses. Prior to BET analyses, samples were in-situ degassed for 12 hours at 60 ℃; nitrogen was used as analyzing gas. A typical measurement curve is shown in **Figure S2.2B**.

For 6PPDQ quantification, a liquid chromatography-tandem mass spectrometry system (Agilent 1290-6460 systems, described previously) was used in ESI+ mode.(Tian et al., 2021a) The analysis was conducted with an Agilent 1290 Infinity ultrahigh performance liquid

chromatography (UHPLC) system coupled to a triple-quadrupole mass spectrometer (Agilent G6460A) with electrospray ionization (ESI+ mode). The chromatographic separation of samples (10 μL) used a reversed-phase analytical column (Eclipse Plus C18, 2.1 mm × 100 mm, 1.8 μm particle size; Agilent, USA) with a Zorbax Eclipse XDB-C18 Guard Column (2.1 mm × 12.5 mm, 5 μm particle size; Agilent, USA) held at 25 °C. Binary mobile phases (0.2 mL/min flow rate) consisting of DI water (A) and methanol (B), both with 0.1% formic acid, were used. The gradient program was set as follows: started at 5% B, held at 5% B to 1 min, 50% B at 4 min, 100% B from 17-20 min, then 5% B from 20–24 min for re-equilibration. For the aqueous stability test sample analyses, the initial 5 minutes and the last 2 minutes of LC eluent were directed to waste to avoid high-concentration salt contamination of the system. Nitrogen was used as nebulizing and dessolvation gas (350 °C, 10 L/min) and capillary voltages were 3500 V. The 6PPDQ was analyzed in multiple reaction monitoring (MRM) mode using two individual ion transitions (299 → 187 and 299 → 215). The ion with higher peak area response was used for quantification (299 → 187) and the other ion for qualification and identity confirmation (maximum tolerance of ± 20% for quantifier-to-qualifier ion ratios). For quality assurance and quality control (QA/QC), samples were run in the following order: calibration curve samples were run at beginning, experimental replicates were analyzed sequentially with a methanol blank frequently run in the middle to check for carryover (no carryover observed), and a QC sample from the calibration curve set was reanalyzed at the end of the sequence (<5% variation of peak area was observed). The data was processed by Agilent MassHunter Quantitative Analysis; a calibration curve for 6PPDQ is shown in **Figure S2.3**. The collected data were processed through a Python-based platform with scipy package

(version 1.6.3) for leaching kinetics modeling. Confidence intervals of regressions were calculated from covariance matrices.(Cai et al., 2015)

2.3 Results and Discussion

2.3.1 Solubility and octanol–water partitioning

Prior to aqueous solubility experiments, 6PPDQ solubility in the stock solvent (ethanol) was tested in a preliminary study (>100 mg/L in ethanol at 20 °C). Solubility tests in water were conducted across a range of concentrations (**Table 2.1**) via a centrifugation method(OECD, 2019); resulting aqueous solutions were equilibrated >12 hrs at room temperature prior to centrifugation. For 6PPDQ solutions at higher nominal concentrations (>50 μg/L), light-pink solids accumulated within centrifuge tubes after centrifugation, which implied that excess solid precipitates existed within the solution. 6PPDQ solubility was estimated by comparison of samples before and after centrifugation. Notably, in contrast to the predicted solubility estimates derived from software (i.e., 2.4 – 140,000 mg/L from EPI-suite(Seung Lim J, 2016), Marvin("ChemAxon - Software Solutions and Services for Chemistry & Biology," n.d.) and SPARC(HILAL, S. H., n.d.)), the experimental results indicated a surprisingly low 6PPDQ solubility of 38±10 μg/L (**Figure 2.1**). This measurement aligned well with the solubility estimate of Hiki et al. (67±5 μg/L)(Hiki et al., 2021) and highlights the potential large inaccuracies of estimates obtained from software predictions. It is an open, and important, question as to why software predictions so diverged from observation for 6PPDQ.

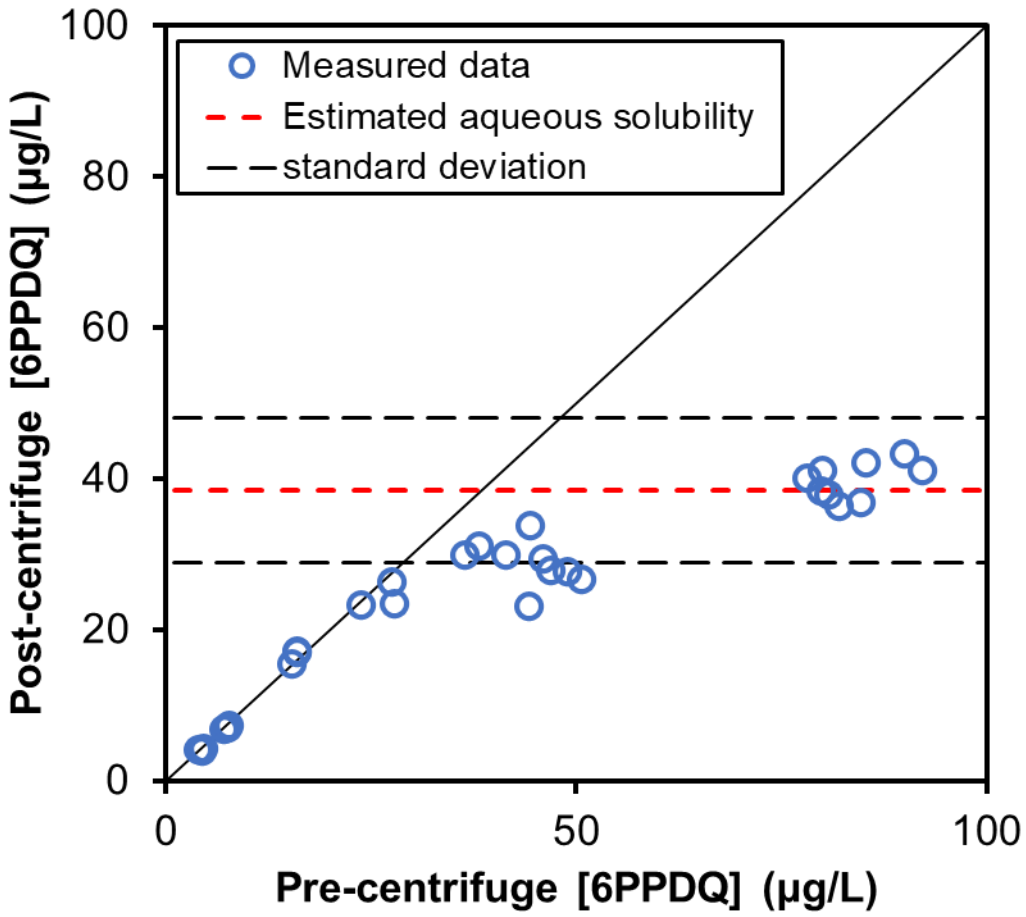


Figure 2.1. 6PPDQ aqueous solubility measurements after overnight (>12 hrs) mixing and sonication (30 min). Each datapoint corresponds to 6PPDQ measurements for an individual sample before and after centrifugation. The diagonal line represents a complete dissolution scenario (1:1 pre- versus post-centrifugation 6PPDQ concentrations). The red dashed line represents the average estimate of 6PPDQ aqueous solubility (38.4 µg/L). The black dashed lines represent one standard deviation above and below the estimated aqueous solubility.

Comparing measured versus nominal concentration of 6PPDQ yielded observations of 99±5% during preliminary screening studies conducted in various organic solvents (acetonitrile, ethyl acetate, and ethanol), indicating that no obvious chemical reactions interfered with solubility observations for 6PPDQ within organic solvents. Note that similar to methods reported by Hiki et al.,(Hiki et al., 2021) our preparation process involved drying of an initial ethanolic stock. We speculate that under the tested conditions, solute-solute intermolecular forces may be especially important for 6PPDQ solid dissolution, given that hydrogen bonding

between 6PPDQ and water should be fairly typical of solvent-solute interactions for similarly polar, similarly structured and sized compounds and serve to promote dissolution. Instead, solute-solute intermolecular forces in the 6PPDQ solid phase (e.g. **Figure S2.4**) may hinder dissolution and even contribute to strong "protective film" formation and stability, an outcome much emphasized in the industrial chemistry literature describing rubber protection by oxidation products of PPD antiozonants(Huntink, 2003, p. 6). Similarly, the strength of such solute-solute interactions would be expected to favor precipitation of 6PPDQ solids from the aqueous phase even at very low dissolved 6PPDQ concentrations, and could therefore also explain the extremely low observed aqueous solubility of 6PPDQ molecules present under equilibrium conditions in solution. This is supported by the negative linear relationship between the measured and the nominal concentration of the reconstituted 6PPDQ aqueous solutions (**Table 2.1**). One potential (albeit speculative) explanation for such low solubility could derive from a mechanism of complementary H-bonding and van der Waals interactions that promote stacked crystalline and other intermolecular interactions amongst molecules of 6PPDQ (and potentially other similar structured compounds) (Burke, 1984; Kuhn et al., 2010; Sieuw et al., 2019; Tuttle et al., 2021) (**Figure S2.4**), whether one is considering 6PPDQ dissolution from the solid phase or its precipitation from the aqueous phase.

Although aqueous solubility is low, 6PPDQ solubility easily exceeds exposure concentrations where coho salmon mortality occurs (LC_{50} = 95 ng/L) and aligns with observations of mortality and transport during storms.(Tian et al., 2021a, 2021b) Given limited solubility and reported 6PPDQ TWP concentrations,(Hu et al., 2022) there should exist other environmental transport pathways for 6PPDQ aside from dissolution and transport in water.

Most reports of 6PPDQ environmental occurrence derive from surface water or airborne particulate matter samples(Huang et al., 2021; Seiwert et al., 2022; Tian et al., 2021b, p.; Zhang et al., 2021), so further investigation of partitioning of 6PPDQ into other environmental medias (e.g., soil, sediments, DOMs) to fully elucidate environmental fate is merited. We also emphasize the importance of understanding solubility limitations in any experimental designs involving 6PPDQ aqueous solutions (e.g., toxicity assessment, photolysis studies) and note that co-solvent effects might be particularly important for higher concentrations of 6PPDQ or behavior in complex solutions.

The logK_{OW} of 6PPDQ was measured at 4.30±0.02, slightly lower than the reported 4.84 for the parent 6PPD.(OSPAR Commission, 2006c) The measured results were similar to the numbers discussed by Cao et al (3.98 for 6PPDQ and 4.47 for 6PPD).(Cao et al., 2022) This "moderate-high" logK_{OW} value indicated that some fraction of 6PPDQ mass would be expected to partition to soil, sediments, rubber or plastic materials in aqueous systems. Notably, chemicals with logK_{OW} values within the range of 3-7 have optimal uptake efficiency through gill tissues,(McKim et al., 1985) an important bioavailability consideration for the highly toxic 6PPDQ, as chemicals with low logK_{OW} values tend to be insoluble in lipids, while high logK_{OW} chemicals bind to lipid membranes and prevent further uptake.(McKim et al., 1985) Further investigation of the bioavailability and uptake of 6PPDQ and other tire rubber-derived roadway contaminants to fish, especially within the context of relative toxicity outcomes,(Hiki et al., 2021) is merited.

2.3.2 Sorption to materials

To assess potential sorption losses during sampling and laboratory processes, sorption to 12 materials was evaluated, including 10 common labware materials (ST-PE, ST-PTFE, ST-Sil, SPE-PTFE, PF, CT-PP, CT-FEP, CT-G, RS, and SS-PP) and 2 conventional materials (BTR and SS). The rubber materials (RS and BTR) were analyzed and confirmed 6PPDQ-free before the experiment; no 6PPDQ (<0.025 ug/L) was detected. For short (5 min) exposures that simulated contact times for transfer between containers or rapid contact with sampling equipment or tubing materials, most materials showed low to moderate sorption (>75% residual 6PPDQ concentrations detected; **Figure 2.2**). Unsurprisingly, for longer 30-minute contact periods, which simulated temporary storage or sample processing, more extensive sorptive losses occurred. Glass, stainless steel and plastics (PTFE, FEP, PE and PP) had zero to moderate sorption tendency (>60% residual 6PPDQ concentrations vs t_0), while rubber (BTR: 35±8.6%; RS: 8.7±3.2%) and silicone (ST-Sil 25±0.1%) promoted substantial concentration reductions. With longer (24 h) contact, even more loss to rubber (BTR: 0.97±0.13%; RS: 0.7±0.21%) and silicone (ST-Sil: 0.34±0.22%) occurred.

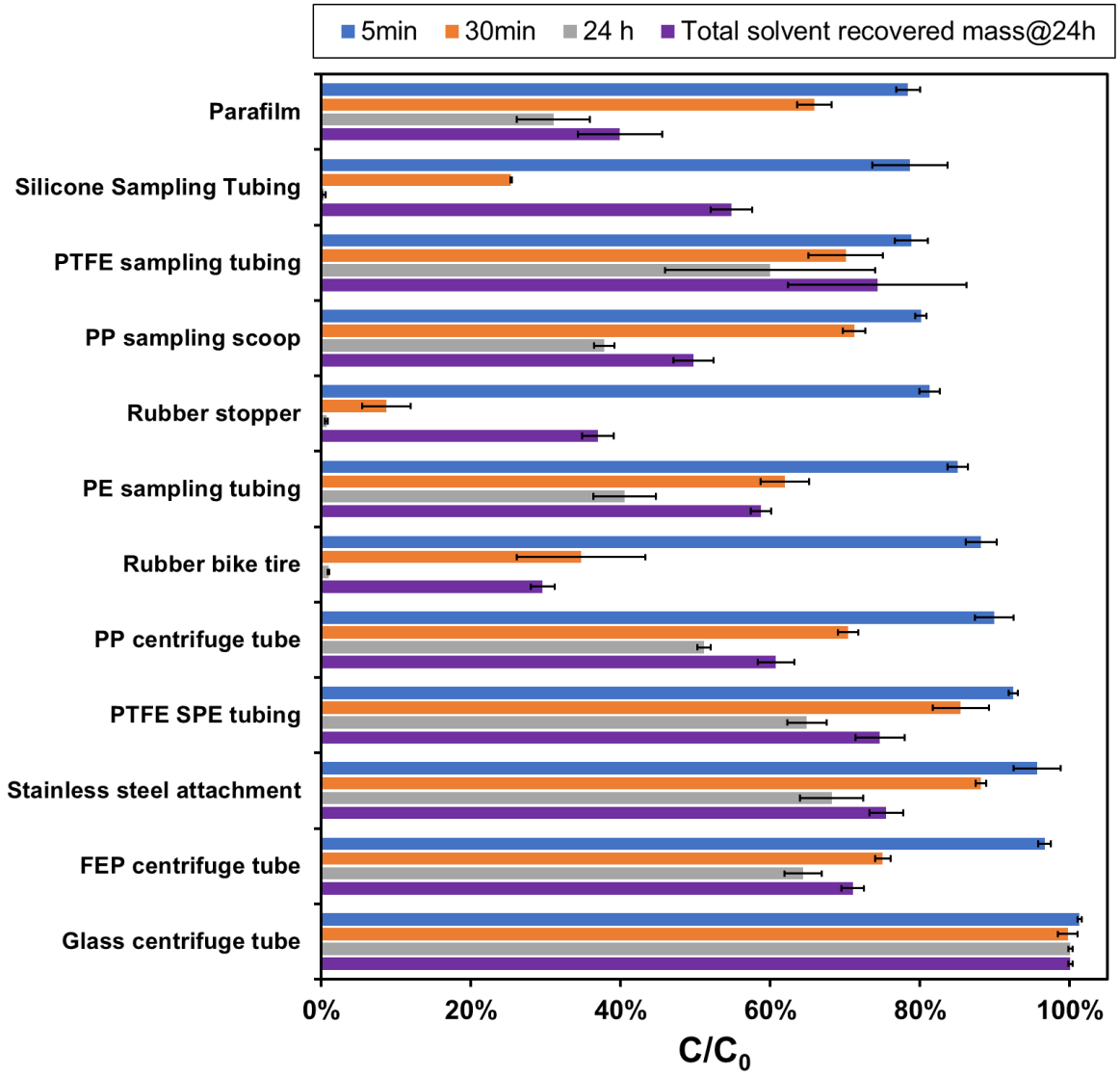


Figure 2.2. Sorption losses of 6PPDQ to different materials over 24 hours. Relative concentrations (C/C_o) were derived by comparing residual aqueous 6PPDQ concentrations to initial 6PPDQ concentrations in the aqueous stock (5 µg/L) used for the tests. Error bars correspond to standard deviations about the mean from three replicate samples.

Similar to outcomes for 6PPD(Huntink, 2003; OSPAR Commission, 2006c), the hydrophobicity ($\log K_{OW}$ = 4.3; aqueous solubility = 38±10 µg/L) of 6PPDQ should promote high sorption losses of 6PPDQ from aqueous solution to rubber, silicone, and plastic materials. For instance, although it is highly variable and depends upon material type and sizes, the surface area of rubbers are typically high (e.g., the surface area of crumb rubbers are measured

at hundreds to thousands of cm^2/g(Fan et al., 2021; Shen et al., 2009; Zhang et al., 2019)). The high specific surface area indicates a porous structure which can retain inaccessible or kinetically limited 6PPDQ mass during equilibration(Angelin et al., 2017; Kasner and Meinecke, 1996; Rostami-Tapeh-Esmaeil et al., 2021) compared to smooth surface materials (e.g., glass). Stainless steel and plastics designed to be more chemically inert (e.g., FEP, PTFE) showed moderate sorption potentials, with higher residual 6PPDQ concentrations after 24-hour contact. Substantial 6PPDQ mass was sorbed by PP (residual 6PPDQ mass; CT-PP: 51±1.4%; SS-PP: 38±1.5%) and parafilm (residual 6PPDQ mass; PF: 31±2.3%). Note that for all contact times, no significant 6PPDQ sorption (t-test; p-value=0.59, 0.32 and 0.12 at 5 min, 30 min and 24 hr vs t_0, respectively) occurred for glass (101±0.3%, 100±1.3% and 100±0.3% residual concentrations at 5 min, 30 min, and 24 hr vs t_0, respectively), suggesting glass as the preferred material for labware and processing.

After 24 h contact periods, aqueous phases were removed and all containers and materials were methanol extracted (30 min sonication) to evaluate recovery of sorbed mass. Mass balances indicated that all materials except glass retained ~25-70% of initial 6PPDQ mass that was not recoverable with methanol (**Figure S2.5**). The rubber materials (6PPDQ total mass loss: BTR: 70±1.6%; RS: 63±2.1%) exhibited substantially higher mass loss than other materials, indicating 6PPDQ has higher and stronger sorption potentials to these materials. The methanol non-recoverable mass loss may result from strong hydrophobic interactions, material microporosities, slow mass transfer, or chemical reactions. During the contact period and subsequent methanol extraction, rubber and silicone materials were especially prone to initial mass loss (>95%) from aqueous phases over 24 h equilibration, but substantial proportions of

the sorbed 6PPDQ were recoverable (30 min methanol extraction with respect to total 6PPDQ mass; BTR: 29±1.5%; RS: 36±1.9%; ST-Sil: 54±2.6%) with methanol, indicating most of the total mass was partitioned onto/into the materials. Seiwert et al.(Seiwert et al., 2022) recently reported that most 6PPDQ mass remained in TWP particles (distribution ratio or water to particle = 0.001) after environmental deposition, consistent with the 6PPDQ sorption data of rubber materials. In contrast, plastic materials were less sorptive during the 24 h tests (~30-60% 6PPDQ mass loss during 24 h equilibration) while much less sorbed 6PPDQ mass was methanol recoverable (30 min methanol extraction: CT-FEP: 6.6±0.99%; SS: 7.3±2.5%; SPE-PTFE: 9.7±0.64%; CT-PP: 9.6±1.6%; ST-PE: 18±5.4%; SS-PP: 12±1.5%; ST-PTFE: 14±2.1%; PF: 8.9±0.85%). These plastic materials also showed a high total recoverable mass (24 h contact solution mass + methanol recoverable mass; ~40-75%), suggesting 6PPDQ has less tendency to sorb to or react with them. This comparison indicates that rubber and silicone materials have much higher sorption capacities, likely resulting from a combination of more porosity and more surface area, in addition to hydrophobic and electrostatic interactions between 6PPDQ and material surfaces, or compounds present on the surfaces of these materials, that affect sorbent capacities.(Joo et al., 2021)

These results provide a general guideline for material selection for aqueous phase samples containing 6PPDQ in lab studies and field sampling events. For short-term contact events, glass, stainless steel and PTFE labware and containers are preferred options. For long-term storage or sample transport, glass would be the best material to handle 6PPDQ aqueous solutions in the lab operations and sampling processes, with solvent rinsing becoming especially important for other materials. We also note a consistent inability (at least by us) to fully recover all spiked

6PPDQ mass in these example systems with various types of materials (e.g., all the tested materials except glass; including plastic, rubber, silicone and stainless steel) following preliminary extraction (30 min methanol sonication extraction). The potential for strong surface interactions or chemical instability, leading to low bias in dissolved concentrations, seems to exist for some experimental conditions and remains a data gap. The results also indicate that suspended particles in the environment, especially various microplastic particles (e.g., plastic debris or waste) may sorb, accumulate, and transport 6PPDQ in the environment.(Müller et al., 2022)

2.3.3 Leaching dynamics

Physical properties of the TWP used here were measured via BET, yielding a specific surface area of 5.88±0.66 m^2/g, median particle diameter (d_{50}) of 232 μm and a uniformity coefficient (d_{60}/d_{10}) of 3.21 (**Figure S2.2**). Note that these measurements are substantially dependent on the tire material, TWP preparation method, TWP shape, size and morphology, and therefore might be subject to large variance for different tested TWP. Nevertheless, Fan et al reported a 0.3-16.65 m^2/g BET-specific surface area measured on TWP during UV aging processes, aligned with current measurements(Fan et al., 2021). The BET-specific surface area indicated that the tire rubber materials were porous,(Milani et al., 2004) and measured particle sizes of this laboratory generated were slightly larger than reports for tire-road wear particles (TRWP; 20-100 μm for most particles) that were agglomerations of tire rubber particles and roadway materials.(Klöckner et al., 2021b; Wagner et al., 2018) The 6 h flow-through leaching

system employed here simulated short-term leaching of TWP; for example, the timescale of a rainstorm event on a roadway surface. Over the 6 hours of leaching 250 mg of TWP, 6PPDQ leachate concentrations decreased 50% (6PPDQ concentration in collected leachates: 125±20 ng/L at 15 min decreasing to 60±9 ng/L at 6 hr, **Figure 2.3A**). Given the complex surface characteristics of the TWP and high flow rate, one dimensional leaching dynamics were assumed and the data was fit to an empirical first-order kinetic model with the following assumptions: a) 6PPDQ dissolution is the dominant surface reaction during leaching; b) the column is an ideal, well mixed reactor with no dispersion; and c) TWP surface-associated 6PPDQ instantaneously equilibrates with the leachate.

Based on these assumptions, leachate concentrations would be calculated:

$$V\frac{dC_{leachate}}{dt} = k_{diss}C_{TWP}A_{TWP} - QC_{leachate} \qquad \text{(Eq. 1)}$$

where V is the reactor volume (mL), Q is leachate flow rate (mL/min), $C_{leachate}$ is the leachate 6PPDQ concentration (ng/mL; assuming column concentration equals outlet concentration), t is the leaching time (min), C_{TWP} is the surface area normalized 6PPDQ concentration (ng/cm²) and A_{TWP} is the surface area of TWP (cm^2); k_{diss} is the empirical first-order kinetic rate constant (min^{-1}). Based on eqn (1) and observed 6PPDQ leachate dynamics yields:

$$M_{leachate} = M_{TWP} \times (1 - exp(-k_{diss}t)) \qquad \text{(Eq. 2)}$$

where M is surface area normalized mass of 6PPDQ (ng/cm^2), $M_{leachate}$ and M_{TWP} are the cumulative mass leached and total surficial mass available on TWP for leaching (converted from 6PPDQ leachate concentration using BET surface area); k_{diss} is the empirical first-order kinetic rate constant (min^{-1}) and t is leaching time (min).

A regression of the cumulative 6PPDQ mass data was fitted to Equation 2 and the confidence interval of the coefficients was calculated from the covariance matrix.(Cai et al., 2015) The measured cumulative leached mass was 0.086±0.005 ng/cm^2 6PPDQ (equivalent to 5.15±0.3 ug 6PPDQ per g TWP) over 6 hours of leaching.(Monaghan et al., 2021) There was good agreement between observed and modeled data, with no statistically significant differences between them (t-test; p-value>0.05; **Figure 2.3B**), indicating the leaching process under tested conditions was dominated by kinetics of 6PPDQ dissolution from the TWP surface. The 6PPDQ leaching dynamics under the studied conditions can be therefore described by Equation 3:

$$M_{Cumulative} = 0.135 \times (1 - \exp(-0.0027t)) \qquad \text{(Eq. 3)}$$

The predicted first-order dissolution rate constant is 0.0027±0.0005 min^{-1}. Based on the regression model and assuming unlimited time for leaching, the total surficial 6PPDQ mass available for leaching is predicted at 0.135±0.021 ng/cm^2 (7.14±1.3 μg 6PPDQ per gram TWP). In comparison, solvent extraction with methanol of this TWP yielded 15.6±1.3 μg 6PPDQ per gram TWP, which translates to 0.295±0.025 ng/cm^2. The solvent extractable 6PPDQ measured here aligned well with reports of Masset et al. (14 μg 6PPDQ per gram cryogenically milled tire tread)(Masset et al., 2022). These data indicate ~46% of the total surficial 6PPDQ mass was water leachable over 6 hours. Note here that these estimates do not consider formation of additional 6PPDQ upon continued 6PPD reaction with ozone or long-term TMP dynamics. Therefore, the total leachable 6PPDQ of TWP over its service life would be much higher than above estimates considering the substantial amount of 6PPD used in tires and expectations for

slow diffusion rates (1.39×10^{-8} - 8.56×10^{-8} cm^2/s) reported for 6PPD through rubber phases.(Huntink, 2003) Such processes will ultimately regenerate 6PPDQ via oxidative reactions on tire rubber surfaces to act as a continuing environmental source of 6PPDQ.(OSPAR Commission, 2006c)

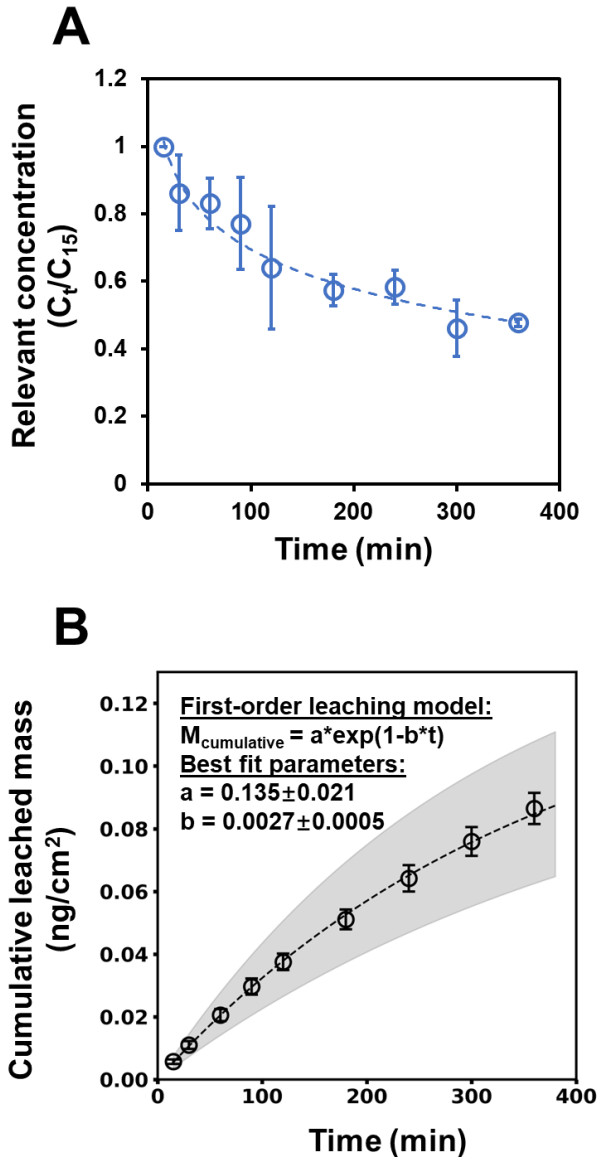


Figure 2.3. 6PPDQ aqueous leaching kinetics from TWPs over 6 hours in a flow through column system (flow rate = 45 mL/min). (A) 6PPDQ aqueous concentration vs time, normalized to the observed "initial" concentration measured at 15 minutes. (B) Cumulative surface area normalized mass of 6PPDQ leached into aqueous solution over time. Error bars represent standard deviation about the mean from three replicate samples; dashed lines in both figures represent regression fit of the data points; shaded boundaries represent 95% confidence intervals.

6PPDQ concentrations in tire leachate over time fit first-order kinetics (k_{diss}=0.0027±0.0005 min^{-1}) without any obvious lag period. Similar rapid leaching dynamics seem likely for storm events where 6PPDQ dissolution into runoff happens quickly (<15min) and concentrations are maintained above effective toxic levels throughout the storm event (a few hours up to a day) by continued leaching. Complementing roadway runoff sources, TWP may themselves transport through the environment and act as a diffuse and continuing source of contaminants.

2.3.4 Aqueous stability and pH

After 6PPDQ is leached, it is expected to transport into surface waters via roadway runoff and presumably minor pathways such as deposition of atmospheric particulates. The stability of 6PPDQ in aqueous solutions such as roadway runoff and surface waters is consequently important to the potential for adverse environmental effects.(Tian et al., 2021b) Here, noting the rapid, pH-dependent transformation in aqueous solution reported for the parent 6PPD(OSPAR Commission, 2006c), the aqueous stability of 6PPDQ in deionized water was tested over 47 days at pH 5, 7 and 9 at 20-22 °C (**Figure 2.4**).

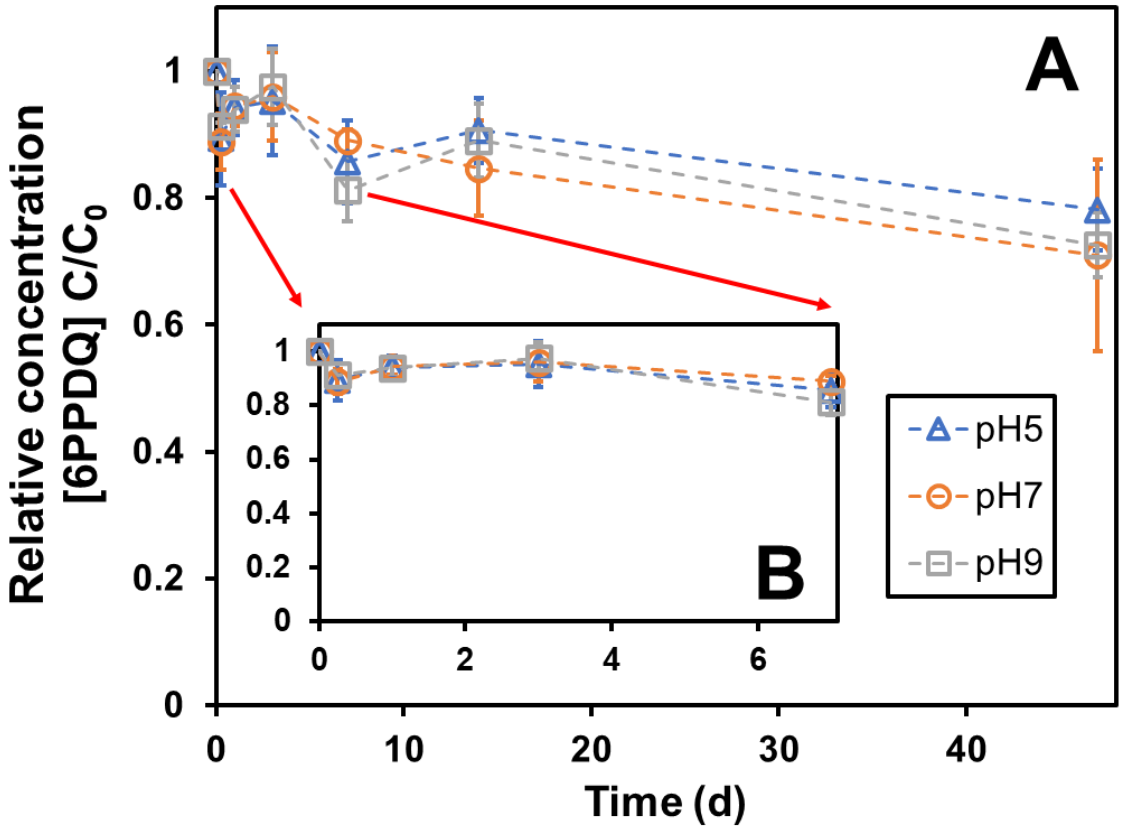


Figure 2.4. 6PPDQ stability in 10 mM aqueous phosphate buffer at pH 5, 7 and 9. (A) Long-term 6PPDQ stability over 47 d (t-test t_0 vs. t_{47}: p-value=0.023, 0.045 and 0.001 for pH 5, 7 and 9, respectively), and (B) short-term 6PPDQ stability over 7 d (t-test t_0 vs. t_7: p-value=0.097, 0.002 and 0.005 for pH 5, 7 and 9, respectively). Error bars represent standard deviations about the mean from triplicate samples.

Over 3 days, no significant loss was observed for any pH condition (two-tailed t-test; *p*-value = 0.84 for pH 5, 0.19 for pH 7 and 0.31 for pH 9). At 14 days, 6PPDQ was generally stable, with no obvious change in concentration in pH 5 and pH 9 (p-value = 0.11 for pH 5; 0.06 for pH 9). The pH 7 group significantly decreased (p-value = 0.03) with 15% concentration reduction. At 47 days, 6PPDQ concentrations exhibited a 26±10% decrease on average (p-values >0.05 for all groups). In general, 6PPDQ appears to be stable in simple aqueous systems, although Hiki et al., reported a much shorter 33 hr half-life for 6PPDQ at

23 °C in dechlorinated tap water.(Hiki et al., 2021) Fohet et al.(Fohet et al., 2023) also reported concentration evolutions of 6PPDQ in tire and road wear particles under different conditions (photoaging τ_{photo}=9.6days, natural aging $\tau_{outdoor}$=9.6days and no thermoaging decrease were observed), suggesting moderate environmental stability of 6PPDQ. Obviously, as solution complexity increases (e.g., co-contaminants, natural organic matter, microbiota, redox active constituents and solid phases), the potential for reactive loss would increase substantially. Solution factors promoting instability remain important research needs for 6PPDQ and other similar roadway contaminants.

2.3.5 6PPDQ roadway mass balances

Kole et al. (Kole et al., 2017) estimated 1,250,000-1,800,000 metric tons TRWP/year release in the U.S. Using the mass distribution of TRWP of Wagner et al. (Wagner et al., 2018), 45-77% of total TRWP will be retained on the roadside. If roadway residual TRWP consists of 50% tire rubber, 280,000-690,000 metric tons TWP/year is deposited to the road and available for localized leaching. Using measurements from this study, 7.14±1.3 (by aqueous leaching model) to 15.6±1.3 (by solvent extraction) μg of surficial 6PPDQ is available per gram TWP, yielding a minimum of 2000 to 10800 kg surficial 6PPDQ/year. According to the 2020 Department of Transportation statistics ("Highway Statistics 2020 - Policy | Federal Highway Administration," n.d.), total road lane length of the U.S highways is 14,200,000 km; with lane width of 3.7 meters (12 feet), the estimated total highway area for the US is 52,600 km^2. At average precipitation rate of 769 mm/year ("Annual 2020 National Climate Report | National

Centers for Environmental Information (NCEI)," n.d.) and assuming a runoff coefficient of 0.85,("Stormwater Design Manual | NC DEQ," n.d.) we estimate an average roadway runoff volume of 3.44×10^{13} L/year on U.S. roadways. Using these estimates, average 6PPDQ concentrations (60-310 ng/L) in roadway runoff would be calculated as follows:

Low-emission scenario:

Annual total 6PPDQ available in mass:

$$1250000\frac{\text{metric ton TWP}}{\text{year}} \times 45\% \times 50\% \times 7.14\frac{\mu\text{g 6PPDQ}}{\text{g TWP}} = 2008\ \frac{\text{kg 6PPDQ}}{\text{year}}$$

High-emission scenario:

Annual total 6PPDQ available in mass:

$$1800000\frac{\text{metric ton TWP}}{\text{year}} \times 77\% \times 15.6\frac{\mu\text{g 6PPDQ}}{\text{g TWP}} = 10811\ \frac{\text{kg 6PPDQ}}{\text{year}}$$

The total highway lane length is 14,200,000 km; with lane width of 3.7 m,("Highway Statistics 2020 - Policy | Federal Highway Administration," n.d.) the total highway area is:

$$14200000 \times 3.7 = 52592\ km^2 = 5.25 \times 10^{10} m^2$$

The annual average roadway runoff can be predicted as:

$$0.769(m/year) \times 0.85 \times 5.25 \times 10^{10} m^2 = 3.44 \times 10^{10}(m^3/year)$$

$$= 3.43 \times 10^{13}(L/year)$$

Consequently, 6PPDQ concentration in roadway runoffs can be predicted for the low-emission scenario as:

$$2008\ \frac{\text{kg 6PPDQ}}{\text{year}} \div \left(3.44 \times 10^{13}\frac{\text{L}}{\text{year}}\right) = 58.4\frac{ng}{L}$$

or for the high-emission scenario as:

$$10811\ \frac{\text{kg 6PPDQ}}{\text{year}} \div (3.44 \times 10^{13}\ \frac{\text{L}}{\text{year}}) = 314.3\ \frac{ng}{L}$$

We note that the above calculation only considers TRWP as a source of TWP-derived contaminants; contributions from whole tire surfaces on vehicles and other sources would increase these values. We also note this estimate does not consider regeneration of 6PPDQ (as 6PPD continues to react with ozone), does not consider possible differences in surface area available for leaching mass across tires, TWP, and TRWP, does not account for non-roadway TRWP, and does not consider any attenuation. The prediction is of the same magnitude as our previous observations.(Tian et al., 2021a) Although such models and assumptions still require substantial further refinement and data collection to address the many data gaps existing for the 6PPDQ environmental mass balance, it is clear that risk thresholds to sensitive species are easily attainable, especially for areas with substantial traffic and roadway runoff impact.(Brinkmann et al., 2022; Feist et al., 2017; Hiki et al., 2021; Tian et al., 2021a)

2.4 Environmental implications

This study investigated basic chemical properties (logK_{OW} and solubility), leaching, and aqueous stabilities of 6PPDQ. Agreeing reasonably well with software predictions, we observed a moderately high logK_{OW} (4.30±0.02) for 6PPDQ, but far lower aqueous solubility (38.4 μg/L) than expected by software prediction. Nevertheless, the aqueous solubility of 6PPDQ easily exceeds the reported LC_{50} value for coho salmon(Tian et al., 2021a). Stability

experiments suggested relatively high stability of 6PPDQ over 14 d but some (26±10%) loss over 47 d in simple deionized water systems.

6PPDQ sorption studies provided a guideline for materials selection during sampling and sample processing. Specifically, glass is recommended for lab operations and storage if possible. For short-term contact, such as transport or field sampling, chemically inert plastics (PTFE and FEP) are likely acceptable as tubing materials or containers (<10% sorption observed during 5 min contact time), although solvent rinsing for recovery of sorbed 6PPDQ may be needed as contact times grow. Extra caution may be needed for rubbers and silicon materials in contact with 6PPDQ (e.g., rubber stoppers, silicone fittings), as these materials can strongly sorb 6PPDQ from aqueous solution (up to 92% 6PPDQ sorption during 30 min) with substantial unrecoverable mass, even using standard methanol rinsing procedures.

During 6 hr leaching tests, a cumulative leached mass of 0.086±0.005 ng/cm^2 6PPDQ (5.15±0.3 ug 6PPDQ per g TWP; k_{diss}=0.0027±0.0005 min^{-1}) was observed, indicating expected reasonably rapid discharge of 6PPDQ into roadway runoff during storms. Extending these data to estimates of TRWP on roadways yielded predictions of 60-310 ng/L of TRWP-derived 6PPDQ in roadway runoff, although there currently exist very few confident and well-validated estimates of TWP, TRWP and tire rubber phases on roadways and none reflecting whole tire sources to the best of our knowledge.(Klöckner et al., 2021b; Kole et al., 2017; Unice et al., 2015)

While the measured attributes above are helpful, there still exist many substantial data gaps around characteristics, environmental fate and transport of 6PPDQ, including the partitioning

behavior of 6PPDQ into sediments and soil, biological availability of 6PPDQ to aquatic organisms and human exposure risks of 6PPDQ. Further investigations are merited to understand the roles of dissolved organic matter and other solution components in modulating 6PPDQ dissolution and partitioning behavior as examined in current studies. Investigation of TWP surface chemistry dynamics during leaching, related sorption processes, and 6PPDQ transformation mechanisms in more complex aqueous solutions (with natural organic matter, co-constituents, and biota) will also provide critical information needed to evaluate 6PPDQ environmental fate and risks. These data will all help to guide field measurements and improved characterization of TWP, roadway runoff, and receiving waters to understand locations and time periods of maximum discharge that will enable the short-term management of environmental risks arising from the occurrence of highly toxic, and ubiquitous, 6PPDQ in roadway runoff.

References

Ahamad, A., Madhav, S., Singh, A.K., Kumar, A., Singh, P., 2020. Types of Water Pollutants: Conventional and Emerging, in: Pooja, D., Kumar, P., Singh, P., Patil, S. (Eds.), Sensors in Water Pollutants Monitoring: Role of Material, Advanced Functional Materials and Sensors. Springer, Singapore, pp. 21–41. https://doi.org/10.1007/978-981-15-0671-0_3

Alhelou, R., Seiwert, B., Reemtsma, T., 2019. Hexamethoxymethylmelamine – A precursor of persistent and mobile contaminants in municipal wastewater and the water cycle. Water Res. 165, 114973. https://doi.org/10.1016/j.watres.2019.114973

Angelin, A.F., Lintz, R.C.C., Gachet-Barbosa, L.A., Osório, W.R., 2017. The effects of porosity on mechanical behavior and water absorption of an environmentally friendly cement mortar with recycled rubber. Constr. Build. Mater. 151, 534–545. https://doi.org/10.1016/j.conbuildmat.2017.06.061

Annual 2020 National Climate Report | National Centers for Environmental Information (NCEI) [WWW Document], n.d. URL https://www.ncdc.noaa.gov/sotc/national/202013 (accessed 12.28.22).

Braukmann, U., Böhme, D., 2011. Salt pollution of the middle and lower sections of the river Werra (Germany) and its impact on benthic macroinvertebrates. Limnologica, Salinisation of running waters 41, 113–124. https://doi.org/10.1016/j.limno.2010.09.003

Brinkmann, M., Montgomery, D., Selinger, S., Miller, J.G.P., Stock, E., Alcaraz, A.J., Challis, J.K., Weber, L., Janz, D., Hecker, M., Wiseman, S., 2022. Acute Toxicity of the Tire Rubber-Derived Chemical 6PPD-quinone to Four Fishes of Commercial, Cultural, and Ecological Importance. Environ. Sci. Technol. Lett. 9, 333–338. https://doi.org/10.1021/acs.estlett.2c00050

Burke, J., 1984. Solubility Parameters: Theory and Application [WWW Document]. URL https://cool.culturalheritage.org/coolaic/sg/bpg/annual/v03/bp03-04.html (accessed 11.22.22).

Cai, T.T., Liang, T., Zhou, H.H., 2015. Law of log determinant of sample covariance matrix and optimal estimation of differential entropy for high-dimensional Gaussian distributions. J. Multivar. Anal. 137, 161–172. https://doi.org/10.1016/j.jmva.2015.02.003

Cao, G., Wang, W., Zhang, J., Wu, P., Zhao, X., Yang, Z., Hu, D., Cai, Z., 2022. New Evidence of Rubber-Derived Quinones in Water, Air, and Soil. Environ. Sci. Technol. 56, 4142–4150. https://doi.org/10.1021/acs.est.1c07376

ChemAxon - Software Solutions and Services for Chemistry & Biology [WWW Document], n.d. URL https://chemaxon.com/ (accessed 8.25.21).

Corsi, S.R., Graczyk, D.J., Geis, S.W., Booth, N.L., Richards, K.D., 2010. A Fresh Look at Road Salt: Aquatic Toxicity and Water-Quality Impacts on Local, Regional, and National Scales. Environ. Sci. Technol. 44, 7376–7382. https://doi.org/10.1021/es101333u

Councell, T.B., Duckenfield, K.U., Landa, E.R., Callender, E., 2004. Tire-Wear Particles as a Source of Zinc to the Environment. Environ. Sci. Technol. 38, 4206–4214. https://doi.org/10.1021/es034631f

Drzyzga, O., 2003. Diphenylamine and derivatives in the environment: a review. Chemosphere 53, 809–818. https://doi.org/10.1016/S0045-6535(03)00613-1

Du, B., Liang, B., Li, Y., Shen, M., Liu, L.-Y., Zeng, L., 2022. First Report on the Occurrence of N-(1,3-Dimethylbutyl)-N′-phenyl-p-phenylenediamine (6PPD) and 6PPD-Quinone as Pervasive Pollutants in Human Urine from South China. Environ. Sci. Technol. Lett. https://doi.org/10.1021/acs.estlett.2c00821

Fan, X., Ma, Z., Zou, Y., Liu, J., Hou, J., 2021. Investigation on the adsorption and desorption behaviors of heavy metals by tire wear particles with or without UV ageing processes. Environ. Res. 195, 110858. https://doi.org/10.1016/j.envres.2021.110858

Feist, B.E., Buhle, E.R., Baldwin, D.H., Spromberg, J.A., Damm, S.E., Davis, J.W., Scholz, N.L., 2017. Roads to ruin: conservation threats to a sentinel species across an urban gradient. Ecol. Appl. 27, 2382–2396. https://doi.org/10.1002/eap.1615

Fohet, L., Andanson, J.-M., Charbouillot, T., Malosse, L., Leremboure, M., Delor-Jestin, F., Verney, V., 2023. Time-concentration profiles of tire particle additives and transformation products under natural and artificial aging. Sci. Total Environ. 859, 160150. https://doi.org/10.1016/j.scitotenv.2022.160150

Highway Statistics 2020 - Policy | Federal Highway Administration [WWW Document], n.d. URL https://www.fhwa.dot.gov/policyinformation/statistics/2020/ (accessed 12.14.22).

Hiki, K., Asahina, K., Kato, K., Yamagishi, T., Omagari, R., Iwasaki, Y., Watanabe, H., Yamamoto, H., 2021. Acute Toxicity of a Tire Rubber-Derived Chemical, 6PPD Quinone, to Freshwater Fish and Crustacean Species. Environ. Sci. Technol. Lett. https://doi.org/10.1021/acs.estlett.1c00453

HILAL, S. H., n.d. Estimation of Physical Properties and Chemical Reactivity Parameters of Organic Compounds for Environmental Modeling by SPARC [WWW Document]. URL https://cfpub.epa.gov/si/si_public_record_report.cfm?Lab=NERL&dirEntryId=203108 (accessed 2.26.22).

Hou, F., Tian, Z., Peter, K.T., Wu, C., Gipe, A.D., Zhao, H., Alegria, E.A., Liu, F., Kolodziej, E.P., 2019. Quantification of organic contaminants in urban stormwater by isotope dilution and liquid chromatography-tandem mass spectrometry. Anal. Bioanal. Chem. 411, 7791–7806. https://doi.org/10.1007/s00216-019-02177-3

Hu, X., Zhao, H.N., Tian, Z., Peter, K.T., Dodd, M.C., Kolodziej, E.P., 2022. Transformation Product Formation upon Heterogeneous Ozonation of the Tire Rubber Antioxidant 6PPD (N-(1,3-dimethylbutyl)-N′-phenyl-p-phenylenediamine). Environ. Sci. Technol. Lett. 9, 413–419. https://doi.org/10.1021/acs.estlett.2c00187

Huang, W., Shi, Y., Huang, J., Deng, C., Tang, S., Liu, X., Chen, D., 2021. Occurrence of Substituted p-Phenylenediamine Antioxidants in Dusts. Environ. Sci. Technol. Lett. https://doi.org/10.1021/acs.estlett.1c00148

Huntink, N.M., 2003. Durability of rubber products: Development of new antidegradants for long-term protection.

Joo, S.H., Liang, Y., Kim, M., Byun, J., Choi, H., 2021. Microplastics with adsorbed contaminants: Mechanisms and Treatment. Environ. Chall. 3, 100042. https://doi.org/10.1016/j.envc.2021.100042

Kasner, A.I., Meinecke, E.A., 1996. Porosity in Rubber, a Review. Rubber Chem. Technol. 69, 424–443. https://doi.org/10.5254/1.3538379

Klöckner, P., Seiwert, B., Wagner, S., Reemtsma, T., 2021a. Organic Markers of Tire and Road Wear Particles in Sediments and Soils: Transformation Products of Major Antiozonants as Promising Candidates. Environ. Sci. Technol. https://doi.org/10.1021/acs.est.1c02723

Klöckner, P., Seiwert, B., Weyrauch, S., Escher, B.I., Reemtsma, T., Wagner, S., 2021b. Comprehensive characterization of tire and road wear particles in highway tunnel road dust by use of size and density fractionation. Chemosphere 279, 130530. https://doi.org/10.1016/j.chemosphere.2021.130530

Kole, P.J., Löhr, A.J., Van Belleghem, F.G.A.J., Ragas, A.M.J., 2017. Wear and Tear of Tyres: A Stealthy Source of Microplastics in the Environment. Int. J. Environ. Res. Public. Health 14. https://doi.org/10.3390/ijerph14101265

Kroiss, H., Rechberger, H., Egle, L., 2011. Phosphorus in Water Quality and Waste Management, Integrated Waste Management - Volume II. IntechOpen. https://doi.org/10.5772/18482

Kuhn, B., Mohr, P., Stahl, M., 2010. Intramolecular Hydrogen Bonding in Medicinal Chemistry. J. Med. Chem. 53, 2601–2611. https://doi.org/10.1021/jm100087s

Kumata, H., Yamada, J., Masuda, K., Takada, H., Sato, Y., Sakurai, T., Fujiwara, K., 2002. Benzothiazolamines as Tire-Derived Molecular Markers: Sorptive Behavior in Street Runoff and Application to Source Apportioning. Environ. Sci. Technol. 36, 702–708. https://doi.org/10.1021/es0155229

Masset, T., Ferrari, B.J.D., Dudefoi, W., Schirmer, K., Bergmann, A., Vermeirssen, E., Grandjean, D., Harris, L.C., Breider, F., 2022. Bioaccessibility of Organic Compounds Associated with Tire Particles Using a Fish In Vitro Digestive Model: Solubilization Kinetics and Effects of Food Coingestion. Environ. Sci. Technol. 56, 15607–15616. https://doi.org/10.1021/acs.est.2c04291

McKim, J., Schmieder, P., Veith, G., 1985. Absorption dynamics of organic chemical transport across trout gills as related to octanol-water partition coefficient. Toxicol. Appl. Pharmacol. 77, 1–10. https://doi.org/10.1016/0041-008X(85)90262-5

McPherson, A.T., Cummings, A.D., n.d. Refractive index of rubber 10.

Milani, M., Pucillo, F.P., Ballerini, M., Camatini, M., Gualtieri, M., Martino, S., 2004. First evidence of tyre debris characterization at the nanoscale by focused ion beam. Mater. Charact. 52, 283–288. https://doi.org/10.1016/j.matchar.2004.06.001

Monaghan, J., Jaeger, A., Agua, A.R., Stanton, R.S., Pirrung, M., Gill, C.G., Krogh, E.T., 2021. A Direct Mass Spectrometry Method for the Rapid Analysis of Ubiquitous Tire-Derived Toxin N-(1,3-Dimethylbutyl)-N′-phenyl-p-phenylenediamine Quinone (6-

PPDQ). Environ. Sci. Technol. Lett. 8, 1051–1056. https://doi.org/10.1021/acs.estlett.1c00794

Müller, A., Kocher, B., Altmann, K., Braun, U., 2022. Determination of tire wear markers in soil samples and their distribution in a roadside soil. Chemosphere 294, 133653. https://doi.org/10.1016/j.chemosphere.2022.133653

OECD, 2019. Guidance Document on Aquatic Toxicity Testing of Difficult Substances and Mixtures. Organisation for Economic Co-operation and Development, Paris.

OSPAR Commission, 2006c. OSPAR background document on 4-(dimethylbutylamino)diphenylamine (6PPD).

Peter, K.T., Hou, F., Tian, Z., Wu, C., Goehring, M., Liu, F., Kolodziej, E.P., 2020. More Than a First Flush: Urban Creek Storm Hydrographs Demonstrate Broad Contaminant Pollutographs. Environ. Sci. Technol. 54, 6152–6165. https://doi.org/10.1021/acs.est.0c00872

Peter, K.T., Tian, Z., Wu, C., Lin, P., White, S., Du, B., McIntyre, J.K., Scholz, N.L., Kolodziej, E.P., 2018. Using High-Resolution Mass Spectrometry to Identify Organic Contaminants Linked to Urban Stormwater Mortality Syndrome in Coho Salmon. Environ. Sci. Technol. 52, 10317–10327. https://doi.org/10.1021/acs.est.8b03287

Puckett, L.J., 1995. Identifying the major sources of nutrient water pollution 7.

Rostami-Tapeh-Esmaeil, E., Vahidifar, A., Esmizadeh, E., Rodrigue, D., 2021. Chemistry, Processing, Properties, and Applications of Rubber Foams. Polymers 13, 1565. https://doi.org/10.3390/polym13101565

Seiwert, B., Nihemaiti, M., Troussier, M., Weyrauch, S., Reemtsma, T., 2022. Abiotic oxidative transformation of 6-PPD and 6-PPD quinone from tires and occurrence of their products in snow from urban roads and in municipal wastewater. Water Res. 212, 118122. https://doi.org/10.1016/j.watres.2022.118122

Seung Lim J, 2016. EPI Suite: A Fascinate Predictive Tool for Estimating the Fates of Organic Contaminants | OMICS International [WWW Document]. URL https://www.omicsonline.org/open-access/epi-suite-a-fascinate-predictive-tool-for-estimating-the-fates-of-organiccontaminants-2155-6199-1000e171.php?aid=70728 (accessed 8.25.21).

Shen, J., Amirkhanian, S., Xiao, F., Tang, B., 2009. Influence of surface area and size of crumb rubber on high temperature properties of crumb rubber modified binders. Constr. Build. Mater. 23, 304–310. https://doi.org/10.1016/j.conbuildmat.2007.12.005

Sheridan, M., 2010. The Vanderbilt Rubber Handbook, 14th edition. R.T. Vanderbilt Company, Inc., Norwalk, CT.

Sieuw, L., Jouhara, A., Quarez, É., Auger, C., Gohy, J.-F., Poizot, P., Vlad, A., 2019. A H-bond stabilized quinone electrode material for Li–organic batteries: the strength of weak bonds. Chem. Sci. 10, 418–426. https://doi.org/10.1039/C8SC02995D

Stormwater Design Manual | NC DEQ [WWW Document], n.d. URL https://deq.nc.gov/about/divisions/energy-mineral-and-land-resources/stormwater/stormwater-program/stormwater-design (accessed 12.28.22).

Su, G., Letcher, R.J., Yu, H., 2016. Organophosphate Flame Retardants and Plasticizers in Aqueous Solution: pH-Dependent Hydrolysis, Kinetics, and Pathways. Environ. Sci. Technol. 50, 8103–8111. https://doi.org/10.1021/acs.est.6b02187

Tchounwou, P.B., Yedjou, C.G., Patlolla, A.K., Sutton, D.J., 2012. Heavy Metals Toxicity and the Environment. EXS 101, 133–164. https://doi.org/10.1007/978-3-7643-8340-4_6

Tian, Z., Gonzalez, M., Rideout, C., Zhao, H., Hu, X., Wetzel, J., Mudrock, E., James, C.A., McIntyre, J.K., Kolodziej, E.P., 2021a. 6PPD-Quinone: Revised Toxicity Assessment and Quantification Method Development with a Commercial Standard. Environ. Sci. Technol. Lett.

Tian, Z., Zhao, H., Peter, K.T., Gonzalez, M., Wetzel, J., Wu, C., Hu, X., Prat, J., Mudrock, E., Hettinger, R., Cortina, A.E., Biswas, R.G., Kock, F.V.C., Soong, R., Jenne, A., Du, B., Hou, F., He, H., Lundeen, R., Gilbreath, A., Sutton, R., Scholz, N.L., Davis, J.W., Dodd, M.C., Simpson, A., McIntyre, J.K., Kolodziej, E.P., 2021b. A ubiquitous tire rubber–derived chemical induces acute mortality in coho salmon. Science 371, 185–189. https://doi.org/10.1126/science.abd6951

Tuttle, M.R., Davis, S.T., Zhang, S., 2021. Synergistic Effect of Hydrogen Bonding and π–π Stacking Enables Long Cycle Life in Organic Electrode Materials. ACS Energy Lett. 6, 643–649. https://doi.org/10.1021/acsenergylett.0c02604

Unice, K.M., Bare, J.L., Kreider, M.L., Panko, J.M., 2015. Experimental methodology for assessing the environmental fate of organic chemicals in polymer matrices using column leaching studies and OECD 308 water/sediment systems: Application to tire and road wear particles. Sci. Total Environ. 533, 476–487. https://doi.org/10.1016/j.scitotenv.2015.06.053

USEPA, 1979. DETERMINATION OF OCTANOL/WATER DISTRIBUTION COEFFICIENTS, WATER SOLUBILITIES, AND SEDIMENT/WATER PARTITION COEFFICIENTS FOR HYDROPHOBIC ORGANIC POLLUTANTS [WWW Document]. URL https://cfpub.epa.gov/si/si_public_record_report.cfm?Lab=NERL&dirEntryId=45876&fed_org_id=770&SIType=PR&TIMSType=&showCriteria=0&address=nerl&view=citation&keyword=Water&sortBy=pubDateYear&count=100&dateBeginPublishedPresented= (accessed 1.24.22).

Wagner, S., Hüffer, T., Klöckner, P., Wehrhahn, M., Hofmann, T., Reemtsma, T., 2018. Tire wear particles in the aquatic environment - A review on generation, analysis, occurrence, fate and effects. Water Res. 139, 83–100. https://doi.org/10.1016/j.watres.2018.03.051

Wang, W., Cao, G., Zhang, J., Wu, P., Chen, Y., Chen, Z., Qi, Z., Li, R., Dong, C., Cai, Z., 2022. Beyond Substituted p-Phenylenediamine Antioxidants: Prevalence of Their Quinone Derivatives in PM2.5. Environ. Sci. Technol. 56, 10629–10637. https://doi.org/10.1021/acs.est.2c02463

Zhang, Y., Liu, Z., Peng, J., 2019. Silane Modification of Crumb Rubber on the Rheological Properties of Rubberized Asphalt. Appl. Sci. 9, 4831. https://doi.org/10.3390/app9224831

Zhang, Y., Xu, C., Zhang, W., Qi, Z., Song, Y., Zhu, L., Dong, C., Chen, J., Cai, Z., 2021. p-Phenylenediamine Antioxidants in PM2.5: The Underestimated Urban Air Pollutants. Environ. Sci. Technol. https://doi.org/10.1021/acs.est.1c04500

Zhou, Q., Yang, N., Li, Y., Ren, B., Ding, X., Bian, H., Yao, X., 2020. Total concentrations and sources of heavy metal pollution in global river and lake water bodies from 1972 to 2017. Glob. Ecol. Conserv. 22, e00925. https://doi.org/10.1016/j.gecco.2020.e00925

Supplementary Materials

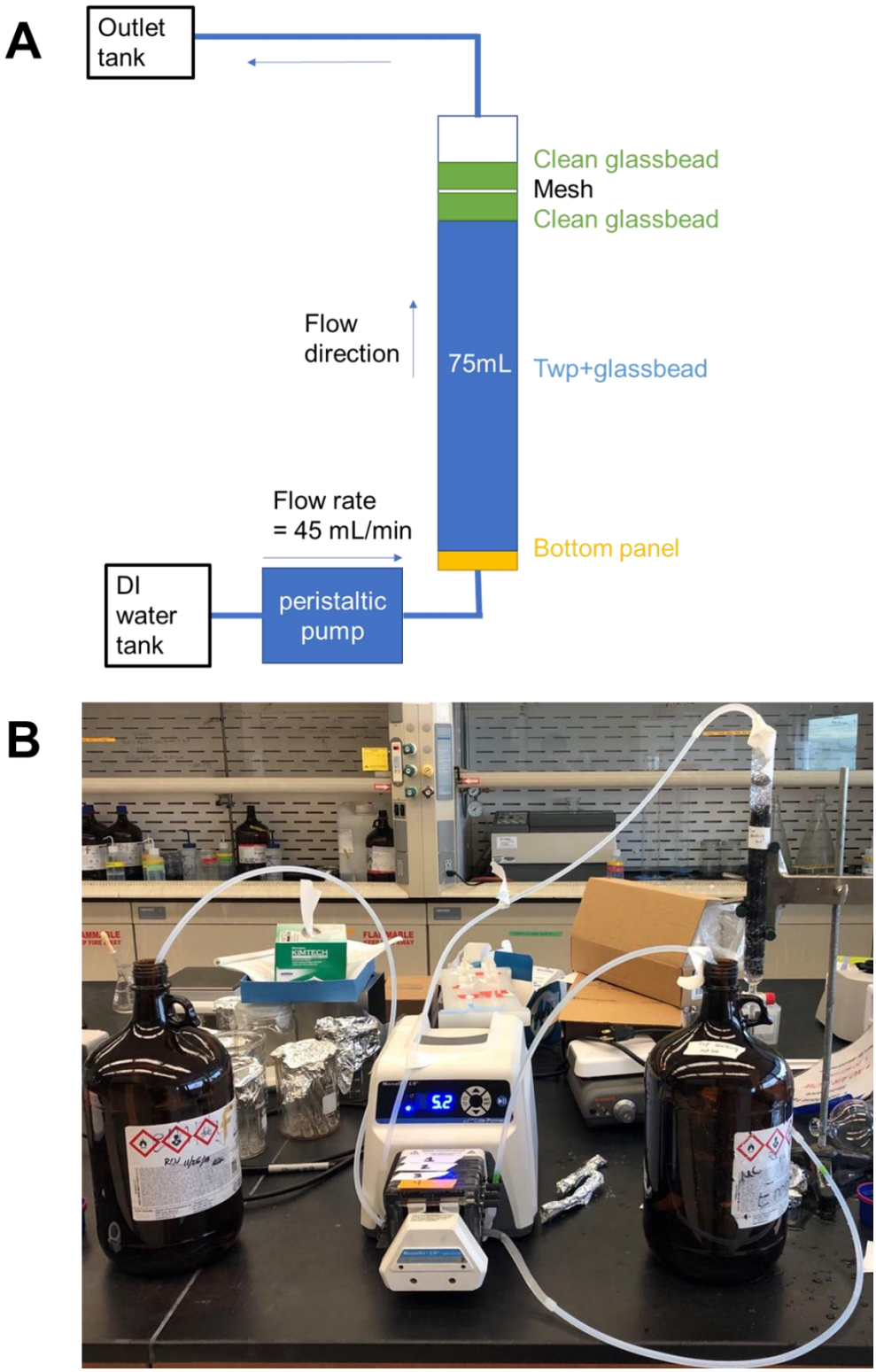


Figure S2.1. The flow-through leaching system setup. (A) Schematic diagram of the setup; and (B) photo of the leaching system setup in the lab.

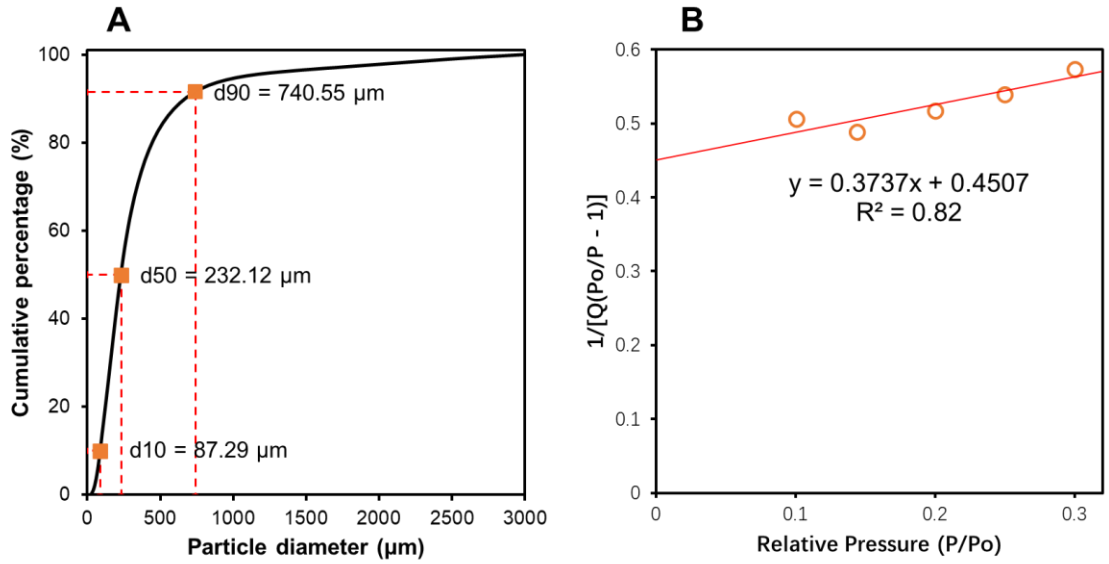


Figure S2.2. Physical characterization of TWPs used for these laboratory studies, including: (A) cumulative percentage of particle diameters measured using particle analyzer; and (B) typical BET surface area plot for the TWPs.

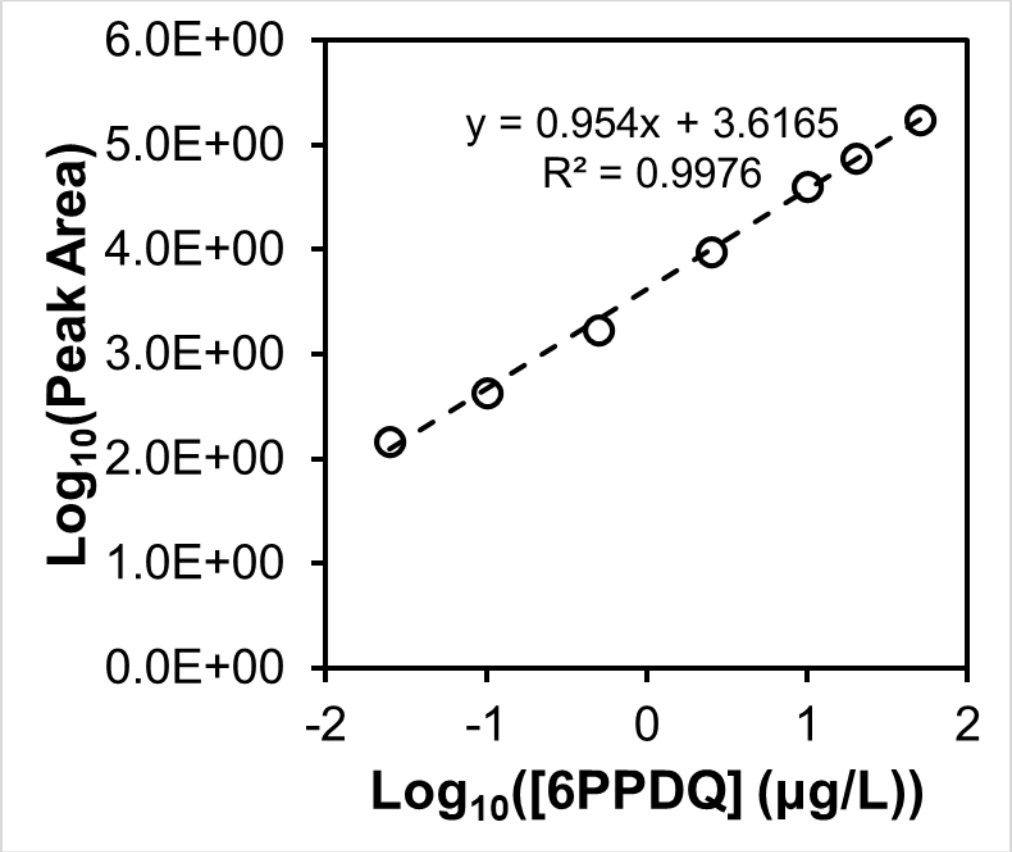


Figure S2.3. Typical calibration curve (log scale) for 6PPDQ on LC-MS/MS system. The peak areas were adjusted with 6PPDQ-d5 ISTD peak area response.

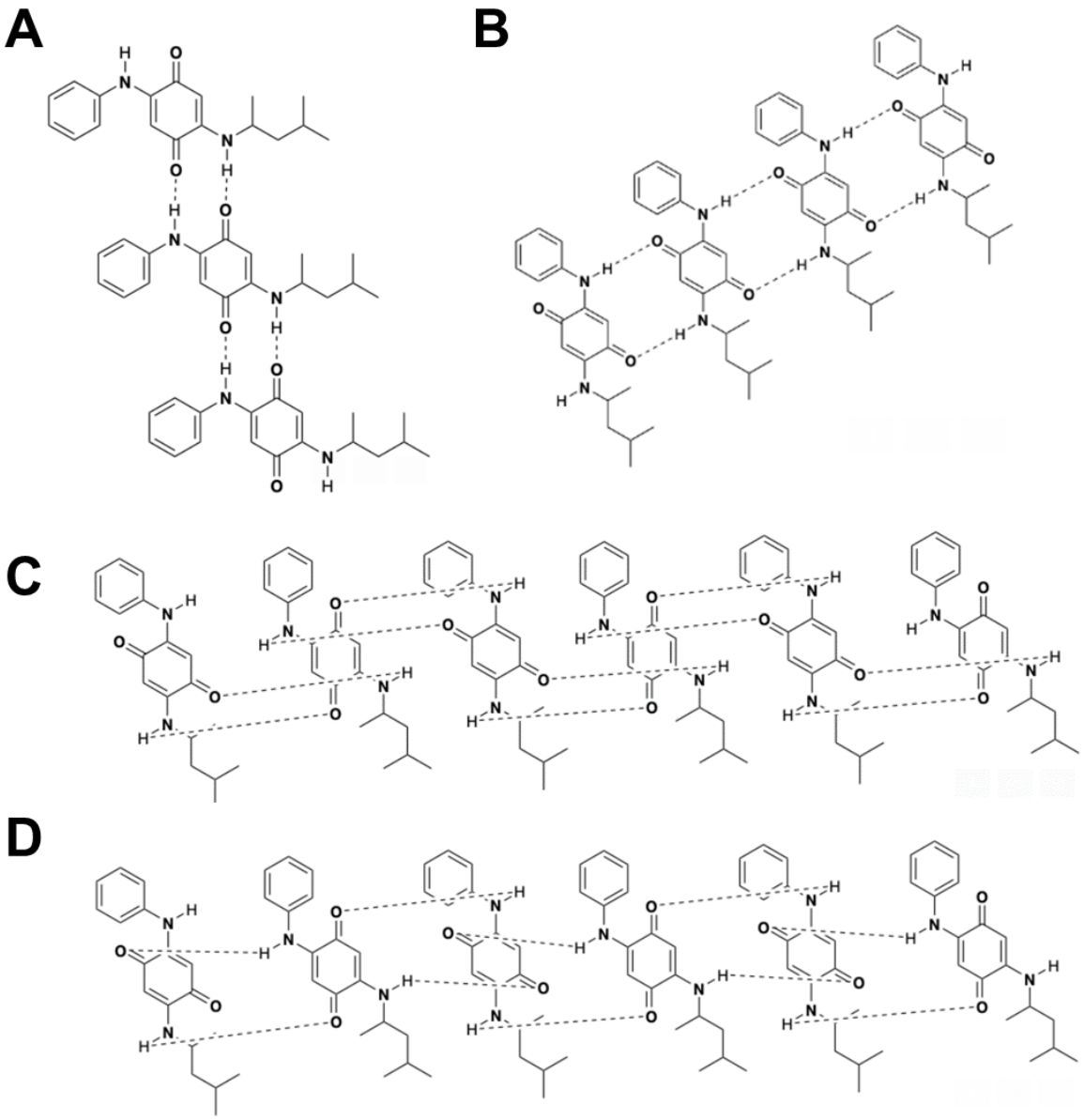


Figure S2.4. Examples of possible H-bonding structures of 6PPDQ contributing to strong solute-solute interactions for 6PPDQ solid phases.

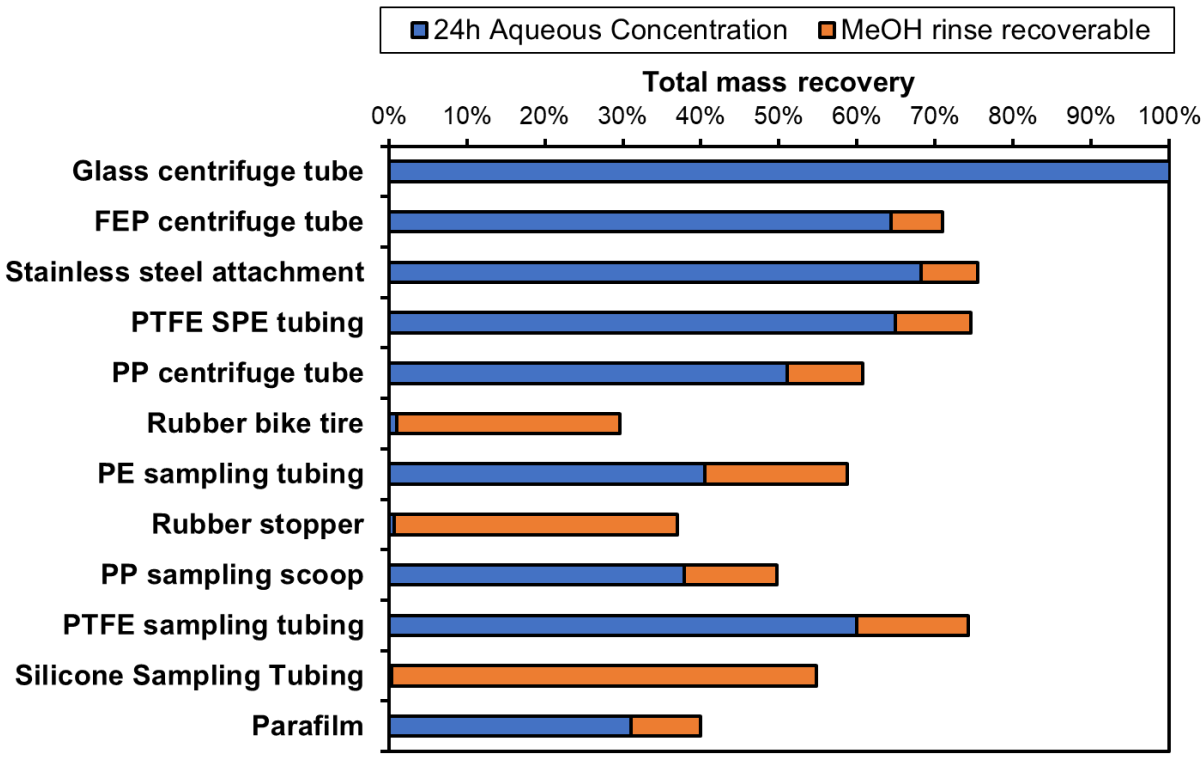


Figure S2.5. Total solvent-recoverable mass of 6PPDQ from containers and test materials after 24 h sorption test. Values represent averages from experimental triplicates.

Chapter 3: Transformation Product Formation upon Heterogeneous Ozonation of the Tire Rubber Antioxidant 6PPD (N-(1,3-dimethylbutyl)-N′-phenyl-p-phenylenediamine)

3.0 Publication and contribution statement

One publication was incorporated for this chapter as below, and XH led most of the research and development effort around these studies. For Hu et al. 2022, XH designed and conducted the experiment, processed the samples, analyzed experimental data and wrote the manuscript draft.

2. X. Hu, H. Zhao, Z. Tian, K. T. Peter, M. C. Dodd, and E. P. Kolodziej. "Transformation Product Formation upon Heterogeneous Ozonation of the Tire Rubber Antioxidant 6PPD (N-(1,3-dimethylbutyl)-N'-phenyl-p-phenylenediamine)". *Environmental Science & Technology Letters*, 2022, **9**(5), 413-419. https://doi.org/10.1021/acs.estlett.2c00187

3.1 Introduction

Tire tread wear particles (TWPs) are increasingly recognized as important sources of contaminants to air, soil, water, and biota. Wagner et al. (Wagner et al., 2018) reported annual generation of ~1,100,000 tons of tire and road wear particles (TRWP), which are hetero-aggregates of TWPs and roadway materials. After environmental release, TWPs enter soils, sediments, or surface waters and become persistent contaminant sources, (Klöckner et al., 2021a; Kole et al., 2017a; Unice et al., 2019, 2015; Wagner et al., 2018) including for some

extensively investigated contaminants like metals and PAHs. (Chen et al., 2007; Rhodes et al., 2012; Sadiktsis et al., 2012) Presently, knowledge concerning TWP-derived emerging contaminants and related transformation products (TPs) is fragmentary and incomplete, though researchers are gradually narrowing this gap. (Capolupo et al., 2020; Müller et al., 2022; Seiwert et al., 2020) Recently, 6PPD-quinone (6PPDQ), a TP of the common antioxidant 6PPD (*N*-(1,3-dimethylbutyl)-*N*'-phenyl-*p*-phenylenediamine), was reported to induce acute toxicity and mortality to coho salmon (*Oncorhynchus kisutch)* at concentrations even below 100 ng/L. (Tian et al., 2021b, 2021a) To our knowledge, 6PPD is used in all passenger and commercial vehicle tires (0.4-2% by rubber mass) (Sheridan, 2010) to provide critical anti-ozonant and antioxidant functions at tire rubber surfaces. 6PPD and some related TPs, including 6PPDQ, were recently reported in roadway-derived particles and dusts. (Huang et al., 2021; Järlskog et al., 2021; Klöckner et al., 2021a, 2021b; Seiwert et al., 2020; Unice et al., 2015; Wu et al., 2020; Zhang et al., 2021) Given the global ubiquity of TWPs in roadway runoff and surface waters (Huber et al., 2016; Seiwert et al., 2020) and their toxicological relevance, (Rochman et al., 2013; Tian et al., 2021b) there exists an urgent need to characterize TWP-derived TPs.

Previous assessments of 6PPD transformation during oxidant exposure have primarily focused on its industrial chemistry within rubber formulations, but generally with limited or no connectivity to environmental systems and using less comprehensive analytical approaches, (Cataldo et al., 2015; Huang et al., 2001; Lattimer et al., 1983) until only recently. (Klöckner et al., 2021a; Seiwert et al., 2022; Tian et al., 2021b, 2021a) The widespread dispersal of 6PPD and the acute toxicity of 6PPDQ clearly merit a deeper understanding of 6PPDQ (and other TP)

formation dynamics to enable effective management. Accordingly, the present study investigated heterogeneous gas-phase ozonation of 6PPD – likely the primary driver of 6PPDQ formation from 6PPD – using pure 6PPD and TWPs containing 6PPD. Specific objectives were to characterize 6PPD transformation dynamics, 6PPDQ generation and stability during ozonation, and to screen for and prioritize additional environmentally-relevant 6PPD ozonation-derived TPs. This study thereby provides critical conceptual and quantitative foundations for continuing efforts to elucidate 6PPD-related TP structures and examine their formation, occurrence, and fate.

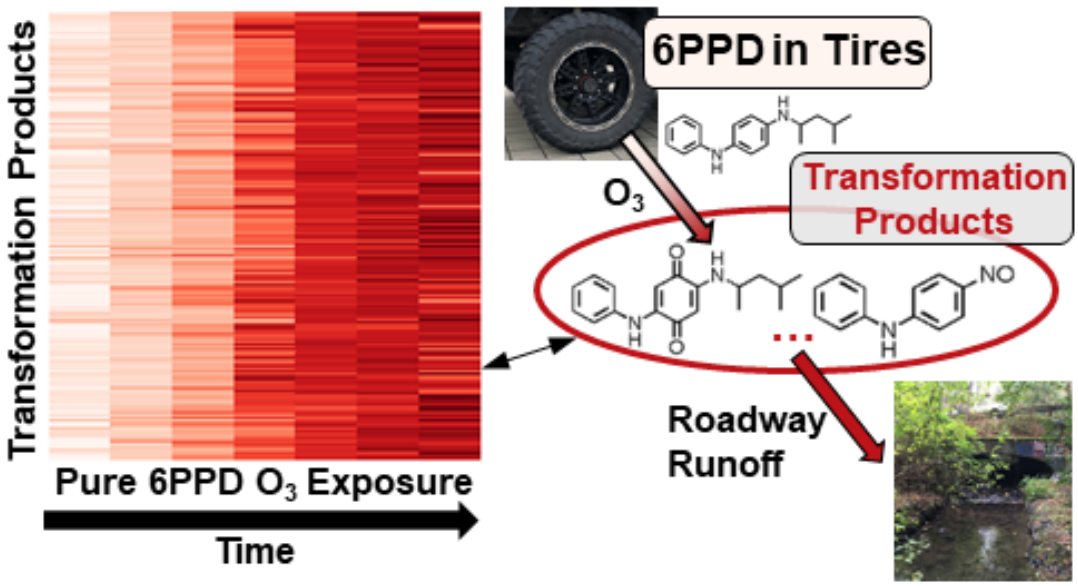


Figure 3.1. Abstract graph for the study.

3.2 Materials and Methods

3.2.1 Reagents and Materials

6PPD (industrial grade, 95%; Usolf Chemicals, Shandong, China) was finely ground with a porcelain mortar and stored in an airtight bag at 4 ºC prior to use. 6PPD-quinone (6PPDQ) (>98.8%) was purchased from HPC Standards Inc. (Atlanta, GA, USA). Methanol, acetone, formic acid, water (all Optima LC/MS grade, ≥99%) and absolute ethanol (~99.5%) were purchased from Fisher Scientific (Waltham, MA, USA) and used as received. A water purification system (Thermo Barnstead Nanopure Diamond UV; Dubuque, IA, USA) provided ultrapure deionized (DI) water. Glass-bodied chromatography columns (20 cm length × 2.5 cm diameter and 50 cm length × 1.0 cm diameter for pure 6PPD ozonation and TWP ozonation, respectively) were purchased from DWK Life Sciences (Vineland, NJ, USA). Glass fiber filters (1.6 μm) were purchased from Fisher Scientific (Pittsburgh, PA, USA) and PTFE syringe filters were purchased from Pall Corporation (Port Washington, NY, USA). A photochemical ozone (O_3) generator (UVP Model 97-0066-01, Upland, CA, USA) – relying on vacuum UV photolysis of O_2 at l = 185 nm to generate O_3 – was used for O_3 generation(Dohan and Masschelein, 1987), and an O_3 detector (IN USA Model IN2000-LC, Needham, MA, USA) – relying on UV absorbance of O_3 at l = 253.7 nm – was used for O_3 concentration measurements. Zero-grade air was purchased from Airgas (Seattle, WA, USA) for O_3-free air control experiments.

3.2.2 6PPD ozonation

6PPD slides were prepared using an approach modified from Zhou et al. (Zhou and Abbatt, 2021). Particulaly, glass slides (7.62×2.54 cm; VWR, Radnor, PA, USA) were pre-cleaned (water/methanol, baked at 400 ºC) and cut to 3.8×2 cm. 6PPD methanolic stock solution (20 µL, ~1 g/L nominal concentration; prepared within 24 h of experiments) was applied evenly to one surface of a slide (~2.6 µg 6PPD/cm^2). We note that 6PPD's substantial instability under oxic conditions(OSPAR Commission, 2006c) hinders its accurate quantification under many circumstances. To mitigate 6PPD loss/variability, analytical calibration curves used identical stock solutions and were analyzed on the same timescales as experimental systems to ensure accurate quantification. Trace 6PPDQ, TP 184 and TP 198 (peak area at 0 h <5% versus 6 h ozonation; **Table S3.1**) were detected in freshly prepared 6PPD stock, likely due to oxidation by ozone (O_3) or other oxidants such as NOx in ambient air prior to use. Peak area corrections were not applied to data. 6PPD-coated slides were prepared and dried (5 min) in triplicate under ambient air, placed into a glass-bodied chromatography column with the 6PPD-side facing up, and exposed to headspace O_3 for up to 6 h (21-23 ºC, **Figure S3.1A-B)** under room light (control experiments undertaken under equivalent conditions, but in a dark room under red light, indicated no significant difference in 6PPD consumption (two-tailed t-test; p-value=0.45) or 6PPDQ formation (p-value=0.60) trends in the presence or absence of room light). O_3 was produced from ambient air (1 L/min) by a flow-through, photochemical O_3 generator. O_3 concentrations at the outlet of the chromatography column were monitored continuously, while inlet O_3 concentrations (stable throughout experiments at 360±12 ppbv) were measured using

column inlet/outlet two-way valves and a bypass line, in each case using a flow-through O_3 detector (**Figure S3.1)**. The 360±12 ppbv inlet O_3 concentration was higher than typical ground-level concentrations (1-84 ppbv for Seattle metropolitan area) (US EPA, 2014) but similar (albeit with shorter exposure time) to that of previous work on 6PPD ozonation by Lattimer et al. (500 ppb for 7 days).(Lattimer et al., 1983) Additional control experiments (6 h; triplicate slides) included exposure of pure methanol (no 6PPD) on glass slides to O_3, and exposure of 6PPD-coated slides to ambient air and O_3-free compressed zero-grade air (Airgas, Seattle, WA, USA) using the same setup as described for ozonation experiments. After reaction, slides were extracted in amber glass bottles by submerging in 10 mL methanol and sonicating (30 min). Extractions of non-O_3 exposed 6PPD (20 µg/glass slide; n=3) and 6PPDQ (2 ug/glass slide; n=3) using identical procedures indicated 88±3% and 97%±4% recoveries, respectively.

Tire tread wear particles (TWP) were made from the treads of nine tires (brand, vehicle type, and use history in **Table S3.2**; mixed in equal weights) by physical abrasion with an angle grinder with a steel carbide disk (Parkes MCM90; Miller Tire, Wauseon, OH, U.S.A.). TWPs were stored in an airtight bag to isolate them from further exposure to ambient air prior to experiments. The treated TWPs (Tian et al., 2021b) were then ozonated (6 h) in triplicate by dispersing TWPs (160 mg) amongst glass beads (3 mm diameter) in a glass-bodied chromatography column with a porous glass frit and PTFE filter (0.2 µm) at the column outlet (**Figure S3.1C-D**). Identical controls were exposed to 1 L/min ambient air and compressed zero grade air in triplicate (6 h). After exposure, TWPs/glass beads were extracted by transferring into an amber glass bottle and sonicating (30 min) in 32 mL methanol. Extracts

were filtered to remove TWPs (glass fiber then PTFE syringe filters), and 100 μL aliquots diluted in methanol (10% v/v) for subsequent analyses. Identical triplicate extractions of TWPs/glass beads spiked with 1.5 μg 6PPDQ indicated ~70±7% recovery.

All pure 6PPD and TWP sample extracts (1 mL) were analyzed within 24 h, except zero grade air control extracts, which were analyzed following 2-month storage (-20 °C).

Archived methanolic extracts from a previous study(Tian et al., 2021b) (stored at -20 ºC) of TWP aqueous leachate (leached with DI water, 24 h, up-flow column; sampled 2018–2019) and of runoff from two multilane roadways were concurrently reanalyzed alongside samples from the present study (**Table S3.1**). (Peter et al., 2018; Tian et al., 2021b) In addition, original 6PPD and 6PPDQ loads in TWPs were estimated via sequential solvent extractions, to explore the maximum extractable mass of 6PPD and 6PPDQ. First, 0.1 g TWP was weighed in a centrifuge tube (15 mL, PYREX). Then, 10 mL of methanol was added to the tube, and the tube was capped and sonicated for 30 minutes. After sonication, 9.5 mL solvent was removed with a glass pipet and the TWPs were dried under nitrogen within the tube. Then, 10 mL of acetone was added, followed by the same procedure of sonication, pipetting solvent out, and drying the TWPs. Lastly, 10 mL of methanol was added into the tube and sonicated for 30 minutes to evaluate the presence of residual chemical mass in the TWPs. All three solvent extracts were then separately analyzed by UHPLC-QTOF-MS for quantification, by methods detailed in the following sections. Based on the mass extracted by the final methanol extraction (3% of total 6PPD mass, 5% of total 6PPDQ mass relative to the total mass of all three extractions), the three-step sequential extraction extracted most of the mass available in TWPs.

The extractable masses of 6PPD (2.8 mg per g TWP) and 6PPDQ (13 and 30 μg per g TWP for pre-ozonation and post-ozonation, respectively) were calculated by combining the measurements from all three extraction steps.

3.2.3 LC-QTOF analysis

All samples were analyzed using an Agilent 1290 Infinity ultrahigh performance liquid chromatograph (UHPLC) coupled to an Agilent 6530 quadrupole time-of-flight high-resolution mass spectrometer (QTOF-HRMS; Santa Clara, CA, USA). A reverse-phase C18 column (Agilent ZORBAX Eclipse Plus 2.1×100 mm, 1.8 μm) with a C18 guard column (2.1×5 mm, 1.8 μm) was used for the UHPLC separation at 45 ºC with 5 μL injection volume. The ozonated products were separated with a gradient elution with mobile phases of 0.1% formic acid in each of deionized water (A) and methanol (B) as follows: 5% B at 0-1 min, 50% B at 4 min, 100% B at 17-20 min, 5% B at 20.1 min; stop time 22.5 min; post-time 2 min. The flow rate was 0.4 mL/min. Full scan data were acquired under 2 GHz Extended Dynamic Range mode at a range of 100-1700 m/z; MS/MS data were collected by data-dependent acquisition at the range of 50-1700 m/z, with collision-induced dissociation at 10, 20, and 40 eV. Quality assurance and quality control (QA/QC) included mass calibration before each analytical batch and mass accuracy correction by continuous infusion of purine and HP-921 calibrants. Solvent blanks and internal standards (details available in previous studies (Du et al., 2017; Peter et al., 2018; Tian et al., 2020)) were each injected every 12 samples (~4.5 h) to assess carry-over (none observed) and variation in instrumental response throughout analysis batches (one analytical batch per experimental system; <30% relative standard deviation of internal standard

peak areas). 6PPD and 6PPDQ were quantified using commercial standards; concentrations were not recovery-normalized given acceptable 6PPDQ recoveries for slides and TWP extractions (97±4% and 70±7%, respectively). Calibration curves (5-200 µg/L for 6PPD and 1-50 µg/L for 6PPDQ; no weighting was used) were prepared by dilution into methanol of methanolic (6PPD) and ethanolic (6PPDQ) analyte stock solutions. 6PPD and 6PPDQ were quantified using commercial standards; concentrations were not recovery-normalized given acceptable 6PPDQ recoveries for slides and TWP extractions (97 ± 4% and 70 ± 7%, respectively). The LODs (0.43 ng/mL and 0.56 ng/mL for 6PPDQ and 6PPD, respectively) and LOQs (1.4 ng/mL and 1.9 ng/mL for 6PPDQ and 6PPD, respectively) were estimated based on the signal-to-noise ratios of the lowest points on the respective calibration curves.

3.2.4 HRMS data processing and clustering analysis

After data acquisition, raw data files (Agilent .d format) were converted to .abf format (Reifycs Abf Converter); MS-DIAL (version 3.46) (Tsugawa et al., 2015) was used for primary data processing (i.e., non-target feature extraction and alignment). Parameter settings were as follows: *m/z* tolerance for feature extraction 0.005; minimum peak height 2000; minimum peak width 5 scans; *m/z* slice 0.01 Da; alignment *m/z* tolerance 0.015 Da; alignment retention time tolerance 0.1 minutes. Initial data reduction used Python programming language (version 3.7.4) with Pandas (version 1.2.0) to isolate features with maximum peak area > 10000, *m/z* < 900 Da, and retention time between 3-18 min; and that were present in 3 of 3 laboratory replicates, and present at peak area 5-fold greater than peak area in any of the solvent blanks or the pure

6PPD ambient air exposure negative controls. Features satisfying all of these criteria were retained.

Chemical features generated from pure 6PPD ozonation that met initial data reduction criteria (936 features) were further prioritized as potential 6PPD ozonation transformation products by the Density-Based Spatial Clustering of Applications with Noise (DBSCAN; **Figure 3.2**) algorithm (Ester et al., n.d.) using Python programming language with scikit-learn package (version 0.23.0).

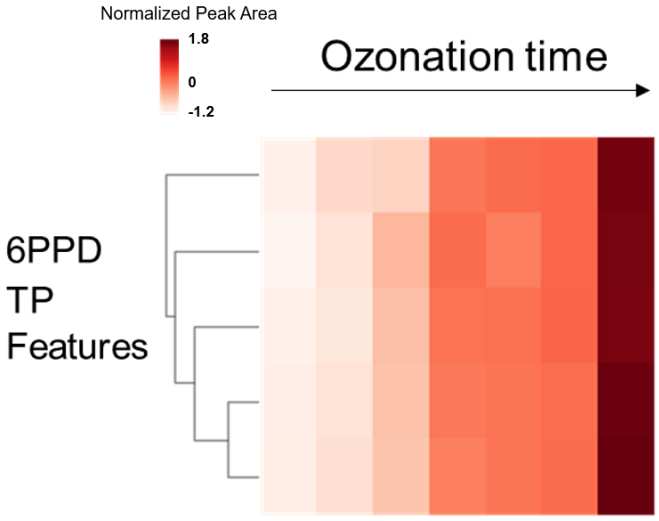


Figure 3.2. Example clustering heatmap from the DBSCAN analysis. Each column represents an individual sample (corresponding to time points of sampling in this example), and each row represents a unique chemical feature that is a potential 6PPD transformation product (TP). The values presented in the plot are z-score normalized peak area values. Full data are available in **Table S3.1**.

Z-score normalization of data was conducted to eliminate the data skewness and kurtosis before the analysis as below:

$$z = (x - \mu) / \sigma$$

where z is z-score, x is the peak area of the feature, μ is the average peak area of the feature across all the samples, and σ is the standard deviation of the peak areas. Algorithm parameters, including *minPts* and *eps*, were tuned iteratively by manually checking the cluster quality to optimize the clustering outcome. A typical clustering heatmap of features with increasing relative abundance is shown in **Figure 3.2**. The determined parameters (*minPts* = 3, *eps* = 0.4) were used to produce final clustering results (297 features). Features in the pure 6PPD post-ozonation samples that possessed a peak area >100000 after 6 h O_3 exposure were retained (98 features). All 98 features were then manually checked to remove software artifacts (e.g., S/N ratio > 20 and peak shape examined; 50 features remained at this stage). Then, features representing multiple adducts from the same compound were collapsed into a single feature using MS-DIAL assigned adduct annotation, retaining only the adduct form with the highest peak area (46 features). The 46 features that passed all of the above criteria were regarded as strong candidate transformation products from 6PPD ozonation.

Analogous TWP ozonation data processing was also conducted following the same criteria as above (comparing with TWP samples prior to gas exposure and TWP ambient air exposure negative control samples), resulting in retention of 1298 features. The residual features were regarded as probable TWP-derived TP features and subjected to statistical testing. Considering that TP formation and recoveries from TWP samples might have been subject to complex matrix effects, features that possessed a peak area > 50000 in TWP samples (either before or after ozonation; 216 features) and that passed a t-test for peak area comparison of

before vs. after ozonation samples with significant differences (p-value < 0.05, from 3 analytical replicates) were retained; resulting in retention of 136 features. Note that more features passed the data reduction filters in TWP ozonation samples (relative to pure 6PPD ozonation samples) due to the presence of other rubber components, even though the final extracts of both pure 6PPD ozonation samples (2 mg/L as precursor 6PPD) and TWP ozonation samples (14 mg/L as precursor 6PPD, calculated from 2.8 mg 6PPD/g TWP) were adjusted to the same nominal in-vial 6PPD concentration (i.e., by diluting the TWP ozonation samples 10-fold using methanol prior to instrument analysis).

Formula assignment was performed with ChemCalc (Patiny and Borel, 2013) using monoisotopic mass-to-charge ratios; elemental constraints of C (0-50), H (0-100), N (0-4), and O (0-5); and a 20 ppm mass error tolerance. Identification confidence was assigned at level 1 (confirmed with commercial standard), level 4 (unequivocal molecular formula) or level 5 (exact mass). (Schymanski et al., 2014)

3.3 Results and Discussion

3.3.1 6PPD transformation

Ozonation reactions during glass slide exposures were assumed to occur at the interface of the gas-phase and 6PPD layers on slides and not in the gas-phase itself, given negligible 6PPD vapor pressure (6.85×10^{-3} Pa) under ambient conditions. Column inlet (i.e., generator output) O_3 concentrations were stable at 360±12 ppbv; column outlet concentrations were

initially 148±7 ppbv (reflecting substantial consumption within the reactor) and increased over time, reaching the inlet level after 4 h. Similarly, 6PPD mass loss plateaued after 4 h (no further transformation between 4 to 6 h), indicating the cessation of detectable transformation reactions. After 6 h ozonation, up to 81% of initial 6PPD mass had reacted (**Figure 3.3A, Figure S3.2A**), suggesting that reactions with O_3 and potentially also O_3-related secondary oxidants (e.g., hydroxyl radical) were a primary driving force for 6PPD oxidative transformation. NO_x formation during passage of ambient or zero grade air through the photochemical O_3 generator used here can be excluded because of negligible spectral overlap and quantum yield for N_2 photolysis at the 185 nm lamp emission wavelength responsible for O_3 generation. (Dohan and Masschelein, 1987; Gürtler et al., 1977; Li et al., 2013; Lofthus and Krupenie, 1977)

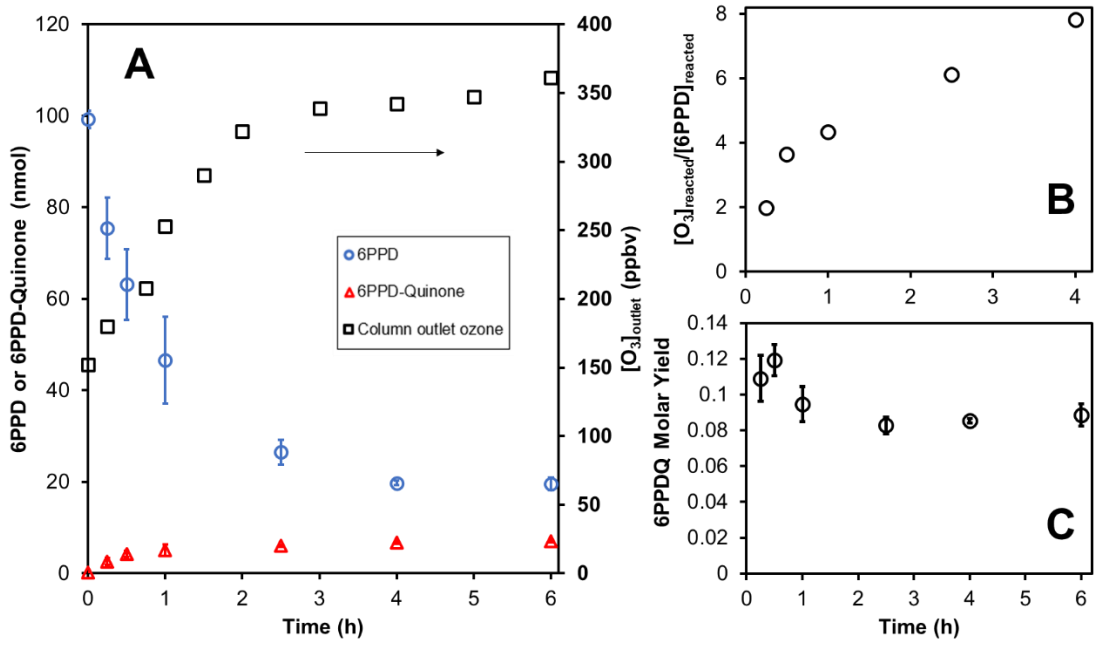


Figure 3.3. Concentration dynamics of 6PPD and 6PPDQ during heterogeneous gas-phase ozonation (360 ppbv) of pure 6PPD on glass slide surfaces. (A) 6PPD and 6PPDQ mass (per individual slide; left axis) and normalized column outlet O_3 concentration (right axis) vs. time during O_3 exposures. (B) The molar ratio of reacted O_3 to reacted 6PPD (total mass from triplicate slides in each experimental batch). Each point represents an average molar ratio over

each sampling interval (i.e., 0-0.25 h, 0.25-0.5 h, 0.5-1 h, 1-2.5 h, 2.5-4 h, respectively); the last sampling interval (4-6 h) was omitted because 6PPD consumption was extremely low (0. 1 nmol compared to 24 nmol over the first sampling interval). (C) Molar yield of 6PPDQ per reacted mole of 6PPD over time (for individual slides). Error bars in (A) and (C) correspond to the standard deviation from three replicate slides.

In O_3-free zero grade air controls, no significant loss of 6PPD mass was observed after 6 h (**Figure S3.3A**; p-value=0.58), showing the stability of 6PPD to molecular oxygen under these conditions, and highlighting the role of O_3 (and/or associated secondary oxidants) in short-term exposures to ozone-containing air. In ambient air controls, where O_3 was not detectable (<10 ppbv in inlet air) but was likely present at trace levels, the mass of 6PPD transformed after 6 h equaled the 6PPD mass reacted during a 1 h O_3 exposure. The lack of observable effect of O_2 in zero grade air controls indicates that ongoing 6PPD reactions in ambient air were likely due to background levels of O_3 or potentially other gas-phase oxidants present in the ambient air (e.g., NO_x species). While beyond the scope of this study, future investigation of these possibilities is recommended.

Similar behavior was observed for TWPs, with 59±2% of initially detectable 6PPD mass reacting over 6 h in ozonated TWP samples (from 2.8 to 1.15 mg 6PPD/g TWP; **Figure S3.4A**) relative to a 10±2% decrease of 6PPD in TWP ambient air controls, while no 6PPD degradation was measured in zero grade air controls (**Figure S3.3**, p-value=0.47).

Both systems exhibited unreacted 6PPD after extended exposures to O_3 (15.5 h and 24 h trials with pure 6PPD and TWP, respectively). Although further investigation of reaction mechanisms is required, the industrial literature reports that 6PPD oxidation generates protective surface films via oxidative polymerization or radical interactions. (Cataldo et al.,

2015; Huang et al., 2001; Huntink, 2003; Lattimer et al., 1983) Such phenomena likely protected deeper 6PPD layers from oxidation in both systems. Thus, although further transformation was possible, some 6PPD mass may have behaved as inert or inaccessible despite high inherent reactivity. This observed surface oxidation phenomenon may also partially explain the extended lifetime of 6PPD used in actual tires. (Huntink, 2003)

The apparent molar O_3:6PPD reaction stoichiometry was initially ~2:1 (**Figure 3.3B**), likely reflecting "true" 6PPD-O_3 interactions, as 6PPD was initially abundant and in excess relative to any O_3-reactive TPs that exert added O_3 demand. The O_3:6PPD molar ratio increased over the entire reaction period, suggesting a growing importance of ozonation consumption by 6PPD-derived TPs, in parallel with decreased 6PPD consumption. (Cataldo et al., 2015; Huntink, 2003)

Average ambient ground-level O_3 concentrations are considerably lower than the concentrations applied here (360 ppbv); thus, we also expect slower 6PPD consumption and associated TP formation kinetics under typical environmental conditions. For example, daily 8 h average O_3 concentrations in urban areas of Washington State are typically 1-84 ppbv (US EPA, 2014) and are regulated at 70 ppbv. (US EPA, n.d.) Episodic pollution events, especially in summer, sometimes yield concentrations >100 ppbv. (Saini et al., 2014; Simon et al., 2015)

3.3.2 6PPDQ formation

The major TPs prioritized from HRMS detections from pure 6PPD ozonation included 6PPDQ (**Figure 3.4**), a highly toxic and environmentally relevant TP. (Klöckner et al., 2021a; Seiwert et al., 2022; Tian et al., 2021b, 2021a)

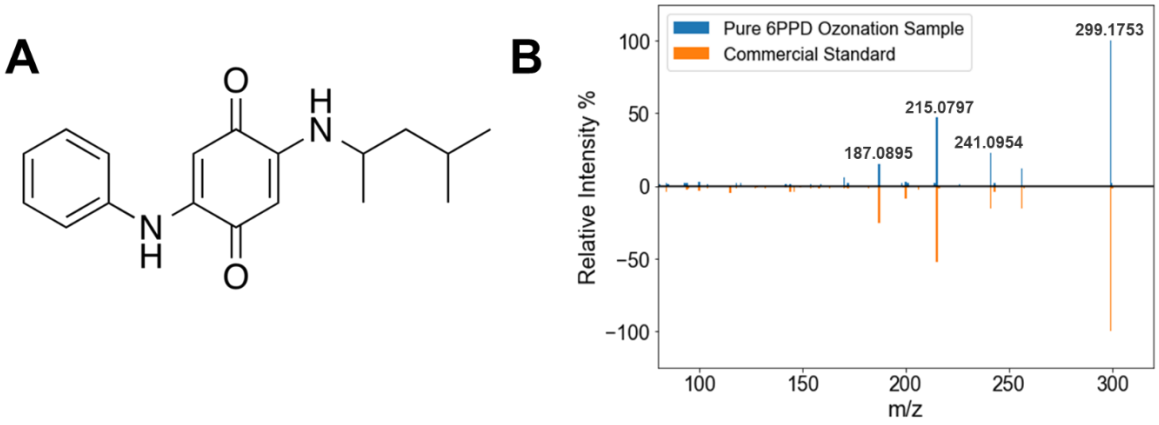


Figure 3.4. Chemical structure and MS/MS spectra of 6PPDQ. (A) Chemical structure of 6PPDQ. (B) MS/MS spectra for 6PPDQ from pure 6PPD ozonation sample (top) and from 6PPDQ commercial standard (bottom) from the average of different CID energies (10, 20, and 40 eV)

During ozonation of pure 6PPD, 6PPDQ formation stabilized after 4 h (**Figure 3.3A, Figure S3.2B**). Short-term 6PPDQ molar yield (per mole 6PPD consumed) averaged ~9.7% and was consistent throughout the reaction (**Figure 3.3C**), indicating 6PPDQ was a stable and potentially terminal TP at these timescales and conditions. No 6PPDQ formation was observed in zero grade air controls, consistent with lack of 6PPD transformation under such conditions. In pure 6PPD ambient air controls, 6PPDQ formation was observed at 6 h (1.9 nmol) at levels similar to those of 15 min O_3 exposures (2.6 nmol), consistent with 6PPD reactivity toward below-LOD levels of background O_3 (or potentially other oxidants) under ambient conditions.

The chemistry investigated here also applies to gas-solid interfaces such as tire surfaces. During TWP ozonation, 6PPDQ concentrations increased >2-fold relative to initial conditions (from ~13-30 µg 6PPDQ/g TWP; **Figure S3.4B**), although some 6PPDQ already existed in the TWPs prior to ozonation, likely from prior exposure of TWP to O_3 in ambient air. The observed molar yield of 6PPDQ from TWP-associated 6PPD was ~0.95% under this condition, representing a minimum estimate due to 6PPD instability, reactive co-contaminants, and the complexity of TWP surface matrices. No 6PPDQ formation was observed in the zero grade air exposed control samples compared to non-exposed TWP samples (**Figure S3.3B**; p-value=0.24). Complemented by other recent studies, (Klöckner et al., 2021a; Seiwert et al., 2022; Tian et al., 2021b) these data suggest that 6PPD reaction with gas-phase O_3 is a primary driver of 6PPDQ formation in TWPs and related environmental systems.

6PPD is used industrially at initial concentrations of 0.4-2% by mass within tires (i.e., 4000-20,000 µg 6PPD/g TWP). (Sheridan, 2010) In rubber materials, 6PPD gradually diffuses to exposed surfaces where it becomes available for gas-phase reaction (the mechanistic basis of its protective antioxidant function) and subsequent environmental release. (Huntink, 2003) While such transport is almost certainly affected by many poorly understood factors inherent to the rubber matrix and associated microenvironments, all 6PPD added to tires is ultimately intended to react with O_3 and other oxidants. (OSPAR Commission, 2006c)

Based on Kole et al.'s estimate of tire and road wear particle (TRWP) emission for the US (1,250,000-1,800,000 metric tons TRWP/year) (Kole et al., 2017b) and industrial tire rubber recipes (0.4-2% 6PPD mass per total tire rubber mass), "low-emission" and "high-

emission" 6PPDQ release potentials were calculated as follows, using the low and high values of the range of 6PPD mass per total tire mass used in tire rubber recipes, the molar 6PPDQ yields determined here from the TWP ozonation experiments (0.95%) and pure 6PPD-coated slide ozonation experiments (9.7%), and the low and high values of the range of annual TWP releases based on the TRWP estimation by Kole et al. (assuming TRWP consists of 50% TWP), respectively:

Low-emission scenario:

Annual total 6PPDQ available (mass):

$$\frac{0.004 \text{ metric ton 6PPD}}{\text{metric ton TWP}} * \frac{0.0095 \text{ mol 6PPDQ}}{\text{mol 6PPD}} * \frac{298.4 \frac{\text{g}}{\text{mol}} \text{6PPDQ}}{268.4 \frac{\text{g}}{\text{mol}} \text{6PPD}} * 1250000 \frac{\text{metric ton TRWP}}{\text{year}} * 50\% \approx 26 \frac{\text{metric ton 6PPDQ}}{\text{year}}$$

High-emission scenario:

Annual total 6PPDQ available (mass):

$$\frac{0.02 \text{ metric ton 6PPD}}{\text{metric ton TWP}} * \frac{0.097 \text{ mol 6PPDQ}}{\text{mol 6PPD}} * \frac{298.4 \frac{\text{g}}{\text{mol}} \text{6PPDQ}}{268.4 \frac{\text{g}}{\text{mol}} \text{6PPD}} * 1800000 \frac{\text{metric ton TRWP}}{\text{year}} * 50\% \approx 1900 \frac{\text{metric ton 6PPDQ}}{\text{year}}$$

As calculated above, the minimum 6PPDQ mass generated from U.S. tire tread wear is anticipated to range from 27-1800 metric tons/year. This estimate is based on TWP surface-associated 6PPDQ formation determined at a single point in time. Therefore, it probably underestimates ultimate 6PPDQ formation potential that will be sensitive to time scales and strongly depend on diffusion rates and lifetimes of tire ingredient chemicals at rubber surfaces.

With respect to transport potential, the computationally predicted logK_{ow} and aqueous solubility (C_w^{sat}) of 6PPDQ are generally lower and higher than those of 6PPD (**Table 3.1**), respectively (estimated ranges of logK_{ow}=1.51-3.98 and C_w^{sat}=2.4-140,000 mg/L for 6PPDQ and logK_{ow}=4.68-4.91 and C_w^{sat}=1.9-11 mg/L for 6PPD from EPI-Suite (Seung Lim J, 2016), Marvin ("ChemAxon - Software Solutions and Services for Chemistry & Biology," n.d.) and SPARC (HILAL, S. H., n.d.)). Although these data suggest easier leaching/transport of the more polar 6PPDQ, they reflect the neutral species of each compound and do not account for acid-base speciation. 6PPD (predicted pK_a=6.46 from Marvin) should be present substantially as the cationic $6PPD^+$ at circumneutral pH, with protonation at the alkylaromatic secondary amine substantially lowering its effective logK_{ow} and raising its effective C_w^{sat}. However, 6PPDQ (predicted pK_a= –4.02 from Marvin) should be almost exclusively in its neutral form. Consistent with this, Hiki et al. recently reported a measured C_w^{sat}=85 µg/L for 6PPDQ, which is substantially lower than the 1 mg/L 6PPD solubility reported by OSPAR (OSPAR Commission, 2006c) (though still ~900x higher than the reported LC_{50} of 95 ng/L for coho salmon (Hiki et al., 2021; Tian et al., 2021a)).

Table 3.1. Computationally predicted chemical properties of 6PPD and 6PPDQ

Chemical	**logK_{ow}**	**C_w^{sat} (mg/L)**	**pK_a**
6PPD	4.68-4.91	1.9-11	6.46
6PPDQ	1.51-3.98	2.4-140,000	-4.02

Note: logK_{ow} and C_w^{sat} values were predicted with EPI-Suite (Seung Lim J, 2016), Marvin ("ChemAxon - Software Solutions and Services for Chemistry & Biology," n.d.) and SPARC (HILAL, S. H., n.d.). pK_a values were predicted with Marvin, and correspond to deprotonation of the cationic (ammonium) forms of the alkylaryl and diaryl nitrogens for 6PPD and 6PPDQ, respectively.

This underscores the critical importance of solution conditions (e.g., pH) with respect to relative 6PPD and 6PPDQ speciation and solubility in the environment. Nevertheless, water is expected to leach away trace amounts of 6PPDQ (and other stable TPs) from intact and dispersed rubber surfaces, leading to long-term discharge over materials lifetimes (including from products made with recycled/repurposed tire rubbers) until 6PPD is fully depleted. Additional work to evaluate 6PPD mass balances and dynamics within tire rubbers and associated recycled rubber products is needed to accurately quantify expected source terms for TPs and to assess potential longer-term impacts on affected environments.

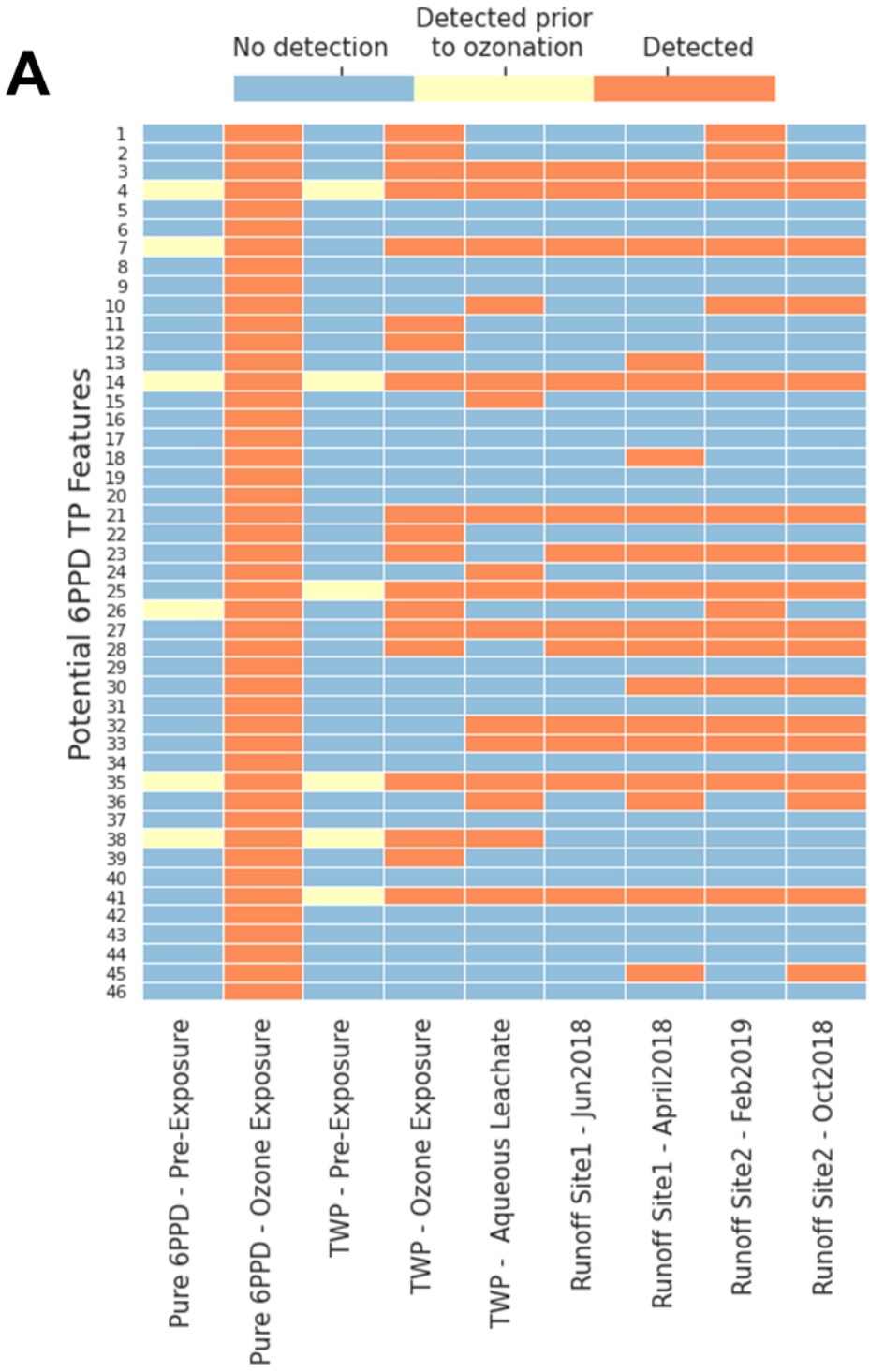
3.3.3 Other 6PPD-Derived TPs

Beyond 6PPDQ, 45 additional LC-HRMS detected features representing potential TPs were formed during pure 6PPD ozonation and isolated by unsupervised clustering (**Table S3.1**). (Cataldo et al., 2015; Lattimer et al., 1983) These included several previously reported TPs (e.g., 4-nitrosodiphenylamine [level 1 identification] (Schymanski et al., 2014), $C_{12}H_{12}N_2$, $C_{16}H_{22}N_2O_2$, and $C_{18}H_{22}N_2O$ [level 4 identifications]), (Cataldo et al., 2015; Huntink, 2003; Klöckner et al., 2021a; Lattimer et al., 1983; Rapta et al., 2009) as well as detections presumably originating from 6PPD industrial impurities, isomers, or secondary products. As discussed below, several of these TPs (6PPDQ, 4-nitrosodiphenylamine, $C_{18}H_{22}N_2O$) were reported in prior studies of 6PPD ozonation, further supporting their identification as ozonation products. (Lattimer et al., 1983; Seiwert et al., 2022)

In TWP ozonation, a statistical analysis described in **section 3.2.4** isolated 136 features with peak areas that significantly (p-value <0.05) increased (n=76) during ozonation. TWP

ozonation-derived features were generally low molecular weight (<400 Da) and contained oxygen, consistent with oxidative TP formation during O_3 exposure. Among these 76 increased features, 18 matched those prioritized as TP candidates from pure 6PPD ozonation. Additionally, *m/z* 213.1025 at a retention time of 6.15 min, whose detected abundance increased during pure 6PPD ozonation, was detected during TWP ozonation without changing abundance. The overlap in TPs from pure 6PPD and TWPs suggested that 6PPD functions as a primary, ongoing reactant that contributes to substantial TP formation potential during TWP ozonation, consistent with its role in protecting rubber from oxidative damage and underscoring the importance of reactive antioxidants like 6PPD as a major source of TPs from tire rubbers.

To prioritize environmentally relevant TPs, features detected in all replicates (n=3/3) after TWP ozonation, in reanalyzed laboratory-generated TWP aqueous leachates (n=4/4),(Tian et al., 2021b) and in reanalyzed roadway runoff extracts from two sites (site1: SR520, Seattle, WA, USA; site2: SR518, Seattle, WA, USA; n=12/12)(Peter et al., 2018; Tian et al., 2021b) were aligned with the 46 presumptive TPs detected in the pure 6PPD system (n=3/3) (**Figure 3.5A**). Among the 19 TP features common across pure 6PPD and TWP ozonation, 9 were detected in all TWP aqueous leachate and roadway runoff samples (**Figure 3.5B**). In addition to 6PPDQ, four TPs with the highest peak area abundance in roadway runoff (TPs 274, 184, 198, and 282) – some of which were previously observed in surface waters (Lattimer et al., 1983; Peter et al., 2020, 2019, 2018) – were tentatively identified (**Table S3.1**).


A
No detection
Detected prior to ozonation
Detected
Potential 6PPD TP Features
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
Pure 6PPD - Pre-Exposure
Pure 6PPD - Ozone Exposure
TWP - Pre-Exposure
TWP - Ozone Exposure
TWP - Aqueous Leachate
Runoff Site1 - Jun2018
Runoff Site1 - April2018
Runoff Site2 - Feb2019
Runoff Site2 - Oct2018

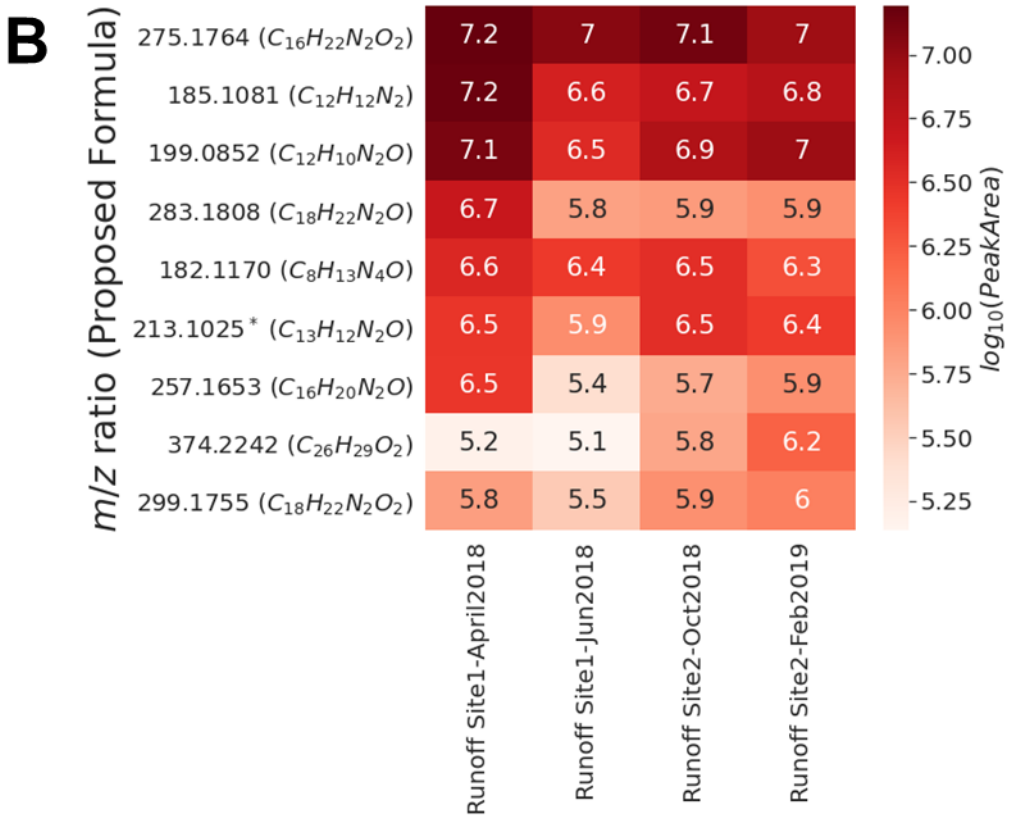


Figure 3.5. Summary of TPs from pure 6PPD-coated slide ozonation experiments and relation to TPs detected in samples of TWPs (pre- and post-ozonation), TWP aqueous leachate, and roadway runoff. (A) Detection heatmap for all 46 presumptive TP features observed by ESI+ non-target screening. Each row represents an individual TP feature; each column represents a group of triplicate samples from the same sampling event or experiment. Colors in the heatmap represent feature presence: blue – no detection; orange – detected; yellow – initially present in pure 6PPD and TWP samples (prior to ozonation) at peak areas at least 5-fold and 2-fold lower, respectively, than after ozonation.*The numerical index for each row in panel (A) corresponds to the feature with the same index number in **Table S3.1**. (B) Peak area heatmap (scaled to $\log_{10}$ transformed observed peak area) of 9 TPs from (A) that were detected in all ozonated TWP samples, TWP aqueous leachate samples, and roadway runoff samples. TPs were ranked (top to bottom in heatmap) by observed maximum peak area in roadway runoff samples. All formulas were annotated in ChemCalc using monoisotopic mass-to-charge ratios; all reported *m/z* ratios represent $[M+H]^+$. *TP *m/z* 213.1025 at a retention time of 6.15 min (TP feature 14 in (A)) was detected in all samples but did not increase in peak area during TWP ozonation.

TP 274 (*m/z* 275.1741 at 10.35 min, $C_{16}H_{22}N_2O_2$, level 4 identification) was the highest peak area TP across all environmental samples, with >10-fold higher peak area than 6PPDQ.

Notably, TP 274 was a reported constituent (in $[M+C_2H_6N_2]+$ adduct form, with *m/z* 332.2212) of the previously reported "coho mortality chemical signature", comprising 57 chemical features co-detected during mortality events. (Peter et al., 2018) Thus, given the benefit of hindsight, the initial chemical signature approach focused on surrogate isolation was itself sufficient to resolve and prioritize a 6PPD ozonation product from thousands of chemicals detected in receiving waters where coho mortality occurred. In retrospect, TP 274 had already provided a direct link to the primary chemical source of coho mortality. (Peter et al., 2018) TP 274 was also reported by the EPA (US EPA, 2019) and by Klöckner et al. (Klöckner et al., 2021a) in both tire rubbers and road dust, suggesting its value as an abundant marker for tire rubber and roadway runoff.

TP 184 (*m/z* 185.1801 at 4.25 min, $C_{12}H_{12}N_2$, level 4 identification) was detected at the second highest abundance in roadway runoff. This TP was reported and annotated as 4-aminodiphenylamine by Seiwert et al. and Lattimer et al. as a side-chain oxidation product. (Lattimer et al., 1983; Seiwert et al., 2022) TP 198 (4-nitrosodiphenylamine, *m/z* 199.0852 at 7.2 min, $C_{12}H_{10}N_2O$, level 1 identification; **Figure 3.6**), which was also identified previously (Lattimer et al., 1983; Seiwert et al., 2022; Tian et al., 2021b), exhibited the third highest peak area abundance.

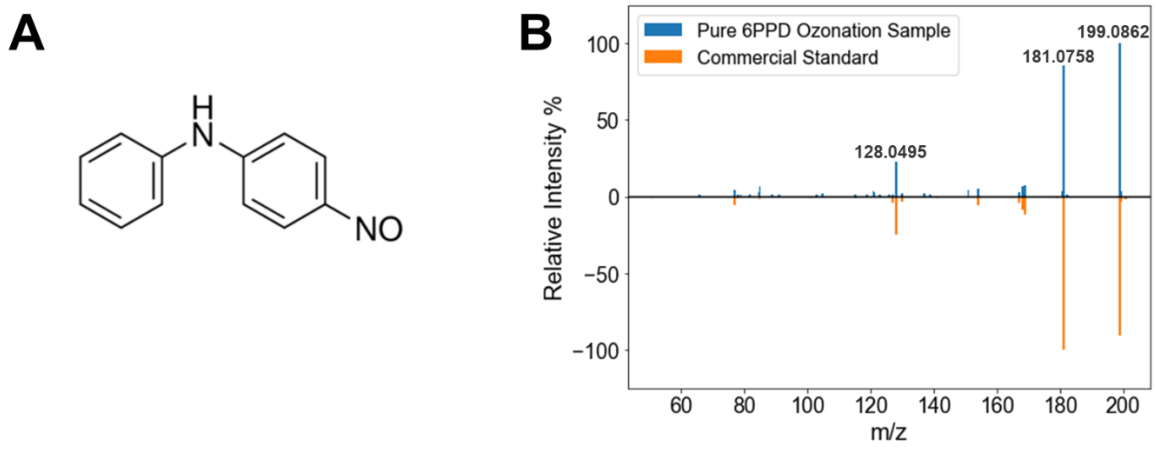


Figure 3.6. Chemical structure and MS/MS spectra of TP 198. (A) Chemical structure of TP198, which is identified as 4-nitrosodiphenylamine. (B) MS/MS spectra for TP198 from pure 6PPD ozonation sample (top) and from TP 198 commercial standard (bottom) from the average of different CID energies (10, 20, and 40 eV).

This TP was reportedly formed via 6PPD hydrolysis ("Registration Dossier - ECHA," n.d.), while Seiwert et al. suggested that it could also derive from hydrolysis of *N*-(4-(1,3-dimethylbutyl)imino)-2,5-cyclohexadiene-1-ylidene (QDI). (Seiwert et al., 2022) TP 198 formation during O_3 exposure could result either from side-chain oxidation or interactions of 6PPD or other TPs with water microlayers on the slides. Presence of a nitroso group and reported detections of 4-nitrosodiphenylamine in such contexts as crumb rubber playing fields suggest potential carcinogenicity and human exposure risks, respectively. (Drzyzga, 2003; Peter et al., 2020; Unice et al., 2015) Finally, TP 282 (*m/z* 283.1804 at 8.4 min, $C_{18}H_{22}N_2O$, level 4 identification) was reported in stormwater-impacted urban creeks and road dust. (Klöckner et al., 2021a; Peter et al., 2019) TPs with the same *m/z* were also reported in previous studies, with Lattimer et al. observing a single TP proposed as a nitrone of 6PPD, and more recent studies (Klöckner et al. and Seiwert et al.) reporting several isomeric TPs – one of which

was proposed as QDI-OH. (Klöckner et al., 2021a; Lattimer et al., 1983; Seiwert et al., 2022) Further investigations are needed to resolve structure(s) for TP(s) 282. All four TPs observed here exhibited formation dynamics similar to 6PPDQ during pure 6PPD ozonation (**Figures 3.7**, **3.8**) and elevated peak areas in ozonated TWPs (**Figure S3.4B**), suggesting multiple parallel TP formation pathways during 6PPD ozonation. Formation mechanisms, structural characterization, relative yields, and environmental occurrence of these unidentified TPs merit further evaluation.

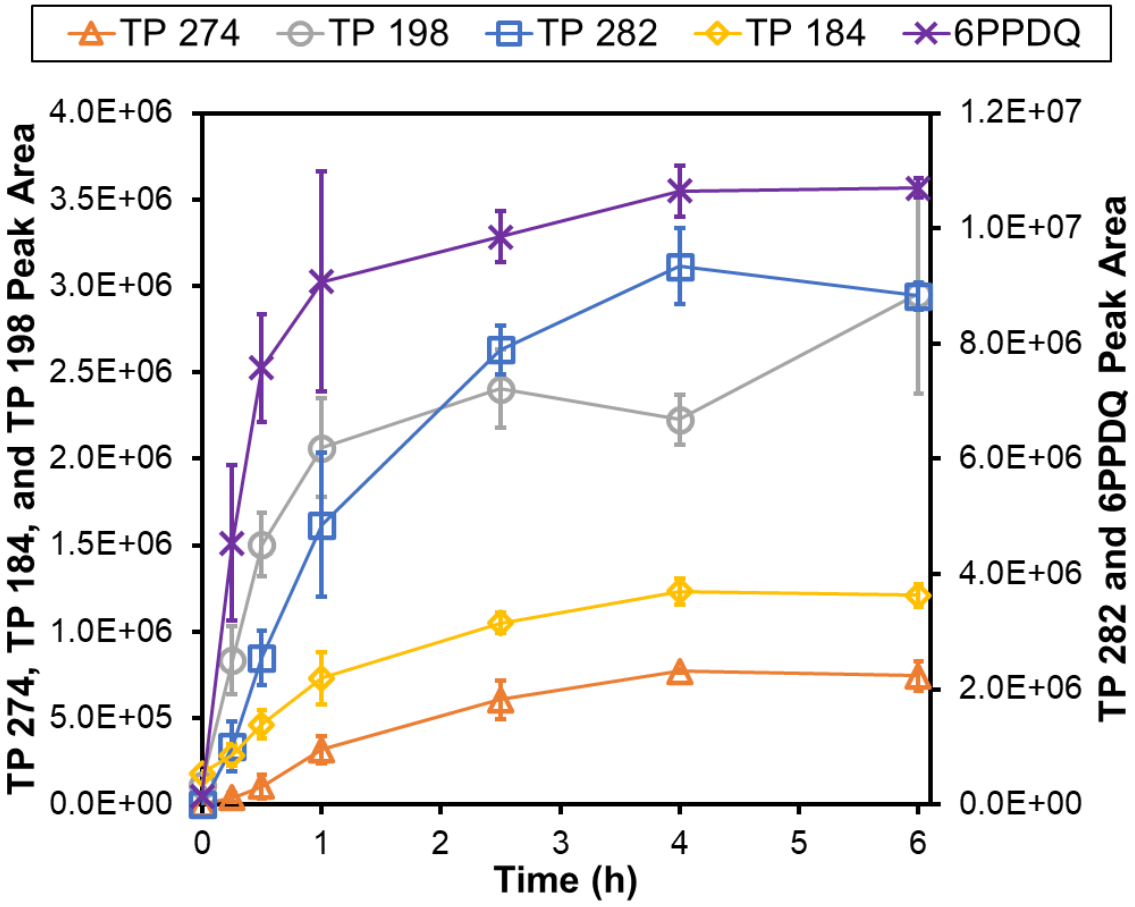


Figure 3.7. Representative time trends for peak areas of five key, stable 6PPD ozonation products generated during heterogeneous gas-phase ozonation of pure 6PPD on 6PPD-coated glass slides. Error bars correspond to the standard deviation from three replicate slides.

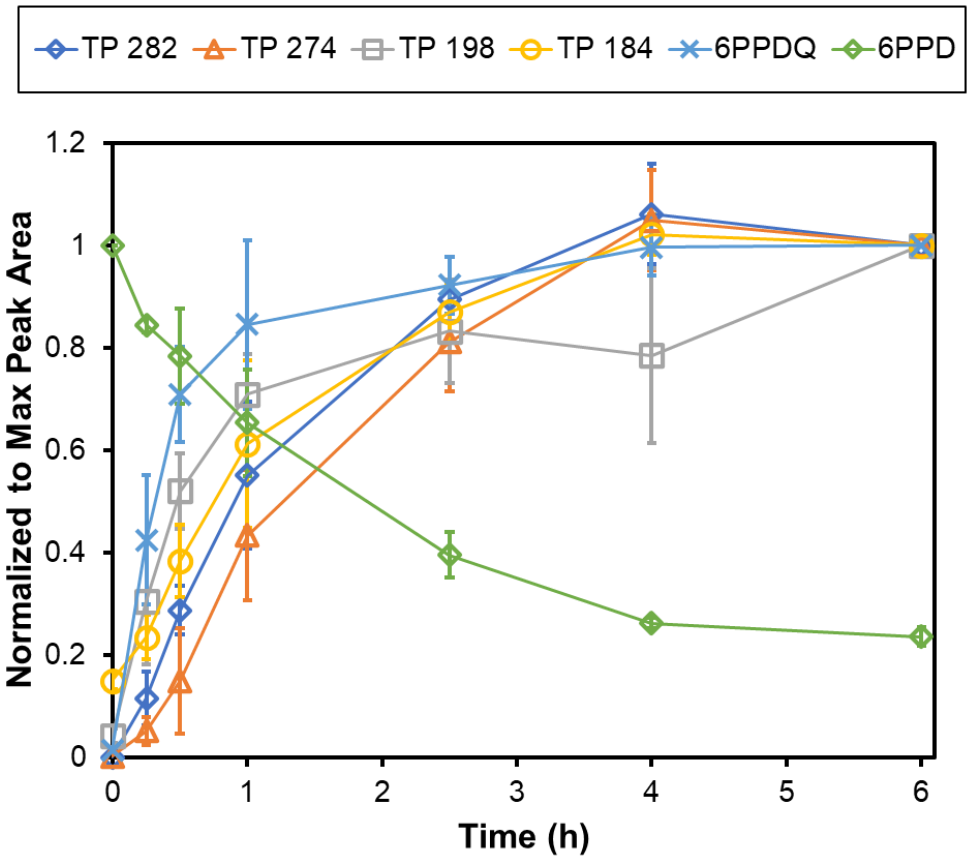


Figure 3.8. Normalized peak areas of 6PPD and 6PPD derived TPs for pure 6PPD ozonation experiments. All the peak areas were normalized to the maximum peak area for a given compound detected during the O_3 exposures. Error bars represent standard deviations of three slides.

3.4 Environmental Implications

Results from these heterogeneous gas-phase ozonation studies are relevant to 6PPD oxidation dynamics on rubber surfaces and consequent formation of the highly toxic TP 6PPDQ and several co-occurring TPs. Notably, 6PPD is designed to react with oxidants, especially ground-level O_3, so these reactions are primary drivers of 6PPD fate in TWPs and related TP formation. In our experimental setup, a primary molar yield of ~9.7% was observed for 6PPDQ formation from pure 6PPD and ~0.95% for 6PPD within TWP, suggesting that a

substantial mass fraction of 6PPD will ultimately be converted to 6PPDQ. Other less-characterized ozonation TPs were observed, indicating that mixtures of 6PPD-derived TPs are likely widely present in roadway-impacted air, soil, sediment, and water. (Huang et al., 2021; Klöckner et al., 2021a; Kole et al., 2017a, 2015) Given the long-lived character of rubbers, expected slow diffusion of chemicals from within rubber matrices and potential for surface O_3 reactions, other materials and systems using recycled tire rubber (e.g., crumb rubbers, building materials, rubber-modified asphalts) also are potentially long-term sources of such TPs. To the best of our knowledge, exposure risks and human health effects of TPs – including 6PPDQ – arising from residual rubbers remain largely uncharacterized. Furthermore, under complex environmental conditions, other antioxidants and reactive chemical additives in rubbers are expected to leach or contribute to the release of mobile TPs into air, water, soil, and sediments.

The current investigation utilized a relatively simple experimental design and screening approach over short time scales relative to tire rubber lifetimes. Further characterization of 6PPD reactivity, transformation mechanisms, and behavior within tire rubber and associated recycled rubber products will improve quantitative assessment of fate and transport potentials for 6PPD, other industrial chemical ingredients and associated TPs over product lifetimes, along with related environmental occurrence and toxicological characterization of TWP-derived contaminants.

Reference

Capolupo, M., Sørensen, L., Jayasena, K.D.R., Booth, A.M., Fabbri, E., 2020. Chemical composition and ecotoxicity of plastic and car tire rubber leachates to aquatic organisms. Water Res. 169, 115270. https://doi.org/10.1016/j.watres.2019.115270

Cataldo, F., Faucette, B., Huang, S., Ebenezer, W., 2015. On the early reaction stages of ozone with N,N′-substituted p-phenylenediamines (6PPD, 77PD) and N,N′,N″-substituted-1,3,5-triazine "Durazone®": An electron spin resonance (ESR) and electronic absorption spectroscopy study. Polym. Degrad. Stab. 111, 223–231. https://doi.org/10.1016/j.polymdegradstab.2014.11.011

ChemAxon - Software Solutions and Services for Chemistry & Biology [WWW Document], n.d. URL https://chemaxon.com/ (accessed 8.25.21).

Chen, S.-J., Su, H.-B., Chang, J.-E., Lee, W.-J., Huang, K.-L., Hsieh, L.-T., Huang, Y.-C., Lin, W.-Y., Lin, C.-C., 2007. Emissions of polycyclic aromatic hydrocarbons (PAHs) from the pyrolysis of scrap tires. Atmos. Environ. 41, 1209–1220. https://doi.org/10.1016/j.atmosenv.2006.09.041

Dohan, J.M., Masschelein, W.J., 1987. The Photochemical Generation of Ozone : Present State–of–the–Art. Ozone Sci. Eng. 9, 315–334. https://doi.org/10.1080/01919518708552147

Drzyzga, O., 2003. Diphenylamine and derivatives in the environment: a review. Chemosphere 53, 809–818. https://doi.org/10.1016/S0045-6535(03)00613-1

Du, B., Lofton, J.M., Peter, K.T., Gipe, A.D., James, C.A., McIntyre, J.K., Scholz, N.L., Baker, J.E., Kolodziej, E.P., 2017. Development of suspect and non-target screening methods for detection of organic contaminants in highway runoff and fish tissue with high-resolution time-of-flight mass spectrometry. Environ. Sci. Process. Impacts 19, 1185–1196. https://doi.org/10.1039/c7em00243b

Ester, M., Kriegel, H.-P., Xu, X., n.d. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise 6.

Gürtler, P., Saile, V., Koch, E.E., 1977. High resolution absorption spectrum of nitrogen in the vacuum ultraviolet. Chem. Phys. Lett. 48, 245–250. https://doi.org/10.1016/0009-2614(77)80308-4

Hiki, K., Asahina, K., Kato, K., Yamagishi, T., Omagari, R., Iwasaki, Y., Watanabe, H., Yamamoto, H., 2021. Acute Toxicity of a Tire Rubber-Derived Chemical, 6PPD Quinone, to Freshwater Fish and Crustacean Species. Environ. Sci. Technol. Lett. https://doi.org/10.1021/acs.estlett.1c00453

HILAL, S. H., n.d. Estimation of Physical Properties and Chemical Reactivity Parameters of Organic Compounds for Environmental Modeling by SPARC [WWW Document]. URL https://cfpub.epa.gov/si/si_public_record_report.cfm?Lab=NERL&dirEntryId=203108 (accessed 2.26.22).

Huang, D., LaCount, B.J., Castro, J.M., Ignatz-Hoover, F., 2001. Development of a service-simulating, accelerated aging test method for exterior tire rubber compounds I. Cyclic

aging. Polym. Degrad. Stab. 74, 353–362. https://doi.org/10.1016/S0141-3910(01)00185-9

Huang, W., Shi, Y., Huang, J., Deng, C., Tang, S., Liu, X., Chen, D., 2021. Occurrence of Substituted p-Phenylenediamine Antioxidants in Dusts. Environ. Sci. Technol. Lett. https://doi.org/10.1021/acs.estlett.1c00148

Huber, M., Welker, A., Helmreich, B., 2016. Critical review of heavy metal pollution of traffic area runoff: Occurrence, influencing factors, and partitioning. Sci. Total Environ. 541, 895–919. https://doi.org/10.1016/j.scitotenv.2015.09.033

Huntink, N.M., 2003. Durability of rubber products: Development of new antidegradants for long-term protection.

Järlskog, I., Strömvall, A.-M., Magnusson, K., Galfi, H., Björklund, K., Polukarova, M., Garção, R., Markiewicz, A., Aronsson, M., Gustafsson, M., Norin, M., Blom, L., Andersson-Sköld, Y., 2021. Traffic-related microplastic particles, metals, and organic pollutants in an urban area under reconstruction. Sci. Total Environ. 774, 145503. https://doi.org/10.1016/j.scitotenv.2021.145503

Klöckner, P., Seiwert, B., Wagner, S., Reemtsma, T., 2021a. Organic Markers of Tire and Road Wear Particles in Sediments and Soils: Transformation Products of Major Antiozonants as Promising Candidates. Environ. Sci. Technol. https://doi.org/10.1021/acs.est.1c02723

Klöckner, P., Seiwert, B., Weyrauch, S., Escher, B.I., Reemtsma, T., Wagner, S., 2021b. Comprehensive characterization of tire and road wear particles in highway tunnel road dust by use of size and density fractionation. Chemosphere 279, 130530. https://doi.org/10.1016/j.chemosphere.2021.130530

Kole, P.J., Löhr, A.J., Ragas, A.M.J., 2015. Autobandenslijtstof: een verwaarloosde bron van microplastics? Milieu Opinieblad Van Ver. Van Milieuprofessionals 39–41.

Kole, P.J., Löhr, A.J., Van Belleghem, F.G.A.J., Ragas, A.M.J., 2017a. Wear and Tear of Tyres: A Stealthy Source of Microplastics in the Environment. Int. J. Environ. Res. Public. Health 14. https://doi.org/10.3390/ijerph14101265

Kole, P.J., Löhr, A.J., Van Belleghem, F.G.A.J., Ragas, A.M.J., 2017b. Wear and Tear of Tyres: A Stealthy Source of Microplastics in the Environment. Int. J. Environ. Res. Public. Health 14. https://doi.org/10.3390/ijerph14101265

Lattimer, R.P., Hooser, E.R., Layer, R.W., Rhee, C.K., 1983. Mechanisms of Ozonation of N-(1,3-Dimethylbutyl)-N′-Phenyl-p-Phenylenediamine. Rubber Chem. Technol. 56, 431–439. https://doi.org/10.5254/1.3538136

Li, X., Heays, A.N., Visser, R., Ubachs, W., Lewis, B.R., Gibson, S.T., Dishoeck, E.F. van, 2013. Photodissociation of interstellar N2. Astron. Astrophys. 555, A14. https://doi.org/10.1051/0004-6361/201220625

Lofthus, A., Krupenie, P.H., 1977. The spectrum of molecular nitrogen. J. Phys. Chem. Ref. Data 6, 113–307. https://doi.org/10.1063/1.555546

Müller, K., Hübner, D., Huppertsberg, S., Knepper, T.P., Zahn, D., 2022. Probing the chemical complexity of tires: Identification of potential tire-borne water contaminants with high-

resolution mass spectrometry. Sci. Total Environ. 802, 149799. https://doi.org/10.1016/j.scitotenv.2021.149799

OSPAR Commission, 2006c. OSPAR background document on 4-(dimethylbutylamino)diphenylamine (6PPD).

Patiny, L., Borel, A., 2013. ChemCalc: A Building Block for Tomorrow's Chemical Infrastructure. J. Chem. Inf. Model. 53, 1223–1228. https://doi.org/10.1021/ci300563h

Peter, K.T., Herzog, S., Tian, Z., Wu, C., McCray, J.E., Lynch, K., Kolodziej, E.P., 2019. Evaluating emerging organic contaminant removal in an engineered hyporheic zone using high resolution mass spectrometry. Water Res. 150, 140–152. https://doi.org/10.1016/j.watres.2018.11.050

Peter, K.T., Hou, F., Tian, Z., Wu, C., Goehring, M., Liu, F., Kolodziej, E.P., 2020. More Than a First Flush: Urban Creek Storm Hydrographs Demonstrate Broad Contaminant Pollutographs. Environ. Sci. Technol. 54, 6152–6165. https://doi.org/10.1021/acs.est.0c00872

Peter, K.T., Tian, Z., Wu, C., Lin, P., White, S., Du, B., McIntyre, J.K., Scholz, N.L., Kolodziej, E.P., 2018. Using High-Resolution Mass Spectrometry to Identify Organic Contaminants Linked to Urban Stormwater Mortality Syndrome in Coho Salmon. Environ. Sci. Technol. 52, 10317–10327. https://doi.org/10.1021/acs.est.8b03287

Rapta, P., Vargová, A., Polovková, J., Gatial, A., Omelka, L., Majzlík, P., Breza, M., 2009. A variety of oxidation products of antioxidants based on N,N′-substituted p-phenylenediamines. Polym. Degrad. Stab. 94, 1457–1466. https://doi.org/10.1016/j.polymdegradstab.2009.05.003

Registration Dossier - ECHA [WWW Document], n.d. URL https://echa.europa.eu/de/registration-dossier/-/registered-dossier/15367/5/6/1 (accessed 2.7.22).

Rhodes, E.P., Ren, Z., Mays, D.C., 2012. Zinc leaching from tire crumb rubber. Environ. Sci. Technol. 46, 12856–12863. https://doi.org/10.1021/es3024379

Rochman, C.M., Hoh, E., Kurobe, T., Teh, S.J., 2013. Ingested plastic transfers hazardous chemicals to fish and induces hepatic stress. Sci. Rep. 3, 3263. https://doi.org/10.1038/srep03263

Sadiktsis, I., Bergvall, C., Johansson, C., Westerholm, R., 2012. Automobile tires--a potential source of highly carcinogenic dibenzopyrenes to the environment. Environ. Sci. Technol. 46, 3326–3334. https://doi.org/10.1021/es204257d

Saini, R., Singh, P., Awasthi, B.B., Kumar, K., Taneja, A., 2014. Ozone distributions and urban air quality during summer in Agra – a world heritage site. Atmospheric Pollut. Res. 5, 796–804. https://doi.org/10.5094/APR.2014.089

Schymanski, E.L., Jeon, J., Gulde, R., Fenner, K., Ruff, M., Singer, H.P., Hollender, J., 2014. Identifying Small Molecules via High Resolution Mass Spectrometry: Communicating Confidence. Environ. Sci. Technol. 48, 2097–2098. https://doi.org/10.1021/es5002105

Seiwert, B., Klöckner, P., Wagner, S., Reemtsma, T., 2020. Source-related smart suspect screening in the aqueous environment: search for tire-derived persistent and mobile

trace organic contaminants in surface waters. Anal. Bioanal. Chem. 412, 4909–4919. https://doi.org/10.1007/s00216-020-02653-1

Seiwert, B., Nihemaiti, M., Troussier, M., Weyrauch, S., Reemtsma, T., 2022. Abiotic oxidative transformation of 6-PPD and 6-PPD quinone from tires and occurrence of their products in snow from urban roads and in municipal wastewater. Water Res. 212, 118122. https://doi.org/10.1016/j.watres.2022.118122

Seung Lim J, 2016. EPI Suite: A Fascinate Predictive Tool for Estimating the Fates of Organic Contaminants | OMICS International [WWW Document]. URL https://www.omicsonline.org/open-access/epi-suite-a-fascinate-predictive-tool-for-estimating-the-fates-of-organiccontaminants-2155-6199-1000e171.php?aid=70728 (accessed 8.25.21).

Sheridan, M., 2010. The Vanderbilt Rubber Handbook, 14th edition. R.T. Vanderbilt Company, Inc., Norwalk, CT.

Simon, H., Reff, A., Wells, B., Xing, J., Frank, N., 2015. Ozone Trends Across the United States over a Period of Decreasing NOx and VOC Emissions. Environ. Sci. Technol. 49, 186–195. https://doi.org/10.1021/es504514z

Tian, Z., Gonzalez, M., Rideout, C., Zhao, H., Hu, X., Wetzel, J., Mudrock, E., James, C.A., McIntyre, J.K., Kolodziej, E.P., 2021a. 6PPD-Quinone: Revised Toxicity Assessment and Quantification Method Development with a Commercial Standard. Environ. Sci. Technol. Lett.

Tian, Z., Peter, K.T., Gipe, A.D., Zhao, H., Hou, F., Wark, D.A., Khangaonkar, T., Kolodziej, E.P., James, C.A., 2020. Suspect and Nontarget Screening for Contaminants of Emerging Concern in an Urban Estuary. Environ. Sci. Technol. 54, 889–901. https://doi.org/10.1021/acs.est.9b06126

Tian, Z., Zhao, H., Peter, K.T., Gonzalez, M., Wetzel, J., Wu, C., Hu, X., Prat, J., Mudrock, E., Hettinger, R., Cortina, A.E., Biswas, R.G., Kock, F.V.C., Soong, R., Jenne, A., Du, B., Hou, F., He, H., Lundeen, R., Gilbreath, A., Sutton, R., Scholz, N.L., Davis, J.W., Dodd, M.C., Simpson, A., McIntyre, J.K., Kolodziej, E.P., 2021b. A ubiquitous tire rubber–derived chemical induces acute mortality in coho salmon. Science 371, 185–189. https://doi.org/10.1126/science.abd6951

Tsugawa, H., Cajka, T., Kind, T., Ma, Y., Higgins, B., Ikeda, K., Kanazawa, M., VanderGheynst, J., Fiehn, O., Arita, M., 2015. MS-DIAL: Data Independent MS/MS Deconvolution for Comprehensive Metabolome Analysis. Nat. Methods 12, 523–526. https://doi.org/10.1038/nmeth.3393

Unice, K.M., Bare, J.L., Kreider, M.L., Panko, J.M., 2015. Experimental methodology for assessing the environmental fate of organic chemicals in polymer matrices using column leaching studies and OECD 308 water/sediment systems: Application to tire and road wear particles. Sci. Total Environ. 533, 476–487. https://doi.org/10.1016/j.scitotenv.2015.06.053

Unice, K.M., Weeber, M.P., Abramson, M.M., Reid, R.C.D., van Gils, J. a. G., Markus, A.A., Vethaak, A.D., Panko, J.M., 2019. Characterizing export of land-based microplastics to the estuary - Part I: Application of integrated geospatial microplastic transport

models to assess tire and road wear particles in the Seine watershed. Sci. Total Environ. 646, 1639–1649. https://doi.org/10.1016/j.scitotenv.2018.07.368

US EPA, O., 2019. July 2019 Report: Tire Crumb Rubber Characterization [WWW Document]. US EPA. URL https://www.epa.gov/chemical-research/july-2019-report-tire-crumb-rubber-characterization-0 (accessed 5.17.21).

US EPA, O., 2014. Air Data: Air Quality Data Collected at Outdoor Monitors Across the US [WWW Document]. US EPA. URL https://www.epa.gov/outdoor-air-quality-data (accessed 6.14.21).

US EPA, R. 01, n.d. Eight-hour Average Ozone Concentrations | Ground-level Ozone | New England | US EPA [WWW Document]. URL https://www3.epa.gov/region1/airquality/avg8hr.html (accessed 5.17.21).

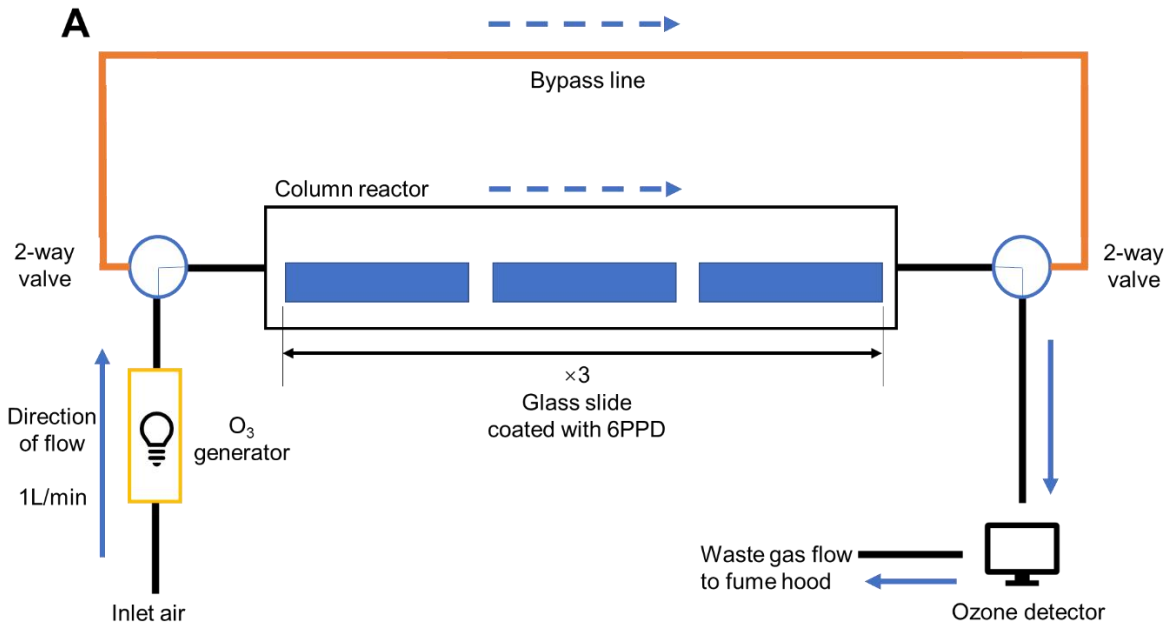
Wagner, S., Hüffer, T., Klöckner, P., Wehrhahn, M., Hofmann, T., Reemtsma, T., 2018. Tire wear particles in the aquatic environment - A review on generation, analysis, occurrence, fate and effects. Water Res. 139, 83–100. https://doi.org/10.1016/j.watres.2018.03.051

Wu, Y., Venier, M., Hites, R.A., 2020. Broad Exposure of the North American Environment to Phenolic and Amino Antioxidants and to Ultraviolet Filters. Environ. Sci. Technol. 54, 9345–9355. https://doi.org/10.1021/acs.est.0c04114

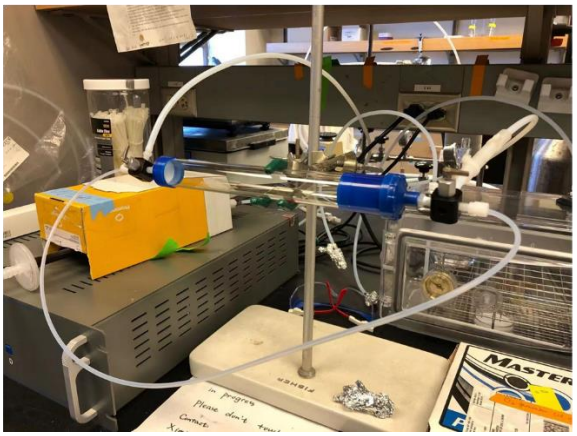
Zhang, Y., Xu, C., Zhang, W., Qi, Z., Song, Y., Zhu, L., Dong, C., Chen, J., Cai, Z., 2021. p-Phenylenediamine Antioxidants in PM2.5: The Underestimated Urban Air Pollutants. Environ. Sci. Technol. https://doi.org/10.1021/acs.est.1c04500

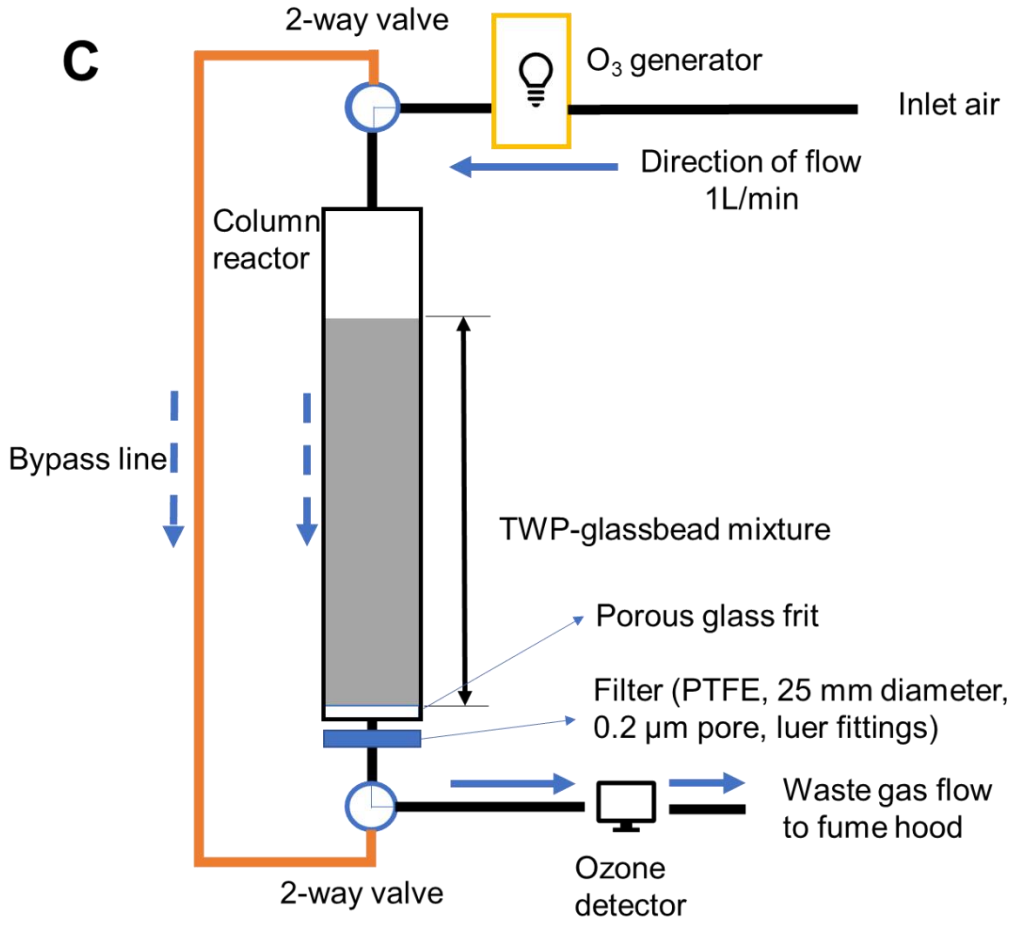
Zhou, Z., Abbatt, J.P.D., 2021. Formation of Gas-Phase Hydrogen Peroxide via Multiphase Ozonolysis of Unsaturated Lipids. Environ. Sci. Technol. Lett. 8, 114–120. https://doi.org/10.1021/acs.estlett.0c00757

Supplementary Materials



B





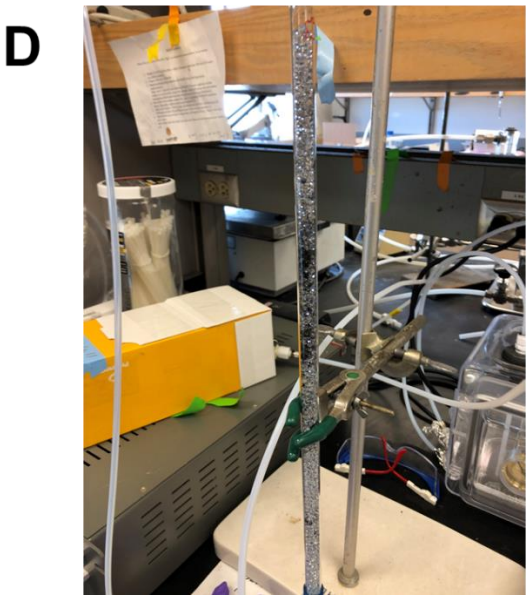


Figure S3.1. Schematic diagrams of experimental system setups. (A) Pure 6PPD-coated slide ozonation setup (B) Photo of the pure 6PPD-coated slide ozonation column and associated valve and tubing connections (C) TWP and glass bead mixture ozonation setup and (D) Photo of the packed TWP/glass bead ozonation column. All tubing in direct contact with O_3 was made from PTFE. The chromatography columns were constructed of borosilicate glass with polypropylene inlets and outlets.

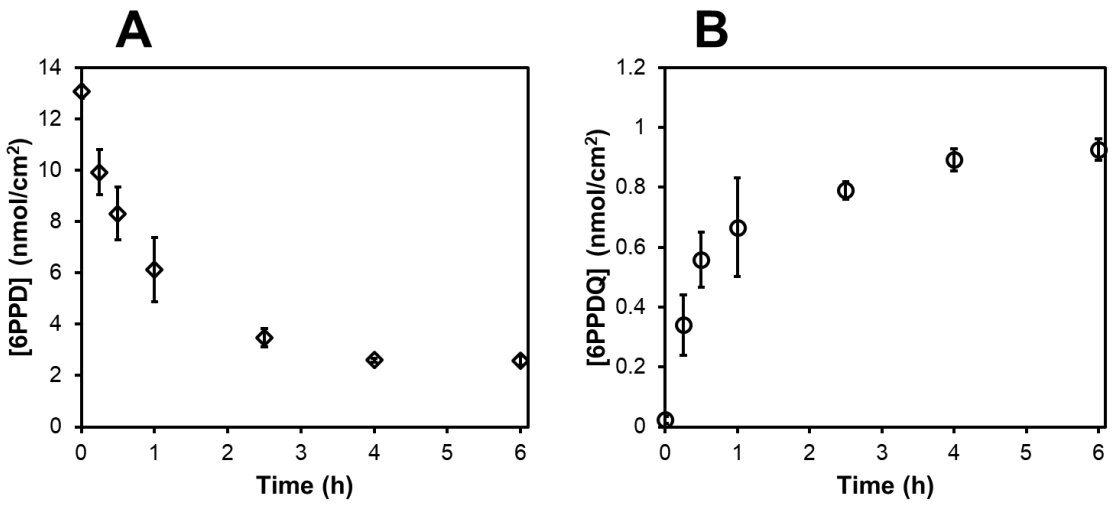


Figure S3.2. 6PPD and 6PPDQ surface area-normalized concentrations vs. time during pure 6PPD O_3 exposures. (A) 6PPD and (B) 6PPDQ. Error bars in both panels represent standard deviations of three replicate slides.

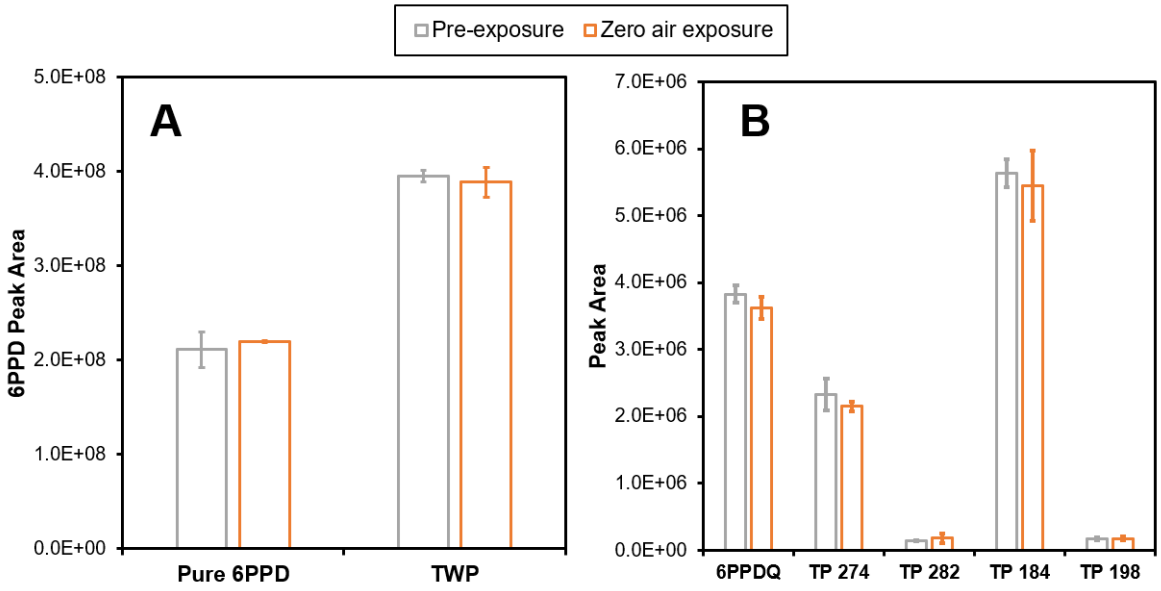


Figure S3.3. 6PPD and 6PPD TP peak area comparison during 6-hour zero grade air exposure experiments. (A) 6PPD peak areas for pure 6PPD and TWP samples prior to and after zero grade air exposures. No significant degradation of 6PPD was observed (p-value = 0.08 and 0.47 for pure 6PPD and TWP samples, respectively). (B) Major TP peak areas for TWP samples prior to and after zero grade air exposures. No significant differences were observed between pre- and post-exposure samples (p-value > 0.05). The results showed the stability of 6PPD to molecular oxygen under tested conditions, though this does not necessarily preclude a role of oxygen over longer periods (months to years) of exposure relevant to tire-service life.

Error bars in both panels represent standard deviations from triplicate samples. The samples were stored at -20 °C for 2 months prior to analysis. 6PPD was confirmed to be stable over the storage period (two-tailed t-test, day 0 vs day 60; p-value = 0.99). All data in Figure S3.3 were obtained from analyses undertaken within a single instrument analytical batch (lasting approximately 12 hours). The differences in magnitudes of analyte peak areas for the sample sets presented in Figure S3.3, compared to the analyte peak areas for the sample sets presented in Figure S3.4, likely resulted primarily from variations in instrument response arising from routine instrumental maintenance and repairs undertaken over the several month time span separating the Figure S3.3 and S3.4 sample analysis batches.

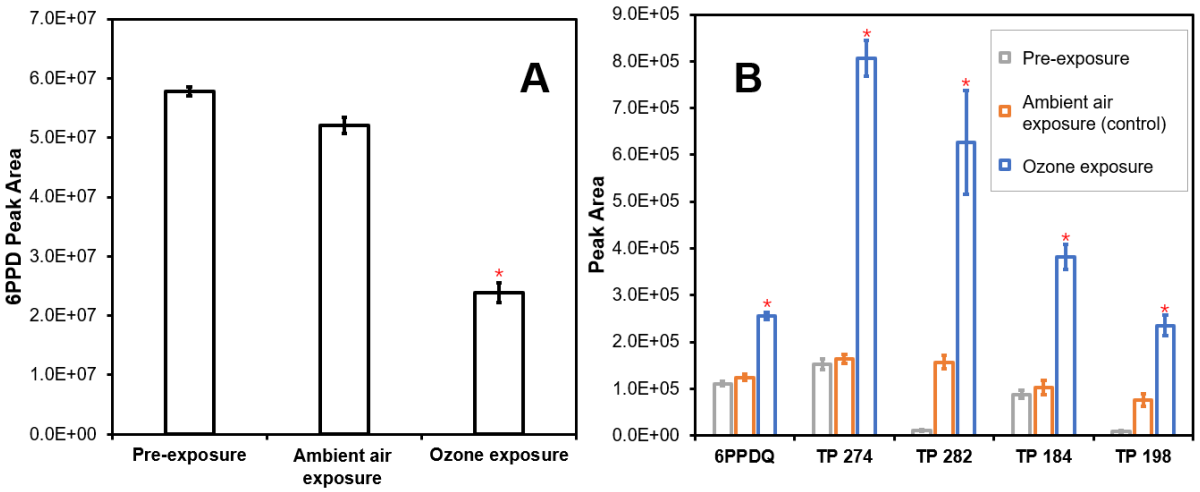


Figure S3.4. 6PPD and 6PPD TP peak area comparison during exposure of TWPs to O_3 and ambient air for 6 hours. The asterisks above the O_3 exposure samples indicate statistically significant changes in TP peak area (two-tailed t-test; $p < 0.05$) compared to the pre-exposure samples and ambient air exposure (control) samples. (A) 6PPD peak areas for TWP samples prior to and after ambient air and O_3 exposures. The peak area decrease observed in the ambient air exposure (control) sample likely resulted from reactions with background levels of ground-level O_3 and/or other gas-phase oxidant species. (B) Major TP peak areas for TWP samples prior to and after ambient air and O_3 exposures. All data in Figure S3.4 were obtained from analyses undertaken within a single instrument analytical batch (lasting approximately 12 hours). The differences in magnitudes of analyte peak areas for the sample sets presented in Figure S3.4, compared to the analyte peak areas for the sample sets presented in Figure S3.3, likely resulted primarily from variations in instrument response arising from routine instrumental maintenance and repairs undertaken over the several month time span separating the Figure S3.3 and S3.4 sample analysis batches.

Table S3.1. All prioritized TPs from pure 6PPD and TWP ozonation.

Note: All the peak areas were averaged from triplicate measurements.

		Feature Information						Observed Peak Areas												
								Pure 6PPD-Ozonation Samples							TWP-Ozonation Samples		TWP aqueous leachate	Roadway Runoff			
Index Number	Chemical Name	Average RT (min)	Average *m/z*	Annotated Formula (if assigned)	Identification level*	Average mass error (ppm)	Adduct	Detection**	0h	0.25h	0.5h	1h	2.5h	4h	6h	TWP-before ozonation	TWP-after ozonation		Runoff Site2-Feb2019	Runoff Site2-Oct2018	Runoff Site1-Jun2018	Runoff Site1-April2018
1		7.35	124.0399	C6H5NO2	4	0.23	[M+H]+	T	0	126686	291649	567654	558618	520444	529986	0	79587	0	108891	0	0	0
2		7.36	152.0711		5		[M+Na]+	T	0	199875	455250	905221	877035	814698	818673	0	212657	0	154385	0	0	0
3		10.38	182.1170	C8H13N4O	4	-0.91	[M+H]+	A	0	0	0	79669	140110	177621	166221	35816	174401	2078180	2030881	3232120	2703990	3704672
4	TP 184	4.25	185.1081	C12H12N2	3	-0.77	[M+H]+	A	540777	854501	1388006	2190125	3149483	3703610	3627729	88313	381888	538037	6477371	5257868	4159811	14782204
5		11.01	187.0859	C8H12NO4	4	-7.08	[M+H]+	P	0	135519	198711	230823	253518	270127	268171	0	0	0	0	0	0	0
6		5.00	196.1325	C9H15N4O	4	0.19	[M+H]+	P	0	88799	219255	474674	488027	506774	1074709	0	0	0	0	0	0	0
7	p-Nitrosodiphen	7.22	199.0852	C12H10N2O	1	-9.73	[M+H]+	A	114968	834047	1503809	2062724	2404868	2227620	2954775	0	235054	661562	9098663	7239276	3502171	12718794
8		4.89	202.0479	C9H5N4O2	4	6.02	[M+H]+	P	0	0	0	0	125866	149009	173930	0	0	0	0	0	0	0
9		4.96	206.0806	C9H9N4O2	4	-0.73	[M+H]+	P	0	0	0	58123	103988	107392	136267	0	0	0	0	0	0	0
10		7.36	208.1332	C12H17NO2	4	3.23	[M+H]+	P	0	468947	1032962	2057812	2000547	1876761	1914305	0	0	203170	323279	112243	0	0
11		5.04	208.1334	C12H17NO2	4	1.88	[M+H]+	T	0	0	0	67512	148618	230288	280123	0	418777	0	0	0	0	0
12		6.08	210.1104	C6H15N3O5	4	-6.36	[M+H]+	T	0	0	0	118911	249460	252139	300315	0	86525	0	0	0	0	0
13		8.45	210.1264	C12H17O3	4	-3.35	[M+H]+	P	0	0	61839	107798	169023	199523	187285	0	0	0	0	0	0	138305
14		6.15	213.1025	C13H12N2O	4	1.62	[M+H]+	A	93491	86526	105560	130989	152688	159024	164603	254519	216268	5423353	2646969	3229045	877775	2881966
15		11.01	215.0811	C12H10N2O2	4	4.94	[M+H]+	P	0	229866	360570	437105	475375	512683	517294	0	0	106067	0	0	0	0
16		9.84	216.0653	C10H7N4O2	4	-2.28	[M+H]+	P	0	0	0	77851	109116	106853	119469	0	0	0	0	0	0	0
17		5.99	224.1276	C10H15N4O2	4	-0.85	[M+H]+	P	0	57130	103968	201877	235198	223080	216902	0	0	0	0	0	0	0
18		8.46	224.1306	C15H15N2	4	3.78	[M+H]+	P	0	90216	174462	270091	407592	472685	449113	0	0	0	0	0	0	296483
19		4.64	224.1634	C11H19N4O	4	1.78	[M+H]+	P	0	0	0	91852	115147	118185	175630	0	0	0	0	0	0	0
20		7.57	237.6569		5		[M+H]+	P	0	0	0	0	87585	102020	100025	0	0	0	0	0	0	0
21		9.52	257.1653	C16H20N2O	4	0.60	[M+H]+	A	0	51871	131554	405140	722166	845952	1187908	0	145125	1122075	847483	558190	254404	2864550
22		11.77	257.1654	C16H20N2O	4	0.36	[M+H]+	T	0	0	0	71358	128308	143328	109877	0	599518	0	0	0	0	0
23		8.46	266.1776	C18H21N2	4	2.77	[M+H]+	T	0	209537	494092	923405	1462549	1724203	1625169	0	60104	0	165174	142072	135040	976304
24		11.24	271.1794	C14H24NO4	4	-3.73	[M+H]+	P	0	0	0	79007	85942	89940	190579	0	0	188907	0	0	0	0
25	TP 274	10.38	275.1764	C16H22N2O2	3	-1.21	[M+H]+	A	0	0	105456	318155	605556	770784	742904	152493	806651	9423247	9014041	13628045	11173567	15503969
26		5.79	281.1644	C18H20N2O	4	3.80	[M+H]+	T	70863	656566	906398	815172	1098773	1087313	1146688	0	64783	0	114897	0	0	0
27	TP 282	8.46	283.1808	C18H22N2O	3	1.04	[M+H]+	A	0	1005856	2544723	4852881	7893370	9337408	8823217	0	627256	127703	833130	710638	654071	5059171
28		9.42	291.1703	C16H22N2O3	4	2.25	[M+H]+	T	0	0	0	102467	205509	244367	225724	0	62999	0	317837	340737	360938	660611
29		13.34	295.1451	C18H18N2O2	4	-1.40	[M+H]+	P	0	0	0	89622	96028	92023	111093	0	0	0	0	0	0	0
30		7.42	297.1606	C18H20N2O2	4	-0.78	[M+H]+	P	0	128844	163867	175226	199773	226920	185158	0	0	0	112187	179209	0	196753
31		11.01	298.1675	C18H21N2O2	4	2.43	[M+H]+	P	0	0	75748	90598	100758	107311	108156	0	0	0	0	0	0	0
32		9.41	299.1744	C15H24NO5	4	-3.56	[M+H]+	P	0	82018	173324	324720	321019	356905	434823	0	0	200612	373292	365180	191097	353676
33		8.18	299.1746	C15H24NO5	4	-4.07	[M+H]+	P	0	58669	93962	149611	160655	187542	170524	0	0	410767	241245	249079	0	392721
34		5.31	299.1754	C18H22N2O2	4	2.17	[M+H]+	P	0	348803	884844	1699938	1461970	1605923	834281	0	0	0	0	0	0	0
35	6PPDQ	11.01	299.1755	C18H22N2O2	1	1.84	[M+H]+	A	148291	4536851	7576168	9075189	9854581	10646981	10699202	111028	256194	2152502	1056790	818986	351146	674812
36		7.08	303.1697	C17H22N2O3	4	4.08	[M+H]+	P	0	0	0	88961	106644	106928	119525	0	0	125793	0	114187	0	320133
37		9.88	311.1751	C19H22N2O2	4	3.15	[M+H]+	P	0	0	0	0	190957	133459	292795	0	0	0	0	0	0	0
38		5.58	313.1907	C19H24N2O2	4	3.13	[M+H]+	T	80955	380109	511710	923972	784001	756031	924481	59586	256416	235296	0	0	0	0
39		7.14	317.1860	C18H24N2O3	4	1.97	[M+H]+	T	0	347692	1041694	2450873	2382699	2671618	1541024	0	261977	0	0	0	0	0
40		10.83	333.1804	C18H24N2O4	4	3.27	[M+H]+	P	0	0	73094	223987	199818	172103	131636	0	0	0	0	0	0	0
41		15.06	374.2242	C26H29O2	4	1.19	[M+H]+	A	0	0	81471	164612	179331	201609	204078	85072	704001	103388	1616625	604496	136013	152911
42		11.01	383.1274	$C_{21}H_{20}N_4O$	4	-0.90	[M+K]+	P	0	83188	129479	139093	155880	156234	156624	0	0	0	0	0	0	0
43		6.55	479.2441	C30H30N4O2	4	1.52	[M+H]+	P	0	0	0	0	61416	88749	101341	0	0	0	0	0	0	0
44		15.64	547.3432	C36H42N4O	4	1.03	[M+H]+	P	0	53642	58076	90939	100450	107207	121186	0	0	0	0	0	0	0
45		10.87	547.3438	C36H42N4O	4	0.02	[M+H]+	P	0	452502	630738	650083	696370	798722	751234	0	0	0	0	251135	0	104836
46		11.01	651.2623	C40H34N4O5	4	-2.23	[M+H]+	P	0	0	53348	73262	88785	98596	102138	0	0	0	0	0	0	0

*Identification level following Schymanski et al: Level 1 -- Confirmed Structure by reference standard; Level 3 -- Tentative candidates; Level 4 -- Unequivocal molecular formula; Level 5 -- Exact mass of interest

**Detection: A -- Detected in all samples (Pure 6PPD ozonation, TWP ozonation, TWP water leachate and environmental runoff samples); T -- Only detected in ozonated pure 6PPD and ozonated TWP samples; P -- Only detected in ozonated pure 6PPD samples

Table S3.2. Information about the nine tires used to produce tire tread wear particles (TWPs).(Tian et al., 2021b) The TWP mixture used herein was an equal-weight mixture of the nine tires.

Tire ID	Manufacturer	Season	New/Used	Vehicle Type	Tire code
1	A	All-season	Used	Car	205/55R16 91H
2	B	All-season	Used	Car	195/70R14 91H
3	C	All-season	Used	Car	P185/65R15 86T
4	D	Winter	Used	Car	175/65R14 82S
5	E	All-season	Used	Car	P185/85R15 86S
6	E	All-season	Used	Light Truck	LT235/85R16 120R
7	F	All-season	New	Car	P205/60R15 90T
8	G	All-season	New	Car	P205/60R16 92H
9	H	All-season	Used	Car	P225/60R17 98T

Chapter 4: Environmental fate of 6PPD and 6PPD-TPs under gas phase and aqueous phase oxidative exposures

4.1 Introduction

Chemical detection methodologies and technologies have substantially advanced over the past decades, including the development of sensitive and powerful mass spectrometry and non-targeted analysis methods (Blum et al., 2017; Dom et al., 2018; Noguera-Oviedo and Aga, 2016). Given these improving capabilities, "emerging" organic contaminants have been receiving more attention as recognition of their pervasive occurrence and potential biological risks grows (Hou et al., 2019; Peter et al., 2018). Stormwater-related emerging organic contaminants also are becoming increasingly recognized as potential hazards, especially as large point sources like wastewater treatment plant effluent have become better controlled. Stormwater runoff generated by both extreme and routine weather events can release and transport chemicals from various anthropogenic sources, eventually impacting water quality or environmental health (Beeson, n.d.; Lin et al., 2020; Spahr et al., 2019). Also, emerging contaminants beyond well-understood classes like pharmaceuticals remain poorly studied, with limited information regarding their environmental fate, toxicity, and risk. Therefore, increasing research effort is being directed toward the identities and environmental fates of stormwater-related emerging contaminants (Carpenter et al., 2019; Challis et al., 2021; Peter et al., 2020; Spahr et al., 2019).

Recently, we reported the discovery of a novel contaminant, 6PPD-quinone (6PPDQ), that is discharged into surface water during storm events and induces acute coho salmon mortality (Tian et al., 2021b). Subsequent research confirmed that 6PPDQ is derived from a tire rubber antioxidant, 6PPD (N-(1,3-dimethylbutyl)-N'-phenyl-p-phenylenediamine), that is widely used in tire rubber products (0.4-2% by mass (Sheridan, 2010)). 6PPD is transformed into 6PPDQ via oxidative reactions, especially gas phase reactions with ozone, which is then deposited into the environment via leaching from wet tires and tire wear particles (TWP) (Councell et al., 2004; Klöckner et al., 2021b; OSPAR Commission, 2006c; Seiwert et al., 2022; Tian et al., 2021a; Wagner et al., 2018). The PPD family of antioxidants and their analogs (e.g., 7PPD, IPPD, DPPD and their quinone products) have now been reported as present world-wide in surface water, soil, air particles and human blood and urine (Cao et al., 2022; Du et al., 2022; Johannessen et al., 2021; Seiwert et al., 2022; Wang et al., 2022; Zhang et al., 2021). Beyond TWP, which is regarded as the major source of 6PPD and 6PPDQ release, Zhao et al. recently reported the presence of PPD chemicals in crumb rubber and consumer products comprised of recycled and remanufactured tire materials (Zhao et al., 2023a), indicating a wider array of sources for these rubber-derived contaminants. Compounding the problem of the adverse environmental effect of 6PPDQ, during the service life of tires, crumb rubbers, and other rubber consumer products and after release of 6PPD into the environment, 6PPD can transform into various TPs (in addition to 6PPDQ) under a range of environmental conditions (e.g., light, oxidative, thermal, and aqueous). Many of these TPs remain uncharacterized or rarely reported (Hu et al., 2022; Zhao et al., 2023b; Klöckner et al., 2021a). Despite the ubiquitous presence

of these chemicals in roadway environments, there is scarce information available regarding their environmental fate.

Given the global presence and potential environmental and health impacts of 6PPD and related TPs, particularly 6PPDQ, there is a growing need to better understand the fate of this antioxidant. This includes investigating the transformation mechanisms and dynamics of 6PPD as well as the formation of TPs under varied environmental conditions. In this study, we conducted a comprehensive analysis of 6PPD environmental fate, including: 1) ozonation of 6PPD under gaseous and aqueous conditions; 2) long-term stability of 6PPD under oxygen exposure without ozone; and 3) short-term aqueous stability of 6PPD under ambient conditions. The results provide important insights into the potential risks associated with environmentally dispersed tire rubbers, 6PPD, and 6PPD byproducts to inform strategies for mitigation.

4.2 Materials and Methods

4.2.1 Chemicals

6PPD (analytical grade, 98%) was purchased from Ambeed, Inc. (Arlington Heights, IL, USA). 6PPD-quinone (6PPDQ) (>98.8%) and 6QDI (≥96%) were purchased from HPC Standards Inc. (Atlanta, GA, USA). 4-nitrodiphenylamine (4-NDPA;99%), 4-aminodiphenylamine (4-ADPA;98%), and 1,3-dimethylbutylamine (1,3-DMBA; 98%) were from Sigma-Aldrich (St.Louis, MO, USA); 4-nitrosodiphenylamine (4s DPA) was from ChemServiceInc. (WestChester, PA, USA), and 4-hydroxydiphenylamine (4-HDPA) was

from FisherScientific (Pittsburgh, PA, USA). 6PPDQ-d5 (solution in acetonitrile, 100mg/L, full deuteration of the aniline ring) was purchased from HPC Standards Inc. Methanol, acetone, formic acid, dimethyl sulfoxide (DMSO), water (all Optima LC/MS grade, ≥99%) and absolute ethanol (~99.5%) were purchased from Fisher Scientific (Waltham, MA, USA) and used as received. Sodium phosphate dibasic anhydrous, sodium phosphate monobasic anhydrous, and sodium hydroxide solution (10 N) were purchased from Fisher Scientific (Waltham, MA, USA). Tert-butanol (ACS reagent grade, ≥99%) and trans-cinnamic acid (CA; ≥99%) were purchased from Sigma-Aldrich (St. Louis, MO, USA). A water purification system (Thermo Barnstead Nanopure Diamond UV; Dubuque, IA, USA) provided ultrapure deionized (DI) water. Glass-bodied chromatography columns (20 cm length × 2.5 cm diameter) were purchased from DWK Life Sciences (Vineland, NJ, USA). A photochemical ozone (O_3) generator (UVP Model 97-0066-01, Upland, CA, USA) – relying on vacuum UV photolysis of O_2 at λ = 185 nm to generate O_3 – was used for O_3 generation, and an O_3 detector (IN USA Model IN2000-LC, Needham, MA, USA) – relying on UV absorbance of O_3 at l = 253.7 nm – was used for O_3 concentration measurements. Zero-grade air was purchased from Airgas (Seattle, WA, USA) and used as the feed gas for ozonation and O_3-free air control experiments.

4.2.2 Gaseous phase ozonation of 6PPD, 6PPDQ and 6QDI

Ozonation of 6PPD family chemicals was investigated, specifically 6PPD, 6PPDQ, and 6QDI. 6QDI was proposed as potential intermediate TP for 6PPDQ formation during 6PPD ozonation (Seiwert et al., 2021a; **Figure S4.1**). Slides coated with 6PPD, 6PPDQ and 6QDI were prepared following methods described previously (Hu et al., 2022). Glass slides

(7.62×2.54 cm; VWR, Radnor, PA, USA) were pre-cleaned (water/methanol, baked at 400 ºC) and cut to 3.8×2 cm. Twenty μL of methanolic stock solution (**Table 4.1**; at different concentrations; prepared within 24 h of experiments) was applied evenly to one surface of a slide with a 50 μL glass syringe. Slides were prepared in triplicate for each time point.

Table 4.1. Gaseous phase ozonation experimental conditions for 6PPD, 6PPDQ, and 6QDI.

Chemical	C_{stock} (mg/L)	Mass on slide (μg)	Inlet Ozone (ppbv)
6PPD	500	10	280
6PPD	200	4	280
6PPD	200	4	140
6PPD	200	4	70
6PPD	100	2	280
6PPD	200	4	0 (Control)
6PPDQ	50	1	280
6PPDQ	50	1	0 (Control)
6QDI	100	2	280
6QDI	100	2	0 (Control)

We note that 6PPD's substantial instability under oxic and aqueous conditions (OSPAR Commission, 2006c) hinders its accurate quantification under many circumstances because timescales for analytical processing and measurement are potentially comparable with timescales for reactive transformation. To mitigate potential 6PPD loss/variability, analytical calibration curves used identical stock solutions and were analyzed on the same timescales as experimental systems to ensure accurate quantification. The experiments were conducted in a dark room equipped with a red light (wavelength = 620-750 nm), to minimize possible competing photochemical reactions. The experimental setup was described elsewhere (Hu et al., 2022). O_3 was produced from feed air (1 L/min) by a flow-through, photochemical O_3

generator, with compressed zero grade air as the feed gas. O_3 concentrations at the outlet of the chromatography column were monitored continuously, while inlet O_3 concentrations (stable throughout experiments at 271±6, 148±7, 73±5 ppbv for nominal conditions of 280, 140, and 70 ppbv, respectively) were measured using column inlet/outlet two-way valves and a bypass line, in each case using a flow-through O_3 detector (**Figure S4.2)**. O_3-free control experiments with zero grade air were also conducted with each test compound (4 μg/plate 6PPD, 1 μg/plate 6PPDQ, 2 μg/plate 6QDI) using the same setup as described for ozonation experiments. After reaction, slides were placed in amber glass jars, 10 mL methanol was added, and the jars were sonicated (30 min). The extracts (950 μL) then spiked with 6PPDQ-d_5 (50 μL; 100 μg/L) for LC-MS/MS analysis.

4.2.3 6PPD aqueous phase ozonation

To determine an apparent 6PPD ozonation reaction rate constant in water, a competitive kinetics approach was deployed (Deborde et al., 2005), due to a presumed fast constant rate of 6PPD-O_3 reaction based on chemicals with similar structures ($k_{O3,\ aniline,\ average} = 2\times10^5\ M^{-1}s^{-1}$, $k_{O3,\ Sulfamethoxazole} = 2\times10^6\ M^{-1}s^{-1}$; (C.H. Kuo, 1985; Jin et al., 2012)). Cinnamic acid (CA; $k_{O3,\ CA} = 1\times10^6\ M^{-1}s^{-1}$; (Leitzke et al., 2001)) was selected as a competitor chemical, given a rate constant similar to the presumed 6PPD constant rates. 6PPD stock solution (in *tert*-butanol; 2mM) and aqueous CA stock (2mM) were prepared within 24 h of the experiment (stored at 4°C). Rate constant determination experiments were performed in phosphate buffer solution (5mM; adjusted with 0.1N sodium hydroxide to pH 7±0.1), at room temperature (25°C). *tert*-butanol ($k_{OH,t\text{-}BuOH} = 6e8\ M^{-1}s^{-1}$) was selected as the organic solvent for 6PPD stock solutions

because it is an •OH radical scavenger (Hoigné and Bader, 1983). For each experiment, the concentration of *tert*-butanol was maintained above 100 mM to achieve excess •OH radical scavenging effect (> 99% scavenging capacity), calculated as follows (Lutz and Kloas, 1999):

$$k_{OH,t-BuOH}[t-BuOH]0 \; > \; 50\,(k_{OH,6PPD}\,[6PPD]_{t,0} \; + \; k_{OH,CA}\,[CA]_{t,0})$$

where $[6PPD]_{t,0}$ and $[CA]_{t,0}$ are the initial concentrations of 6PPD and CA, respectively. The ozone stock was prepared by sparging ozone gas into 1-L LC-MS grade water samples. Ozone was generated from pure oxygen by a corona-discharge O_3 generator (AC2025, IN USA, Norwood, MA, USA) and pumped into the ozone stock solution at 500 mL/min. The ozone concentration (1.1 mM) was determined with UV spectrometer at wavelength = 258 nm, multiple measurements (n>3) were taken before the experiment to ensure the ozone solution was fully equilibrated. Control samples were prepared in triplicate with identical buffer solution and 6PPD and CA stocks without spiking the O_3 stock solution. The tested experimental conditions are listed in **Table 4.2A**.

Table 4.2. Experimental conditions for 6PPD aqueous ozonation reactions. The tested conditions in each row were performed in duplicate. Control samples were not spiked with ozone stock.

A. Experimental condition for 6PPD-CA competitive kinetics experiments.

$6PPD_{t,0}$ (µM)	**$CA_{t,0}$ (µM)**	**O_3 stock vol (µL)**	**O_3 dose (µM)**	**O_3:6PPD**
26	11	0	0	Control
26	11	250	14	0.5
26	11	300	17	0.6
26	11	350	19	0.7
26	11	400	22	0.8
26	11	450	25	0.9

26	11	500	28	1.1

B. Experimental conditions for 6PPD aqueous ozonation experiments. The 6PPD concentration at t_0 and O_3:6PPD values were calculated from the actual detections. O_3 stock concentration is 0.23 mM for the experiments.

$6PPD_{t,0}$ (μM)	**•OH Scavenger (100 mM)**	**Reaction volume (mL)**	**ozone stock vol (μL)**	**O_3:6PPD**
1.38	NA	50	0	0
1.38	NA	50	77	0.25
1.38	NA	50	130	0.44
1.38	NA	50	261	0.88
1.38	NA	50	522	1.75
1.29	Acetone	20	0	0
1.29	Acetone	20	20	0.18
1.29	Acetone	20	45	0.40
1.29	Acetone	20	87	0.78
1.29	Acetone	20	175	1.56
1.42	DMSO	20	0	0
1.42	DMSO	20	20	0.16
1.42	DMSO	20	45	0.37
1.42	DMSO	20	87	0.71
1.42	DMSO	20	175	1.42
1.35	tBuOH	20	0	0
2.35	tBuOH	20	20	0.17
3.35	tBuOH	20	45	0.38
4.35	tBuOH	20	87	0.74
5.35	tBuOH	20	175	1.49

For each tested condition, experiments were performed in 24 mL amber bottles containing 20 mL buffer solution and spiked with CA and 6PPD. Ozone stock solutions were then spiked into the bottles at different volumes with a glass gas-tight syringe. After O_3 solution was spiked, the bottles were immediately capped to avoid ozone degassing in the solution. PTFE-coated stir bars were placed in the bottles and vigorously mixed the solution throughout the reaction period. Pilot studies suggested one-to-one dilution of 6PPD water solution with methanol would prohibit 6PPD aqueous degradation. Therefore, after 1 min reaction, 500 μL

of solution was withdrawn from the amber bottles, diluted with 450 μL methanol and 50 μL 6PPDQ-d_5, and then analyzed by LC-MS/MS within 6 hours.

In addition to the 6PPD-CA competitive kinetics reactions, 6PPD ozonation (1 μM; stock solution prepared within •OH scavenger solvents) without a competitor was also investigated with or without an •OH scavenger (tert-butanol, acetone and DMSO; [•OH scavenger] = 100 mM; **Table 4.2B**). Different O_3:6PPD ratios were tested (0~2) by adjusting spiked ozone stock volumes. The experimental process was the same as for 6PPD-CA competitive reaction experiments described above. For each tested condition, experiments were performed in 20 mL amber bottles. Ozone was added to each bottle containing buffered solution spiked with 6PPD and CA with a glass gas-tight syringe. The bottles were immediately capped to avoid ozone degassing in the solution. PTFE-coated stir bars were placed in the bottles and used to vigorously mix the solution throughout the reaction. After 1 min, 500 μL of solution was drawn from the amber bottles, diluted with 450 μL methanol and 50 μL 6PPDQ-d_5 and then analyzed by LC-MS/MS within 6 hours.

4.2.4 Gaseous and aqueous stability of 6PPD

6PPD stability under aerobic conditions in air and water was examined. Gas phase stability experiments were conducted similarly to the ozonation experiments, albeit with gas-tight glove bags (Cole-Parmer, Vernon Hills, IL, USA) containing desiccants instead of glass-bodied columns, under no-detectable-ozone conditions. For an N_2 control experiment, an oxygen absorbent (PackFreshUSA, Ontario, CA, USA) was added to remove residual oxygen. 6PPD-coated slides (25 μg/slide) were placed into the gas-tight glove bags at different times

(1, 2, 3, 7, 14, 21, 28, 35, and 42 d); bags were purged with zero grade air or nitrogen (control group) before and after putting in the samples. On day 42, all slides were extracted with methanol (10 mL methanol per slide; 30 min sonication). Extracts were spiked with 50 μL 6PPD-d_5 and analyzed by LC-MS/MS.

6PPD aqueous stability was also investigated following OECD 111 protocol ("Test No. 111: Hydrolysis as a Function of pH; OECD Guidelines for the Testing of Chemicals, Section 1: Physical-Chemical properties | OECD iLibrary," n.d.) with slight alteration by using an atmospherically controlled environment within a glove box (50800 Model, Labconco, Kansas City, MO, USA). The glove box was purged with compressed air or nitrogen (control group) prior to the experiment (>50 cycles). Before the experiment, potassium phosphate buffer solution (5 mM, pH 7) was degassed under vacuum (>1 h, -30psi) and sparged with compressed air or nitrogen overnight within the glove box. Dissolved oxygen test kits (CHEMetrics, Midland, VA, USA) were used to determine the solution dissolved oxygen concentration in the nitrogen control group. The buffer solution was spiked with 6PPD methanolic stock solution (10 g/L; final concentration 2 mg/L) then split into triplicate aliquots (100 mL) in 250 mL flasks. The flasks were covered with aluminum foil and shaken (100 rpm) for 72 hours. At different time points (0, 0.5, 2, 4, 8, 24, 36, 48, and 72h), 500 μL subsamples from each container were taken, diluted with 450 μL methanol, spiked with 50 μL 6PPDQ-*d5* and analyzed by LC-MS/MS.

4.2.5 Instrument methods

The quantification of 6PPD, 6PPDQ, 6QDI, CA and major TPs used an Agilent 1290 Infinity ultrahigh performance liquid chromatography (UHPLC) coupled to a triple-quadrupole mass spectrometer (Agilent G6460A) with electrospray ionization (ESI+ mode). The chromatographic separation of samples (10 μL) used a reversed-phase analytical column (Eclipse Plus C18, 2.1 mm × 100 mm, 1.8 μm particle size; Agilent, USA) with a Zorbax Eclipse XDB-C18 Guard Column (2.1 mm × 12.5 mm, 5 μm particle size; Agilent, USA) held at 25 °C. For ozonated samples including 6PPD, 6PPDQ, 6QDI and major TPs analysis, binary mobile phases consisting of DI water (A) and methanol (B), both with 0.1% formic acid, were used. The gradient program was: start at 5% B, 5% B at 1 min, 50% B at 4 min, 100% B at 17-20 min, then 5% B at 20–24 min for re-equilibration. For CA-6PPD competitive kinetics samples analysis, binary mobile (0.2 mL/min flow rate) phases consisting of LC-MS grade water (A) and methanol (B), both with 0.3 g/L ammonium acetate, were used. The gradient program was start at 50% B, 100% B at 6-8 min, 50% B at 9–12 min for re-equilibration. Nitrogen was used as nebulizing and desolvation gas (350 °C, 10 L/min) and capillary voltages were 3500 V. The ion transitions used for chemical analysis were summarized in **Table 4.3**. All the chemicals were scanned under multiple reaction monitoring (MRM) mode.

Table 4.3. LC-MS/MS dMRM parameters for 6PPD, and the TPs, including retention time (RT, min), precursor ion (*m/z*), quantifier and qualifier ions (*m/z*; collision energy values provided in parentheses, eV), and fragmentor voltage (V). The table below following Zhao et al. (Zhao et al. 2023)

Target analyte	Precursor ion	Quantifier ion	Qualifier ion one	Qualifier ion two	Fragmentor voltage (V)
6PPD	269.2	184.0 (20)	107.1 (50)	93.1 (40)	100

6PPDQ	299.2	215.1 (14)	187.1 (26)	-	110
6QDI	267.2	183.0 (10)	166.0 (30)	-	110
4-ADPA	185.1	93.1 (20)	167.0 (30)	108.0 (25)	80
4s DPA	199.0	181.1 (25)	169.0 (20)	128.0 (50)	100
1,3-DMBA	102.1	85.1 (5)	57.2 (10)	-	80
4-NDPA	215.0	198.0 (15)	168.0 (25)	-	100
4-HDPA	186.0	109.0 (25)	92.0 (20)	-	100
TP 282b	283.2	266.1 (10)	224.1 (20)	210.1 (20)	80
TP 274	275.1	182.1 (5)	174.0 (20)	98.0 (20)	80
CA	147.0	102.9 (8)	102.9 (8)	-	60
6PPDQ-d_5	304.2	220.1 (14)	192.1 (26)	-	110

During the sample analysis, large 6PPD signals were detected within the 6QDI standard. Therefore, complementary analysis for 6QDI was performed to determine the sample purity, as described by Zhao et al. (**Appendix A**) (Zhao et al. 2023). The instrumental analysis and communications with the manufacturer (HPC) confirmed 6QDI standard purity of ≥96%, the 6PPD detections from the instrumentation were analytical artifacts (which were 6QDI responses).

The ion with higher peak area response (if more than one transition exists) was used for quantification and the second for qualification and identity confirmation (maximum tolerance of ± 20% for quantifier-to-qualifier ion ratios). For quality assurance and quality control (QA/QC), the samples were run in the following order: calibration curve samples were run at beginning, and experimental replicates were analyzed sequentially with methanol blank analyzed after every 12 sample injections to check for carry-over (no carry over observed), a QC sample (Level 4) from the calibration curve was reanalyzed at the end of the sequence (<5% variation of peak area was observed). The data were processed with Agilent Quan Browser (v08.00.00) and Excel (Microsoft 365).

4.3 Results and Discussion

4.3.1 Heterogeneous gas phase ozonation of 6PPD

6PPD ozonation kinetics were investigated with varied mass loadings (10, 4, and 2 µg/plate; inlet O_3 280 ppbv) and varied inlet O_3 concentrations (70, 140, and 280 ppbv; 6PPD mass load 4 µg/plate). Ozonation reactions during the heterogeneous exposures were assumed to occur on the surface of the slide rather than within the gas phase, following the assumption of limited to no volatility of 6PPD or similar non-volatile chemicals (Hu et al., 2022). Ozone dosing was maintained until 6PPD mass degradation rates slowed or a concentration plateau was observed. O_3 inlet concentrations were stable at 271±6, 148±7, and 73±5 for different O_3 concentrations, the O_3 inlet concentrations were tracked throughout the experiments (**Figure 4.1**). During the control experiment, no 6PPD degradation occurred with short-term (4 h) air exposure at ambient temperature and pressure (4 ug/plate, t_0 vs. 4 h, p-value = 0.26), indicating 6PPD stability with exposure to molecular oxygen under tested conditions near 20°C.

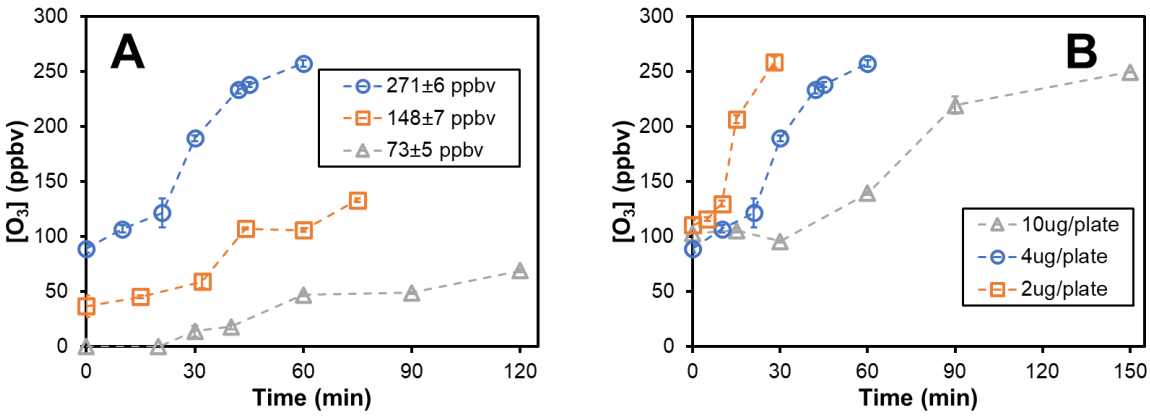


Figure 4.1. Column outlet O_3 concentration measurement vs. reaction time. (A) 4 ug/plate 6PPD mass load tested under different inlet O_3 concentrations and (B) different 6PPD mass load tested under 280 ppbv inlet O_3 concentration. The error bars in each panel represent the standard deviation from three replicate readings.

Under all tested conditions, 6PPD transformation slowed or ceased over 0.5 to 3 h of reaction time (**Figure 4.2**). Several factors, including decreased reactant mass at the reaction interface (i.e., the surface of the 6PPD-coated slide) and protective film formation, might contribute to this phenomenon, slowing the reaction by limiting contact between O_3 and 6PPD (Cataldo et al., 2015; Huntink, 2003; Lattimer et al., 1983). This hypothesis was supported by the observation that across all groups, higher ozone inlet concentrations and less 6PPD mass yielded faster reaction rates (**Figure 4.2E-F**). Across all groups, 93 ± 2 % of 6PPD mass reacted during the experiments, underscoring that O_3 or O_3-related oxidants (e.g., hydroxyl radicals) facilitate 6PPD transformation.

Empirical first-order reaction rate constants were calculated for all 6PPD ozonation conditions (**Table 4.4**; rate constant for 6PPDQ ozonation also included, with experimental results detailed in **Section 4.3.2)**. These calculated rate constants only reflect reaction kinetics under the tested conditions in the current study. Due to the complexity of the reaction system, further investigations are merited to measure the actual reaction rate constants for such systems. Notably, the normalized reaction kinetics of 6PPD ozonation (**Figure 4.2C-D**) across different conditions showed similar patterns, suggesting that 6PPD ozonation follows identical empirical reaction orders under different conditions.

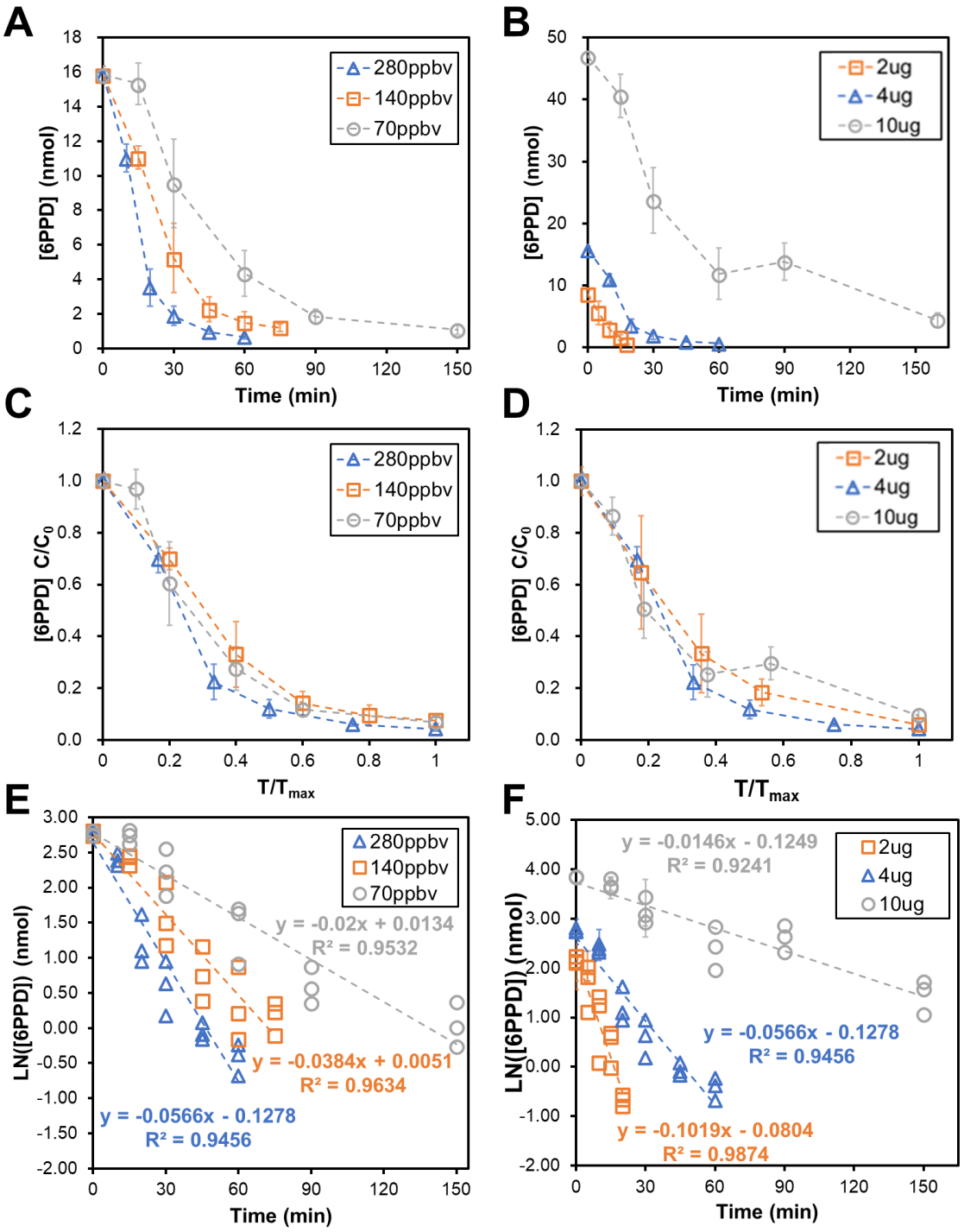


Figure 4.2. 6PPD concentration over time during the ozonation reactions. (A) 4 ug/plate 6PPD mass load tested under different inlet O_3 concentrations; (B) different 6PPD mass loadings reacted at 280 ppbv inlet O_3 concentration; (C) normalized 4 ug/plate 6PPD mass load tested under different inlet O_3 concentrations over time; (D) normalized different 6PPD mass load tested under 280 ppbv inlet O_3 concentration over time; (E) natural log of 6PPD concentration overtime, derived from panel A, for empirical first order rate constant and; (F) natural log of 6PPD concentration overtime, derived from panel B, for empirical first order rate constant. In panels C to E, all measured 6PPD masses were normalized to the initial measured sample mass,

and in panels D to E, reaction time was normalized to the maximum sampling time point in each group. Error bars in each panel represent the standard deviation from three replicate samples.

Table 4.4. Gaseous phase ozonation experimental conditions and empirical first-order reaction rate constants for 6PPD and 6PPDQ.

Reactant	Mass load (μg/plate)	Mass load (nmol/plate)	Inlet O_3 Concentration (ppbv)	Empirical rate constant (min^{-1})	Half-life (min)
6PPD	4	14.9	280	0.057	12.2
6PPD	4	14.9	140	0.038	18.2
6PPD	4	14.9	70	0.020	34.7
6PPD	2	7.5	280	0.102	6.8
6PPD	10	37.3	280	0.015	46.2
6PPDQ	1	3.3	280	0.059	11.7

Major LC-HRMS detectable 6PPD ozonation TPs were also investigated across O_3 exposure conditions. Average molar yield of 6PPDQ, the highly toxic TP (Tian et al., 2021b) was 12±4% across all samples (**Figure S4.3**; mass generated over time summarized in **Figure 4.3**). No 6PPDQ formation was detected in zero grade air controls (with no presence of O_3), consistent with the lack of any apparent 6PPD transformation under such conditions. The estimated molar yield of 6PPDQ was consistent with our previous report (9.7%; (Hu et al., 2022)), underscoring that 6PPDQ was a fairly stable TP during short-term 6PPD ozonation reactions. However, 6PPDQ molar yields tend to decrease over time upon further ozone exposure, implying that 6PPDQ is itself an intermediate TP that is somewhat ozone sensitive, with other terminal TPs formed during O_3 longer exposures.

While conditions in the current study (70-280 ppbv O_3) were intended to mimic environmental conditions, these concentrations are still slightly to moderately higher than typical environmental conditions (e.g., 1-84 ppbv for Washington State (US EPA, 2014)).

Under lower O_3 concentrations, we would expect slower 6PPD consumption and TP formation dynamics.

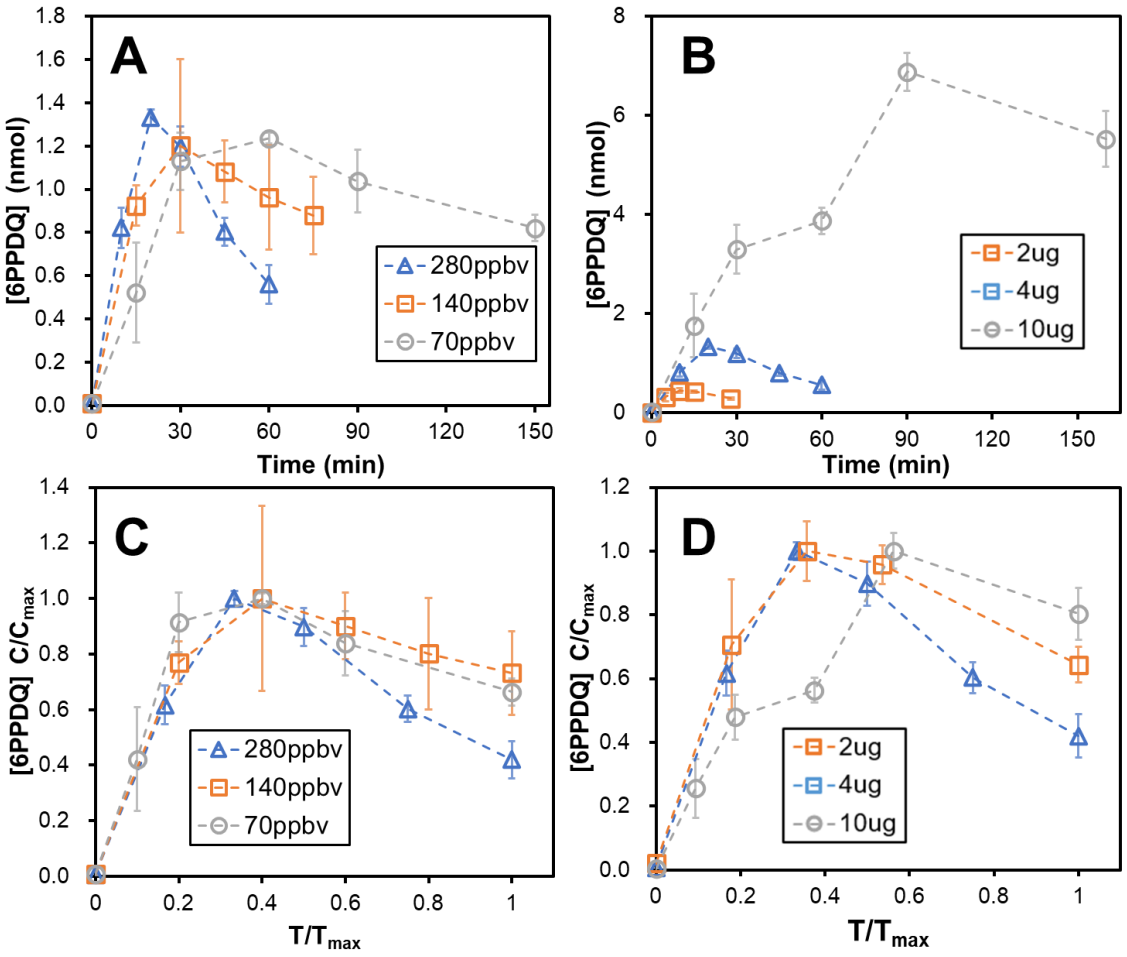
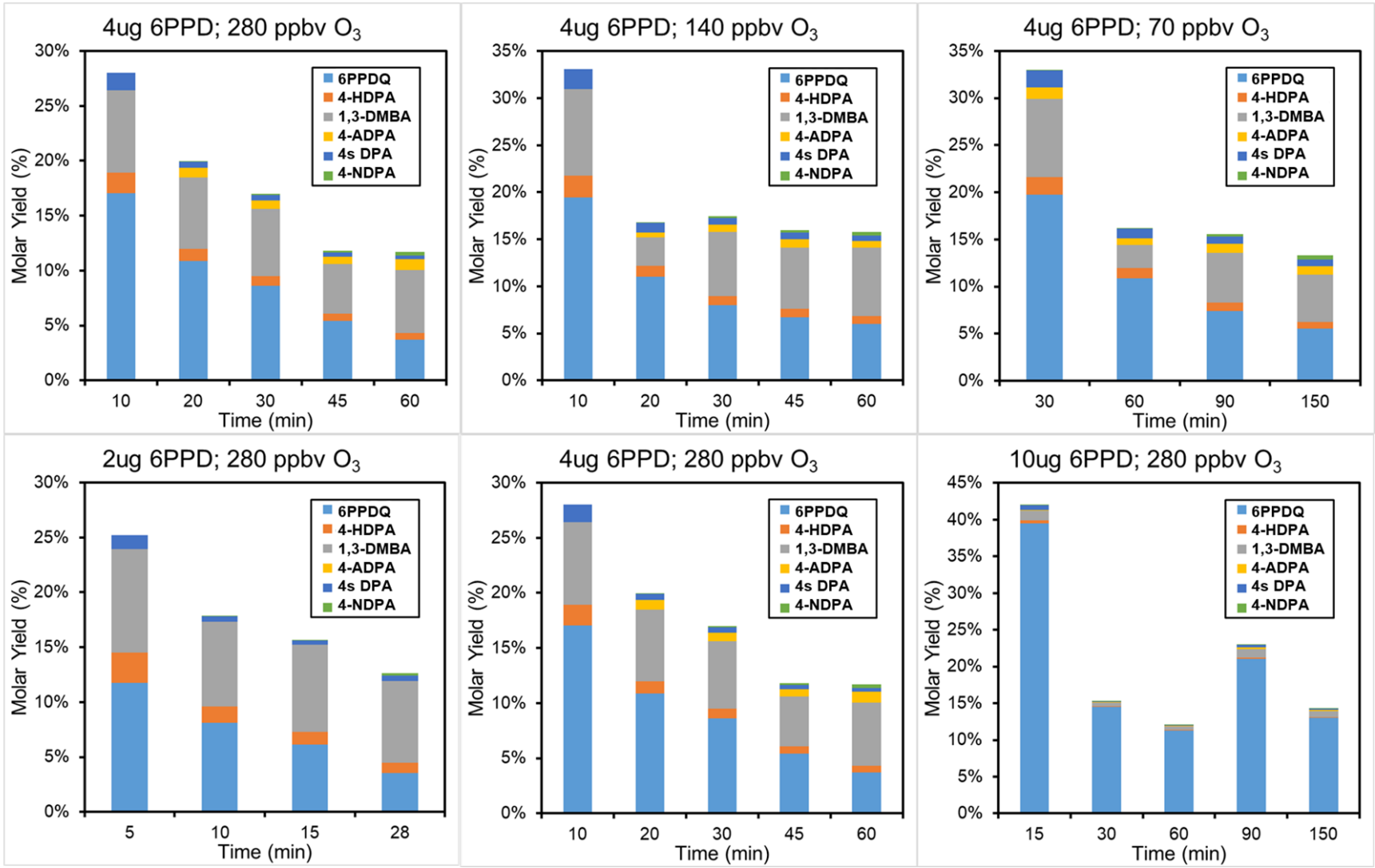


Figure 4.3. 6PPDQ concentration over time during the ozonation reactions. (A, C) 4 ug/plate 6PPD mass load tested under different inlet O_3 concentrations and (B, D) different 6PPD mass load tested under 280 ppbv inlet O_3 concentration. Concentrations and time in (C) and (D) were normalized by normalizing all measured 6PPDQ masses to the maximum mass detected in each group and normalizing the reaction time to the maximum sampling time point in each group. Error bars in each panel represent the standard deviation from three replicate samples.

Beyond 6PPDQ, other TPs (4-HDPA: TP 185; 4s DPA: TP 198; 1,3-DMBA: TP 101; 4-ADPA: TP 184; 4-NDPA: TP 214) identified by Zhao et al. (Zhao et al., 2023b) were detected and quantified over time (**Figure 4.4**).

A


4ug 6PPD; 280 ppbv O_3
4ug 6PPD; 140 ppbv O_3
4ug 6PPD; 70 ppbv O_3
2ug 6PPD; 280 ppbv O_3
4ug 6PPD; 280 ppbv O_3
10ug 6PPD; 280 ppbv O_3
Molar Yield (%)
Time (min)
6PPDQ
4-HDPA
1,3-DMBA
4-ADPA
4s DPA
4-NDPA

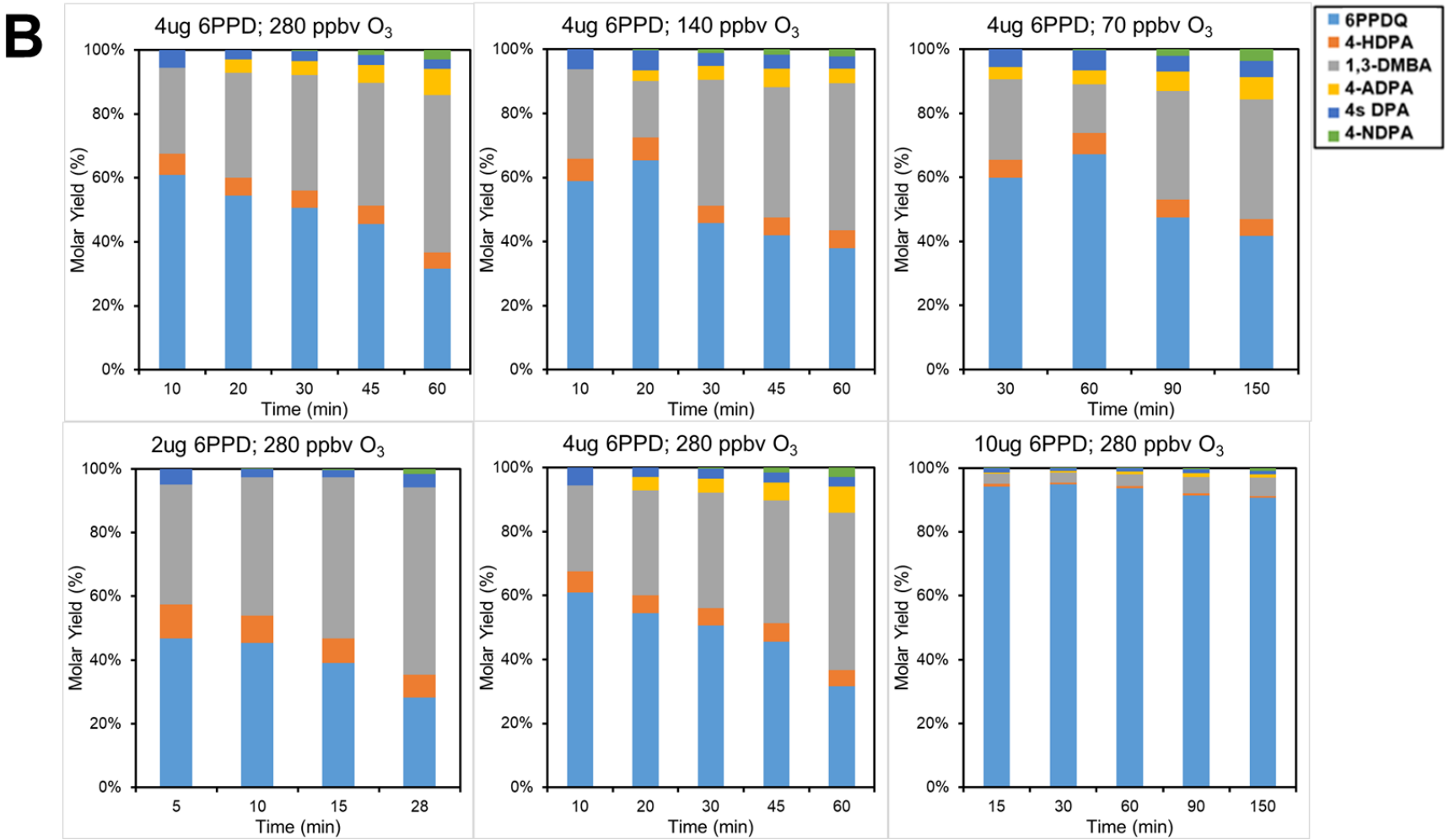


Figure 4.4. TP composition of ozonated 6PPD solids under different exposure conditions. (A) Molar yields of 6PPD converted into the quantified TPs. (B) Normalized molar yield of all the TPs. Each data bar was averaged from triplicate samples.

For identical mass loads with different inlet O_3 concentrations, TPs compositions were consistent, suggesting that only 6PPD surface layers interacted with O_3, while the majority of 6PPD mass residing in deeper layers didn't react or reacted incompletely with O_3. However, at different mass loadings for the same inlet O_3 exposure, the TP compositions obviously shifted. A higher 6PPD mass load yielded substantially more 6PPDQ while a lower mass load yielded more 4-HDPA and 1,3-DMBA (**Figure 4.4B**). As reported in several studies (Wylie et al., 2020; Zhou et al., 2019), high 6PPD mass loads result in more residual 6PPD and thicker layers of 6PPD and TPs mixture. The shifting of the TP composition across different mass loads suggests that reducing available 6PPD, a preferential ozone reaction pathway, promoted secondary reactions of the TPs with ozone, with further transformation to terminal products. These data also indicated that most of the 6PPD TPs detected are intermediate TPs that undergo further transformation with extended O_3 exposure.

Two additional TPs (TP 274 and TP 282) reported by Zhao et al. (Zhao et al. 2023) were also detected after 6PPD ozonation (**Figure 4.5**). Both TPs exhibited similar formation dynamics as 6PPDQ, suggesting that multiple parallel reaction pathways of 6PPD oxidative transformation exist (Hu et al., 2022). These TPs are sometimes abundant by peak area in environmental samples (Hu et al., 2022), although further investigation is required for full structural identification and quantification.

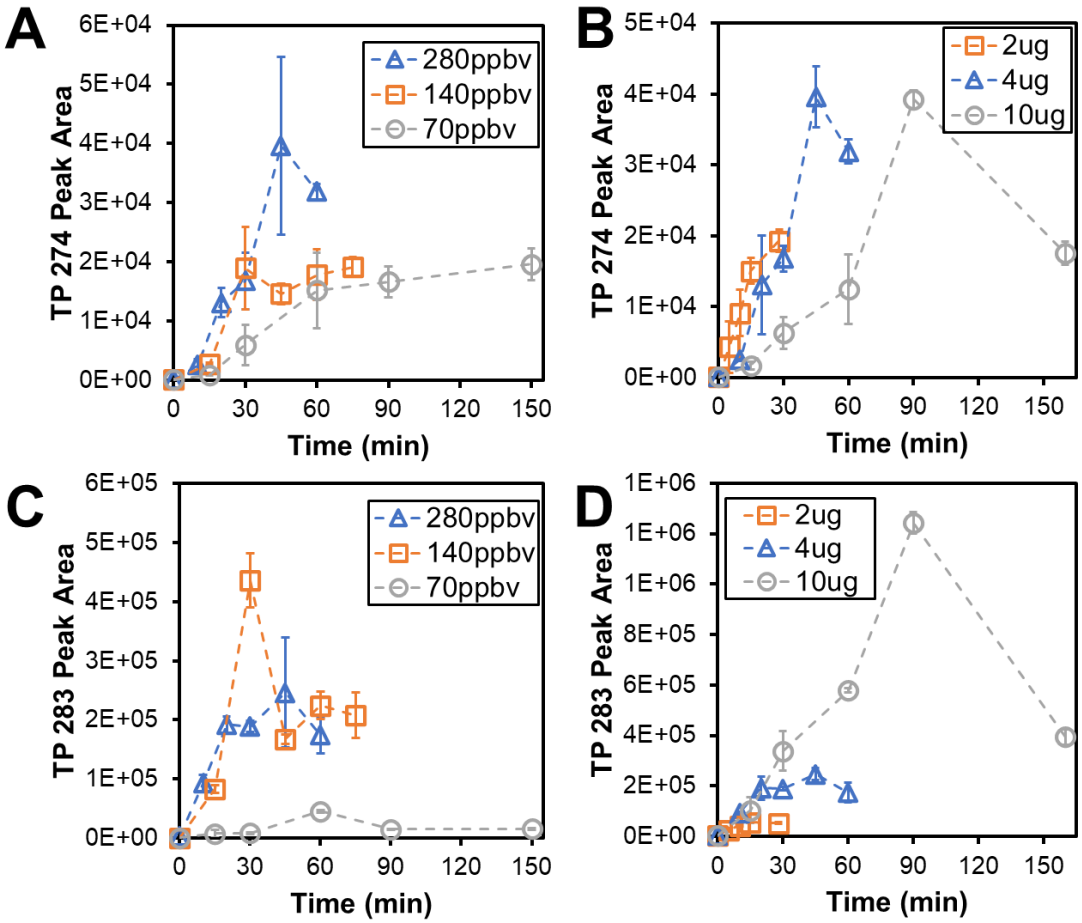


Figure 4.5. Peak areas of TP 274 and TP 282 over time during the ozonation reactions. (A) TP 274 peak areas detected from 4 ug/plate 6PPD mass load samples tested under different inlet O_3 concentrations and (B) TP 274 peak areas detected from different 6PPD mass load samples tested under 280 ppbv inlets O_3 concentration. (C) TP 282 peak areas detected from 4 ug/plate 6PPD mass load samples tested under different inlet O_3 concentrations and (D) TP 282 peak areas detected from different 6PPD mass load samples tested under 280 ppbv inlets O_3 concentration. The error bars in each panel represent the standard deviation from three replicate samples.

4.3.2 Heterogeneous gas phase ozonation of 6PPDQ and 6QDI

To better understand 6PPD ozonation, including reaction mechanism and end products, the ozonation kinetics of 6PPDQ and 6QDI were also investigated. 6QDI was previously proposed as an intermediate TP during 6PPDQ formation from 6PPD ozonation (Seiwert et al., 2021a). Inlet ozone concentrations were 282±5 ppbv. During exposure of both 6PPDQ and 6QDI, measured column outlet O_3 concentrations were initially lower (~200 ppbv), suggesting

rapid initial reaction of both compounds, but returned to the inlet concentration within two min (**Figure 4.6**), suggesting limited interactions between O_3 and 6PPDQ or 6QDI.

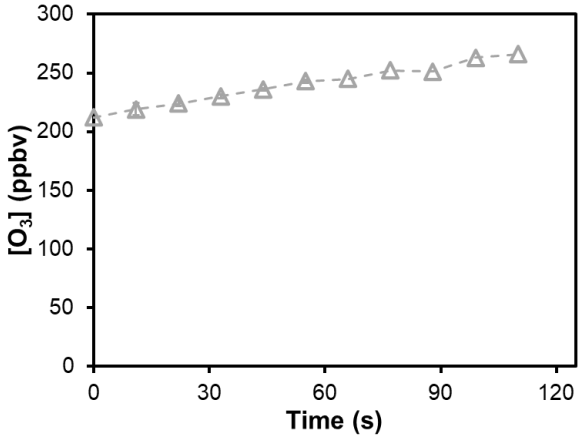


Figure 4.6. O_3 reading from 6PPDQ ozonation. The reading reached input concentration (265-275 ppbv) by 2 min. Each data point represents a single reading from the O_3 monitor.

Similar to 6PPD ozonation, 6PPDQ concentrations declined and then stabilized over 240 min (60% of 6PPDQ mass reacted), potentially due to protective film formation or otherwise unavailable mass for reaction. Among TPs detected after 6PPD ozonation, only 1,3-DMBA was also detected in ozonated 6PPDQ samples, suggesting that 1,3-DMBA is a potential terminal TP of 6PPD and 6PPDQ ozonation (**Figure 4.7**). Of transformed 6PPDQ, 16% was converted (by molar yield) into 1,3-DMBA. No 6PPDQ loss occurred in zero grade air controls (t_0 vs. 4 h, p-value = 0.99), suggesting that 6PPDQ transformation was mainly driven by O_3 or O_3-related secondary oxidants (e.g., hydroxyl radicals) under these conditions, similar to reaction of 6PPD.

The mass loads used for 6PPDQ ozone exposure were lower than 6PPD studies (1 μg vs 4 μg per plate), implying thinner mass films on surfaces and potentially faster (compared to 6PPD under same conditions) reaction kinetics. Because 6PPDQ showed a relatively slow

reaction kinetics with much lower mass load compared to 6PPD (empirical rate constant under 280 ppbv exposures: $k_{6PPDQ,O3,\ 1\mu g/plate}=0.059min^{-1}$, $k_{6PPD,O3,\ 4\mu g/plate}=0.057min^{-1}$), it seems to be more stable (compared to 6PPD) under environmental conditions (US EPA, 2014, n.d.) and with higher potential to induce effects over longer time periods.

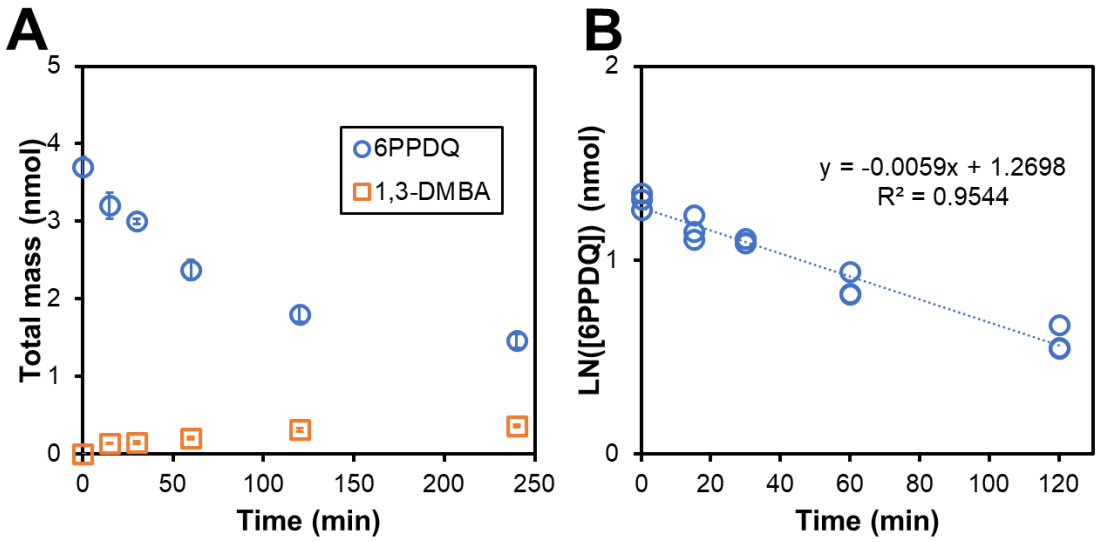


Figure 4.7. Reaction dynamics of 6PPDQ and 1,3-DMBA during heterogeneous gas-phase ozonation (280 ppbv; O_3 generated from zero air or ambient air with O_3 generator). (A) 6PPDQ and 1,3-DMBA mass versus reaction time and; (B) natural log of 6PPDQ concentration (nmol) versus reaction time, calculated from panel A. Error bars correspond to the standard deviation from three replicate slides.

The ozonation of 6QDI, on the other hand, seems to be more complicated. 6QDI showed differential transformation under all 4 experimental conditions (2µg/plate, 180 min exposures; zero grade air control, ambient air control, O_3 produced from zero grade air, O_3 produced from ambient air) (**Figure 4.8**). During the 180 min exposure, 6QDI transformed by 86±3.4% (O_3 produced from zero air), 97±0.5% (O_3 produced from ambient air), and 42±5.2% for air controls (both zero grade air and ambient air without O_3 generator turned on), respectively. The results indicate that while O_3 effectively reacts with 6QDI, 6QDI is

also prone to transformation from oxygen exposure. 6QDI has been co-applied with 6PPD for anti-degradant protection in tire rubbers (Ray, 1998), suggesting that 6QDI might be more protective with respect to oxygen while 6PPD serves primarily as an antiozonant. The synergistic antioxidant effect of both compounds is reported to provide decent protection against O_3 and oxygen in ambient air and extend tire rubber lifetimes (Huntink, 2003).

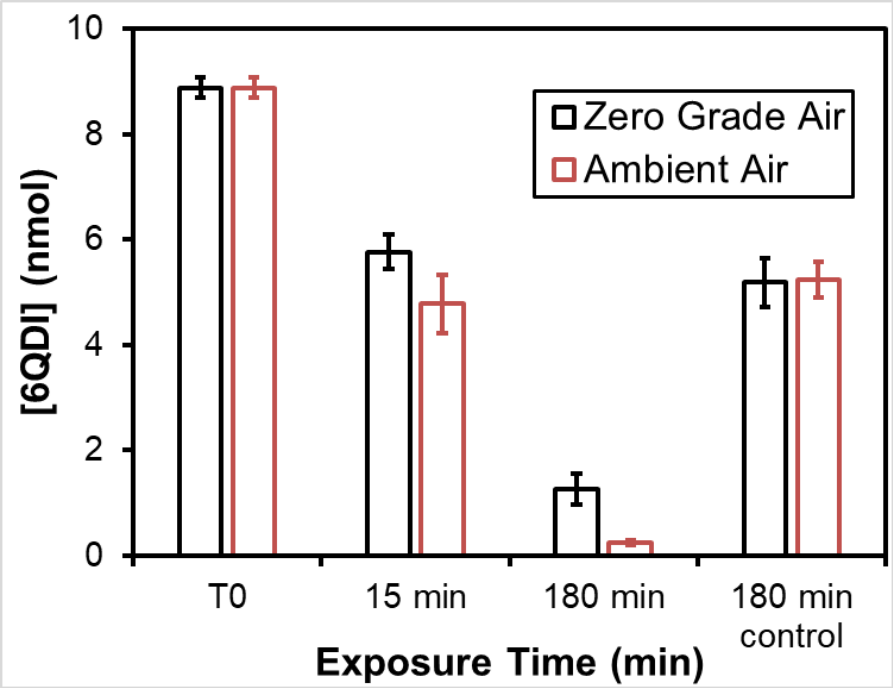


Figure 4.8. 6QDI mass overtime during the heterogeneous gas-phase ozone exposure (280 ppbv) and control experiments (with zero grade air or ambient air) of 6QDI on glass slide surfaces. Error bars correspond to the standard deviation from triplicate slides.

Among TPs detected in ozonated 6PPD samples, 4-HDPA, 1,3-DMBA, 4-ADPA, TP 274, and TP 282 were all detected in ozonated 6QDI samples. 4-HDPA and 4-ADPA were both detected in the 6QDI standard and then decreased during the ozonation process, suggesting they exist as impurities in the chemical standard (the manufacturer declared 4% impurity) or as instrument artifacts due to in-source fragmentation (**Appendix A**). As observed for ozonation of 6PPDQ, 1,3-DMBA seems to be a terminal product of 6PPD-

derived or -related chemicals upon ozonation. No significant formation of 1,3-DMBA occurred in air controls (t_0 vs. 180min, p-value=0.09 for zero grade air control and 0.97 for ambient air control). This shared end product between 6PPD, 6PPDQ, and 6QDI ozonation indicated a similar side chain oxidation reaction pathway across this group of chemicals. TP 274 and TP 282 also formed from 6QDI ozonation and were not detected in air controls, indicating that these TPs formed from O_3-6QDI interactions. Notably, 6PPDQ was not detected in any of the 6QDI samples, suggesting that 6QDI is likely not an intermediate TP during 6PPD to 6PPDQ transformation under these conditions (Seiwert et al., 2022).

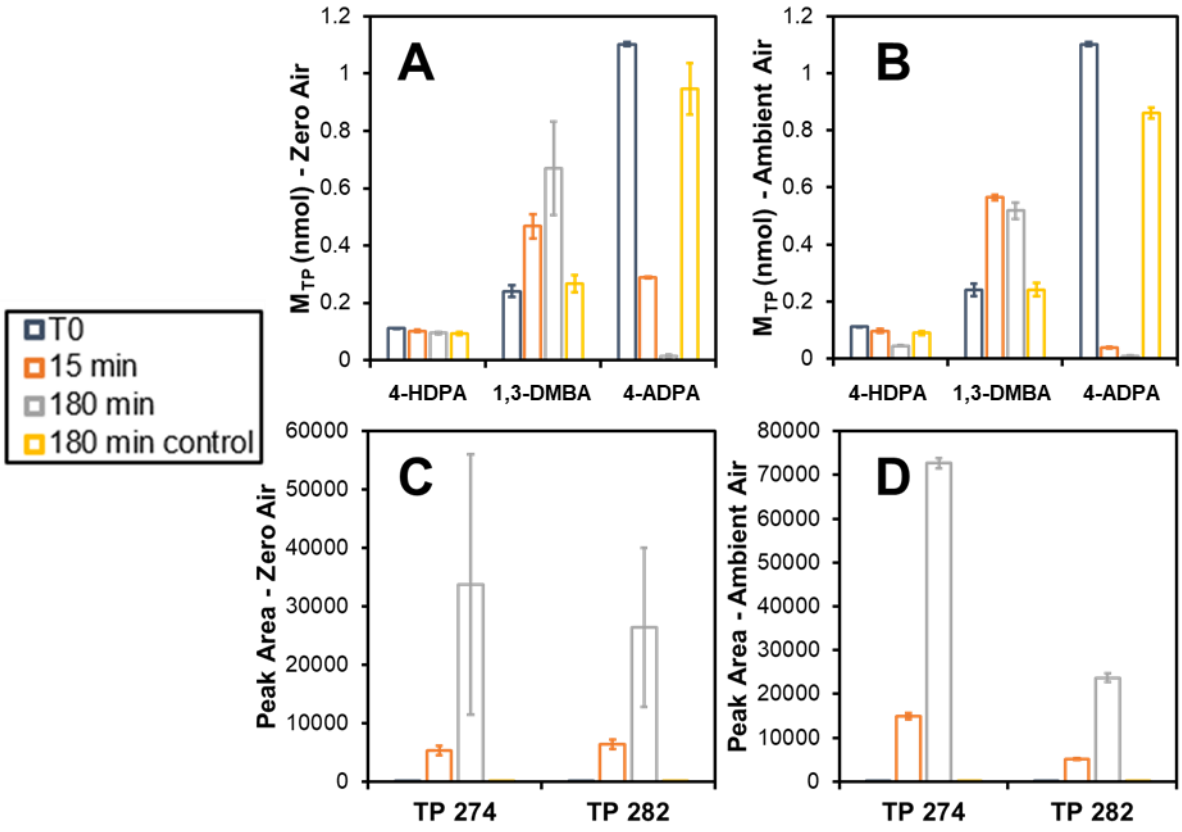


Figure 4.9. Detected TPs over time during heterogeneous ozone exposure and control experiments of pure 6QDI on glass slide surfaces. (A-B) Quantified mass on individual slides for 4-HDPA, 1,3-DMBA and 4-ADPA for O_3 exposure with (A) O_3 generated from zero grade air or (B) O_3 generated from ambient air. (C-D) Detected peak area of TP 274 and TP 282 on the glass slide during the O_3 exposure with (A) O_3 generated from zero air or (D) O_3

generated from ambient air. The O_3 exposure experiment was conducted with 280 ppbv inlet O_3; both O_3 exposure and O_3 free controls used zero air or ambient air as feed gas. TP 274 and TP 282 were reported as detected peak areas because they lack commercial standards. Error bars correspond to the standard deviation from three replicate slides.

4.3.3 Aqueous ozonation of 6PPD

Aqueous phase ozonation of 6PPD was also investigated, with the 6PPD-O_3 reaction rate constant determined to be $(1.18\pm0.16)\times10^6$ $M^{-1}s^{-1}$ via a competitive kinetics experiment (**Figure 4.10**). The measured rate constant suggested 6PPD aqueous ozonation is an extremely rapid process that occurs over timescales of milliseconds under the tested conditions. While such reactions are unlikely in surface waters because dissolved ozone concentrations are very low, O_3 reactors are likely highly effective for 6PPD removal during engineered treatment. No significant differences were observed across systems with or without different hydroxyl radical scavengers (t-test; p-value > 0.05; **Figure 4.11**), revealing that 6PPD ozonation in water was dominated by molecular O_3 interactions with 6PPD rather than hydroxyl radical interactions.

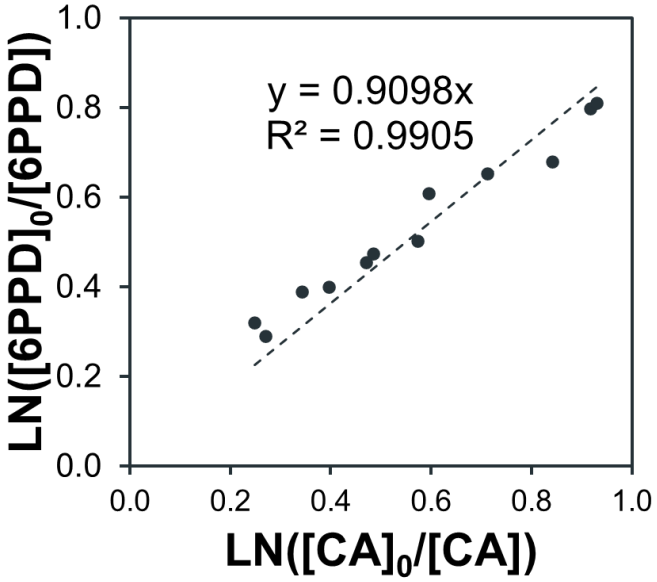


Figure 4.10. Determination of the $k_{app,6PPD,O3}/k_{app,CA,O3}$ for simultaneous reactions of ozone with 6PPD and CA at pH 7.

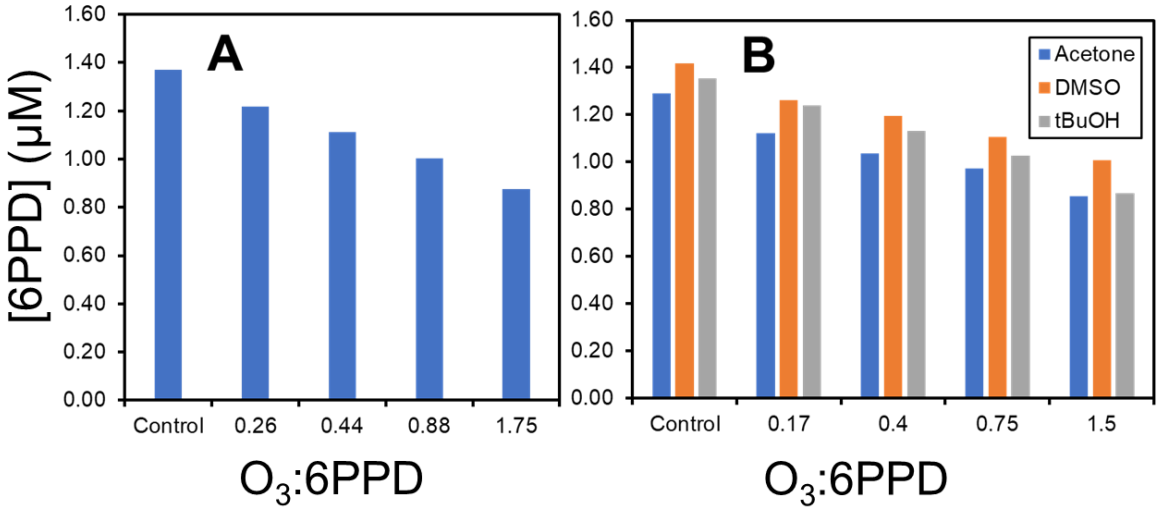


Figure 4.11. Aqueous phase ozonation of 6PPD with different •OH scavengers under different O_3:6PPD ratios. (A) 6PPD ozonation in pure water without •OH scavenger agents. (B) aqueous 6PPD ozonation with acetone, DMSO or tBuOH as •OH scavengers.

Notably, 6PPDQ was not detected under any conditions tested in water, underscoring that gas phase ozonation seems to be the major driver of 6PPDQ formation from 6PPD. While additional experiments are needed to understand whether any 6PPDQ formation can occur from 6PPD in saturated systems, it remains clear that ozone mediated reactions in the aqueous phase are relatively unimportant. An alternative formation pathway for 6PPDQ from 6PPD was also reported by Xu et al. recently, suggesting the existence of other 6PPDQ formation pathways in the environment (Xu et al., 2023).

4.4.4 Gaseous and aqueous stability of 6PPD

The long-term stability of 6PPD (25 μg/plate) under oxidative environmental conditions with O_3-free zero grade air exposure (i.e., O_2 exposure assuming no 6PPD-N_2 interactions) was investigated over 42 days in dark conditions. Gradual loss of 6PPD was observed (**Figure 4.12A**), with 41±3% mass degraded over the 42 days of exposure. Fitting the

reaction data to first order kinetics yielded an apparent first order reaction rate constant for 6PPD-air reaction of 0.01 d^{-1}, implying a 6PPD half-life of 69.3 days under the tested conditions (**Figure 4.12B**). Under nitrogen-only exposure, 6PPD showed similar dynamics (t-test; p-value=0.48, **Figure 4.12C**) as oxygen exposure. Over these time scales, we cannot exclude the potential presence of trace oxygen or ozone residuals entering the glove bags. Further investigations under better controlled conditions would be required to evaluate slow reactions, or those with trace constituents (e.g., other trace gases or constituents).

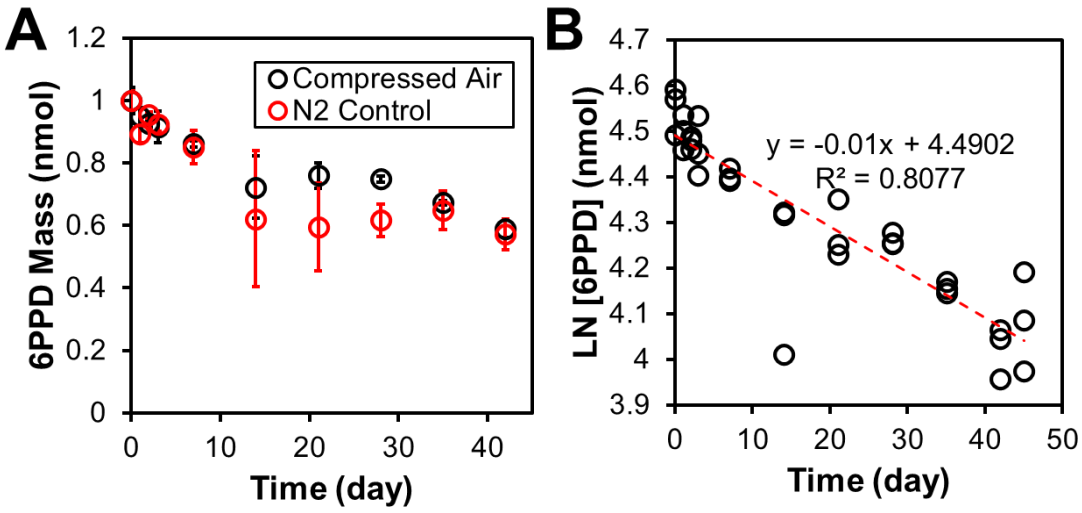


Figure 4.12. Long term (42 days) air exposure of glass slides coated with 6PPD. (A) normalized 6PPD mass (C/C_0) vs. exposure time for both compressed air and N_2 exposure experiments and; (B) natural log of 6PPD mass over time for first order reaction constant estimation. The data points and error bars in panel A were calculated from the experimental triplicates. The data points in panel B represent individual samples, and the red dashed line represents regression of the data points.

Compared to the O_3 exposure, any transformation process of 6PPD under air (O_2) was slow (6PPD-O_2 reaction half-life: 69.3 days, 6PPD-O_3 (10µg/plate, 280 ppbv) reaction half-life: 46.2 min), indicating 6PPD is relatively stable with O_2 exposure for extended periods of time. As it is used ubiquitously in production of tire rubber (0.4-2% by mass) (Sheridan, 2010)

and widely reported as an antioxidant and antiozonant in rubber materials, the current results suggest that 6PPD transformation within rubber materials is primarily driven by ozonation reactions at ambient temperatures. Further investigation would be necessary for 6PPD interactions with other oxidants (e.g., NOx or •OH), reactivity with molecular oxygen at high temperatures (e.g., >50°C), or matrix-promoted reactions involving other reactive chemicals or rubber materials. Additionally, under the conditions that were tested, no gas phase ozonation TPs reported previously (Zhao et al., 2023) or described above, including 6PPDQ, were detected (or were below the detection limit) for 6PPD compressed air and nitrogen gas exposures. Thus, 6PPD reaction processes under O_2 or N_2 exposure seem to have different reaction mechanisms than during ozonation, which merits further investigation.

The aqueous stability of 6PPD was investigated using an atmosphere controlled (with zero grade air or nitrogen) glove box for timescales up to 48 hours with buffer solution ([6PPD]=2 mg/L with methanol as co-solvent at v:v = 0.02%; reaction solution adjusted to pH 7 with dibasic or monobasic potassium phosphate) which is purged with zero grade air (overnight, the headspaces were also filled with zero air from the pre-purged glove box) prior to the addition of 6PPD, 90.5±0.6% 6PPD reacted over 48 hours (**Figure 4.13A**). Fitting the reaction data yielded an apparent first-order reaction rate constant for 6PPD aqueous degradation at 0.069 h^{-1}, implying a 6PPD half-life of 10 hours under these conditions (**Figure 4.13B**). The observation also aligned with previously documented 6PPD half-life values of 5 to 8 hours in water (Hiki et al., 2021; OSPAR., 2006).

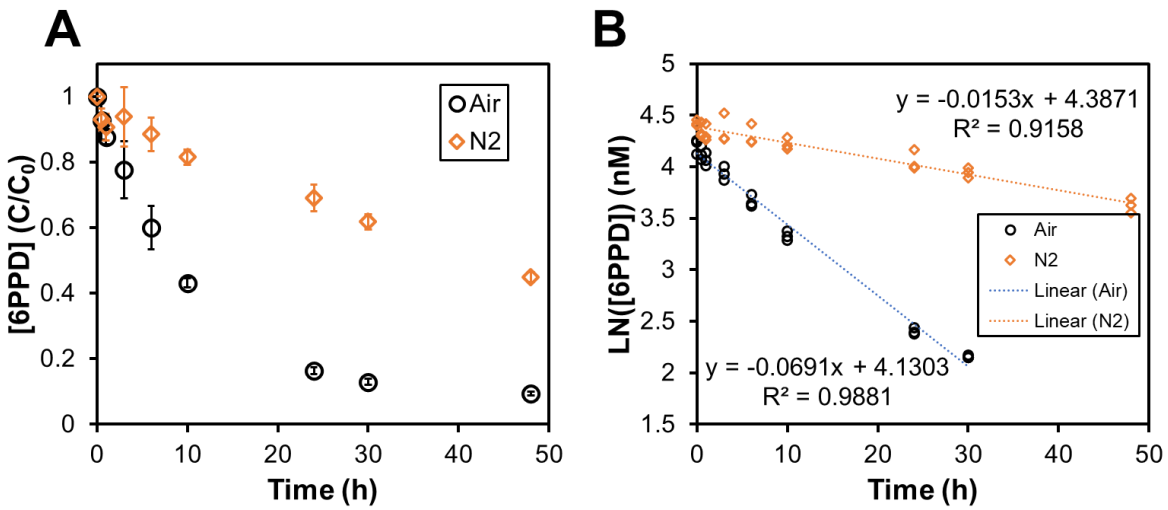


Figure 4.13. Aqueous stability experiment of 6PPD under aerobic (air) and anaerobic (nitrogen) conditions. (A) normalized 6PPD concentration (C/C_0) vs. exposure time in air and nitrogen groups and; (B) natural log of 6PPD concentration over time in air and nitrogen groups. The data points and error bars in panel A were calculated from the experimental triplicates. The data points in panel B represent individual samples, and the dashed line represents regressions of the data points. Samples from the last time point (48h) were excluded for first-order rate constant calculation to reduce skewed residuals, as the reaction plateaued.

For anoxic (nitrogen environment) conditions, dissolved oxygen was measured as below 0.05 mg/L (at t_0) to ~0.2 mg/L (48 h) (**Figure 4.14**), indicating a trace amount of dissolved oxygen was present in the glove box. For nitrogen experimental group, 55±1.5% 6PPD reacted by the end of the 48 hours (**Figure 4.13A**). The apparent first-order reaction rate constant for 6PPD under the nitrogen headspace was calculated at 0.015 h^{-1} (half-life = 46.2 h), indicating a much slower reaction under anaerobic conditions (**Figure 4.13B**). This observation suggested the presence of dissolved oxygen affects 6PPD transformation under aqueous conditions.

Figure 4.14. Dissolved oxygen measurement for the reaction solution under nitrogen conditions. (A) measurement of dissolved oxygen at t_0 and (B) measurement of dissolved oxygen at 48 hrs. The tested tubes were highlighted in red boxes.

These data suggest substantial instability for 6PPD under tested aqueous conditions, with reaction mechanisms such as oxidative degradation likely dominating 6PPD fate in water relative to oxidant mediated reactions, given that ozone concentrations are negligible in natural surface waters. However, only limited conditions (zero grade air and nitrogen; pH 7 and dark conditions) were tested in the current study, other variables that contribute to 6PPD transformation, such as different pHs and light conditions, require further investigation.

Aqueous samples were screened for 6PPD ozonation TPs with available commercial standards. No 6PPDQ was detected (or below detection limit) from all the samples analyzed,

indicating none of the air exposure or aqueous reactions would generate 6PPDQ. Among all the 6PPD-derived TPs, only 4-HDPA and 1,3-DMBA were detected (**Figure 4.15**).

Over 48 h under aerobic conditions, 31.9±2.3% of reacted 6PPD converted into 4-HDPA, while 74.6±5.0% converted into 1,3-DMBA (calculated by molar yield; **Figure 4.15A**). Under anaerobic conditions, 24.9±2.1% of reacted 6PPD converted into 4-HDPA, while 55.9±3.8% converted into 1,3-DMBA (**Figure 4.15B**). The observation suggested that, similar to the gas phase reactions, side chain oxidation plays an important role in aqueous systems. Because 6PPDQ was not detected in either gas or aqueous phase reaction systems that clearly lacked any measurable ozone, we reiterate the importance of gas phase ozonation as a primary mechanism for 6PPDQ formation.

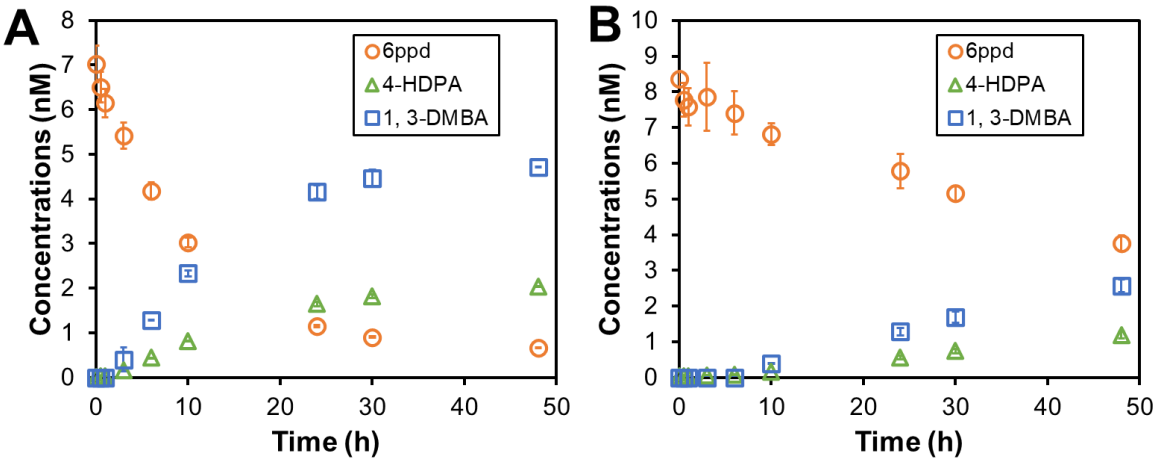


Figure 4.15. Aqueous stability experiment of 6PPD under (A) aerobic (air) and (B) anaerobic (nitrogen) conditions, showing concentrations of 6PPD, 4-HDPA and 1,3-DMBA vs. exposure time (48 h). The data points and error bars were calculated from experimental triplicates.

4.4 Conclusion

6PPD transformation and environmental fate were investigated under varied conditions, including gas phase and aqueous phase ozonation, and stability in aqueous systems and during long-term air exposure. The results provide a comprehensive understanding of 6PPD and related TPs transformation and formation kinetics. Notably, although 6PPD functions as both antioxidant and antiozonant in tire rubber, ozonation was determined to be the primary driving force behind 6PPD transformation and 6PPDQ formation in ambient environments (distinct from transformations occurring within tire rubbers and at high temperatures). Under the tested conditions, 6PPD mass progressively reacted during short term gas phase (93±2% 6PPD degraded during 2.5 h) and aqueous ozone reactions (empirical reaction constant = $(1.18\pm0.16)\times10^{6}$). Stability tests of 6PPD in simple aqueous systems indicated that 90.5±0.6% 6PPD degraded over 48 hours (half-life: 10.0 hours), indicating rapid attenuation of 6PPD in water. 6PPD was much more stable in ambient air (41±3% 6PPD degraded over 42 days; half-life: 69.3 days). Among all tested conditions, 6PPDQ formation was only observed from 6PPD gaseous ozonation process, indicating gas phase 6PPD-O_3 reaction is the primary 6PPDQ formation mechanism. We expect that this process would often occur within or at the surface of TWP or rubber products both before and after they are subject to environmental dispersal. Several TPs, including 4-HDPA and 1,3-DMBA, were formed in water, indicating a common oxidative chain cleavage reaction mechanism in both water and gaseous ozonation conditions.

To the best of our knowledge, this investigation covered major TPs formed from 6PPD abiotic reactions in the gas or aqueous phases. The possibility for significant microbial or redox

mediated processes also contributing to environmental transformations and 6PPD TPs should also be investigated. Given the long-lived character of rubbers and the slow diffusion of chemicals from these phases, 6PPD may continue to act as a source of such TPs, including potentially bioactive or bio-accumulative compounds, to surface water, soil and air particles over long time periods (Cao et al., 2022; Du et al., 2022; Huang et al., 2021; Klöckner et al., 2021a; Kole et al., 2017, 2015). However, to the best of our knowledge, the environmental fate and transport, ecological effect, exposure risks and human health effect of most 6PPD TPs remain largely uncharacterized. In particular, a pressing need exists for structural characterization and risk assessment for the suite of 6PPD TPs derived from different environmental or treatment processes. In addition, further characterization of 6PPD reactivity under other environmental conditions, including thermal stability, biological processes, or photo-transformation would provide critical information to inform tire rubber composition and facilitate management of tire rubber residuals over product lifetimes.

References

Beeson, S., n.d. Pollution Discharge from Hurricane Florence: Examining how North Carolina is Impacted by Modern day Storms 36.

Blum, K.M., Andersson, P.L., Renman, G., Ahrens, L., Gros, M., Wiberg, K., Haglund, P., 2017. Non-target screening and prioritization of potentially persistent, bioaccumulating and toxic domestic wastewater contaminants and their removal in on-site and large-scale sewage treatment plants. Sci. Total Environ. 575, 265–275. https://doi.org/10.1016/j.scitotenv.2016.09.135

Cao, G., Wang, W., Zhang, J., Wu, P., Zhao, X., Yang, Z., Hu, D., Cai, Z., 2022. New Evidence of Rubber-Derived Quinones in Water, Air, and Soil. Environ. Sci. Technol. 56, 4142–4150. https://doi.org/10.1021/acs.est.1c07376

Carpenter, C.M.G., Wong, L.Y.J., Johnson, C.A., Helbling, D.E., 2019. Fall Creek Monitoring Station: Highly Resolved Temporal Sampling to Prioritize the Identification of Nontarget Micropollutants in a Small Stream. Environ. Sci. Technol. 53, 77–87. https://doi.org/10.1021/acs.est.8b05320

Cataldo, F., Faucette, B., Huang, S., Ebenezer, W., 2015. On the early reaction stages of ozone with N,N′-substituted p-phenylenediamines (6PPD, 77PD) and N,N′,N″-substituted-1,3,5-triazine "Durazone®": An electron spin resonance (ESR) and electronic absorption spectroscopy study. Polymer Degradation and Stability 111, 223–231. https://doi.org/10.1016/j.polymdegradstab.2014.11.011

C.H. Kuo, 1985. Reactions of ozone with organics in aqueous solutions.

Challis, J.K., Popick, H., Prajapati, S., Harder, P., Giesy, J.P., McPhedran, K., Brinkmann, M., 2021. Occurrences of Tire Rubber-Derived Contaminants in Cold-Climate Urban Runoff. Environ. Sci. Technol. Lett. 8, 961–967. https://doi.org/10.1021/acs.estlett.1c00682

Councell, T.B., Duckenfield, K.U., Landa, E.R., Callender, E., 2004. Tire-Wear Particles as a Source of Zinc to the Environment. Environ. Sci. Technol. 38, 4206–4214. https://doi.org/10.1021/es034631f

Deborde, M., Rabouan, S., Duguet, J.-P., Legube, B., 2005. Kinetics of Aqueous Ozone-Induced Oxidation of Some Endocrine Disruptors. Environ. Sci. Technol. 39, 6086–6092. https://doi.org/10.1021/es0501619

Dom, I., Biré, R., Hort, V., Lavison-Bompard, G., Nicolas, M., Guérin, T., 2018. Extended Targeted and Non-Targeted Strategies for the Analysis of Marine Toxins in Mussels and Oysters by (LC-HRMS). Toxins 10, 375. https://doi.org/10.3390/toxins10090375

Du, B., Liang, B., Li, Y., Shen, M., Liu, L.-Y., Zeng, L., 2022. First Report on the Occurrence of N-(1,3-Dimethylbutyl)-N′-phenyl-p-phenylenediamine (6PPD) and 6PPD-Quinone as Pervasive Pollutants in Human Urine from South China. Environmental Science & Technology Letters. https://doi.org/10.1021/acs.estlett.2c00821

Hiki, K., Asahina, K., Kato, K., Yamagishi, T., Omagari, R., Iwasaki, Y., Watanabe, H., Yamamoto, H., 2021. Acute Toxicity of a Tire Rubber-Derived Chemical, 6PPD

Quinone, to Freshwater Fish and Crustacean Species. Environ. Sci. Technol. Lett. https://doi.org/10.1021/acs.estlett.1c00453

Hoigné, J., Bader, H., 1983. Rate constants of reactions of ozone with organic and inorganic compounds in water—I: Non-dissociating organic compounds. Water Research 17, 173–183. https://doi.org/10.1016/0043-1354(83)90098-2

Hou, F., Tian, Z., Peter, K.T., Wu, C., Gipe, A.D., Zhao, H., Alegria, E.A., Liu, F., Kolodziej, E.P., 2019. Quantification of organic contaminants in urban stormwater by isotope dilution and liquid chromatography-tandem mass spectrometry. Anal Bioanal Chem 411, 7791–7806. https://doi.org/10.1007/s00216-019-02177-3

Hu, X., Zhao, H.N., Tian, Z., Peter, K.T., Dodd, M.C., Kolodziej, E.P., 2022. Transformation Product Formation upon Heterogeneous Ozonation of the Tire Rubber Antioxidant 6PPD (N-(1,3-dimethylbutyl)-N′-phenyl-p-phenylenediamine). Environ. Sci. Technol. Lett. 9, 413–419. https://doi.org/10.1021/acs.estlett.2c00187

Huang, W., Shi, Y., Huang, J., Deng, C., Tang, S., Liu, X., Chen, D., 2021. Occurrence of Substituted p-Phenylenediamine Antioxidants in Dusts. Environmental Science & Technology Letters. https://doi.org/10.1021/acs.estlett.1c00148

Huntink, N.M., 2003. Durability of rubber products: Development of new antidegradants for long-term protection.

Jin, X., Peldszus, S., Huck, P.M., 2012. Reaction kinetics of selected micropollutants in ozonation and advanced oxidation processes. Water Research 46, 6519–6530. https://doi.org/10.1016/j.watres.2012.09.026

Johannessen, C., Helm, P., Metcalfe, C.D., 2021. Detection of selected tire wear compounds in urban receiving waters. Environmental Pollution 287, 117659. https://doi.org/10.1016/j.envpol.2021.117659

Klöckner, P., Seiwert, B., Wagner, S., Reemtsma, T., 2021a. Organic Markers of Tire and Road Wear Particles in Sediments and Soils: Transformation Products of Major Antiozonants as Promising Candidates. Environ. Sci. Technol. https://doi.org/10.1021/acs.est.1c02723

Klöckner, P., Seiwert, B., Weyrauch, S., Escher, B.I., Reemtsma, T., Wagner, S., 2021b. Comprehensive characterization of tire and road wear particles in highway tunnel road dust by use of size and density fractionation. Chemosphere 279, 130530. https://doi.org/10.1016/j.chemosphere.2021.130530

Kole, P.J., Löhr, A.J., Ragas, A.M.J., 2015. Autobandenslijtstof: een verwaarloosde bron van microplastics? Milieu : opinieblad van de Vereniging van Milieuprofessionals 39–41.

Kole, P.J., Löhr, A.J., Van Belleghem, F.G.A.J., Ragas, A.M.J., 2017. Wear and Tear of Tyres: A Stealthy Source of Microplastics in the Environment. Int J Environ Res Public Health 14. https://doi.org/10.3390/ijerph14101265

Lattimer, R.P., Hooser, E.R., Layer, R.W., Rhee, C.K., 1983. Mechanisms of Ozonation of N-(1,3-Dimethylbutyl)-N′-Phenyl-p-Phenylenediamine. Rubber Chemistry and Technology 56, 431–439. https://doi.org/10.5254/1.3538136

Leitzke, A., Reisz, E., Flyunt, R., von Sonntag, C., 2001. The reactions of ozone with cinnamic acids: formation and decay of 2-hydroperoxy-2-hydroxyacetic acid. J. Chem. Soc., Perkin Trans. 2 793–797. https://doi.org/10.1039/b009327k

Lin, Y., Sevillano-Rivera, M., Jiang, T., Li, G., Cotto, I., Vosloo, S., Carpenter, C.M.G., Larese-Casanova, P., Giese, R.W., Helbling, D.E., Padilla, I.Y., Rosario-Pabón, Z., Vélez Vega, C., Cordero, J.F., Alshawabkeh, A.N., Pinto, A., Gu, A.Z., 2020. Impact of Hurricane Maria on Drinking Water Quality in Puerto Rico. Environ. Sci. Technol. 54, 9495–9509. https://doi.org/10.1021/acs.est.0c01655

Lutz, I., Kloas, W., 1999. Amphibians as a model to study endocrine disruptors: I. Environmental pollution and estrogen receptor binding. Sci Total Environ 225, 49–57. https://doi.org/10.1016/s0048-9697(99)80016-3

Noguera-Oviedo, K., Aga, D.S., 2016. Lessons learned from more than two decades of research on emerging contaminants in the environment. Journal of Hazardous Materials 316, 242–251. https://doi.org/10.1016/j.jhazmat.2016.04.058

OSPAR Commission, 2006c. OSPAR background document on 4-(dimethylbutylamino)diphenylamine (6PPD).

Peter, K.T., Hou, F., Tian, Z., Wu, C., Goehring, M., Liu, F., Kolodziej, E.P., 2020. More Than a First Flush: Urban Creek Storm Hydrographs Demonstrate Broad Contaminant Pollutographs. Environ. Sci. Technol. 54, 6152–6165. https://doi.org/10.1021/acs.est.0c00872

Peter, K.T., Tian, Z., Wu, C., Lin, P., White, S., Du, B., McIntyre, J.K., Scholz, N.L., Kolodziej, E.P., 2018. Using High-Resolution Mass Spectrometry to Identify Organic Contaminants Linked to Urban Stormwater Mortality Syndrome in Coho Salmon. Environ. Sci. Technol. 52, 10317–10327. https://doi.org/10.1021/acs.est.8b03287

Seiwert, B., Nihemaiti, M., Troussier, M., Weyrauch, S., Reemtsma, T., 2022. Abiotic oxidative transformation of 6-PPD and 6-PPD quinone from tires and occurrence of their products in snow from urban roads and in municipal wastewater. Water Research 212, 118122. https://doi.org/10.1016/j.watres.2022.118122

Sheridan, M., 2010. The Vanderbilt Rubber Handbook, 14th edition. R.T. Vanderbilt Company, Inc., Norwalk, CT.

Spahr, S., Teixidó, M., Sedlak, D.L., Luthy, R.G., 2019. Hydrophilic trace organic contaminants in urban stormwater: occurrence, toxicological relevance, and the need to enhance green stormwater infrastructure. Environ. Sci.: Water Res. Technol. 6, 15–44. https://doi.org/10.1039/C9EW00674E

Test No. 111: Hydrolysis as a Function of pH | OECD Guidelines for the Testing of Chemicals, Section 1 : Physical-Chemical properties | OECD iLibrary [WWW Document], n.d. URL https://www.oecd-ilibrary.org/environment/test-no-111-hydrolysis-as-a-function-of-ph_9789264069701-en (accessed 4.4.23).

Tian, Z., Gonzalez, M., Rideout, C., Zhao, H., Hu, X., Wetzel, J., Mudrock, E., James, C.A., McIntyre, J.K., Kolodziej, E.P., 2021a. 6PPD-Quinone: Revised Toxicity Assessment and Quantification Method Development with a Commercial Standard. Environmental Science & Technology Letters.

Tian, Z., Zhao, H., Peter, K.T., Gonzalez, M., Wetzel, J., Wu, C., Hu, X., Prat, J., Mudrock, E., Hettinger, R., Cortina, A.E., Biswas, R.G., Kock, F.V.C., Soong, R., Jenne, A., Du, B., Hou, F., He, H., Lundeen, R., Gilbreath, A., Sutton, R., Scholz, N.L., Davis, J.W., Dodd, M.C., Simpson, A., McIntyre, J.K., Kolodziej, E.P., 2021b. A ubiquitous tire rubber–derived chemical induces acute mortality in coho salmon. Science 371, 185–189. https://doi.org/10.1126/science.abd6951

US EPA, O., 2014. Air Data: Air Quality Data Collected at Outdoor Monitors Across the US [WWW Document]. US EPA. URL https://www.epa.gov/outdoor-air-quality-data (accessed 6.14.21).

US EPA, R. 01, n.d. Eight-hour Average Ozone Concentrations | Ground-level Ozone | New England | US EPA [WWW Document]. URL https://www3.epa.gov/region1/airquality/avg8hr.html (accessed 5.17.21).

Wagner, S., Hüffer, T., Klöckner, P., Wehrhahn, M., Hofmann, T., Reemtsma, T., 2018. Tire wear particles in the aquatic environment - A review on generation, analysis, occurrence, fate and effects. Water Res 139, 83–100. https://doi.org/10.1016/j.watres.2018.03.051

Wang, W., Cao, G., Zhang, J., Wu, P., Chen, Y., Chen, Z., Qi, Z., Li, R., Dong, C., Cai, Z., 2022. Beyond Substituted p-Phenylenediamine Antioxidants: Prevalence of Their Quinone Derivatives in PM2.5. Environ. Sci. Technol. 56, 10629–10637. https://doi.org/10.1021/acs.est.2c02463

Wylie, A.D.L., Abbatt, J.P.D. 2020. Heterogeneous Ozonolysis of Tetrahydrocannabinol: Implications for Thirdhand Cannabis Smoke. Environ Sci Technol. 17;54(22):14215-14223. doi: 10.1021/acs.est.0c03728.

Xu, Q., Li, G., Fang, L., Sun, Q., Han, R., Zhu, Z., Zhu, Y.G. 2023. Enhanced Formation of 6PPD-Q during the Aging of Tire Wear Particles in Anaerobic Flooded Soils: The Role of Iron Reduction and Environmentally Persistent Free Radicals. Environ Sci Technol. 57(14):5978-5987. doi: 10.1021/acs.est.2c08672.

Zhang, Y., Xu, C., Zhang, W., Qi, Z., Song, Y., Zhu, L., Dong, C., Chen, J., Cai, Z., 2021. p-Phenylenediamine Antioxidants in PM2.5: The Underestimated Urban Air Pollutants. Environ. Sci. Technol. https://doi.org/10.1021/acs.est.1c04500

Zhao, H.N., Hu, X., Gonzalez, M., Rideout, C.A., Hobby, G.C., Fisher, M.F., McCormick, C.J., Dodd, M.C., Kim, K.E., Tian, Z., Kolodziej, E.P., 2023a. Screening p-Phenylenediamine Antioxidants, Their Transformation Products, and Industrial Chemical Additives in Crumb Rubber and Elastomeric Consumer Products. Environ. Sci. Technol. 57, 2779–2791. https://doi.org/10.1021/acs.est.2c07014

Zhao, H.N., Hu, X., Tian, Z., Gonzalez, M., Rideout, C.A., Peter, K.T., Dodd, M.C., Kolodziej, E.P., 2023b. Transformation Products of Tire Rubber Antioxidant 6PPD in Heterogeneous Gas-Phase Ozonation: Identification and Environmental Occurrence. Environ. Sci. Technol. https://doi.org/10.1021/acs.est.2c08690

Zhou, S., Hwang, B.C.H., Lakey, P.S.J., Zuend, A., Abbatt, J.P.D., Shiraiwa, M. 2019. Multiphase reactivity of polycyclic aromatic hydrocarbons is driven by phase

separation and diffusion limitations. Proc Natl Acad Sci U S A. 116(24):11658-11663. doi: 10.1073/pnas.1902517116.

Supplementary Materials

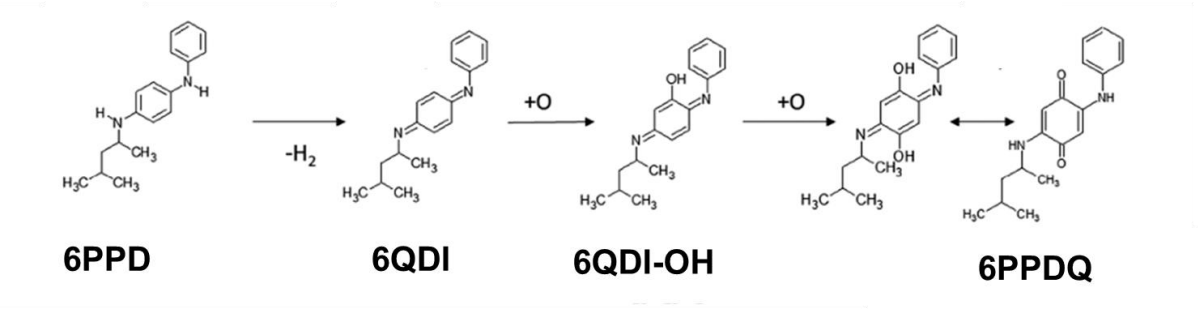
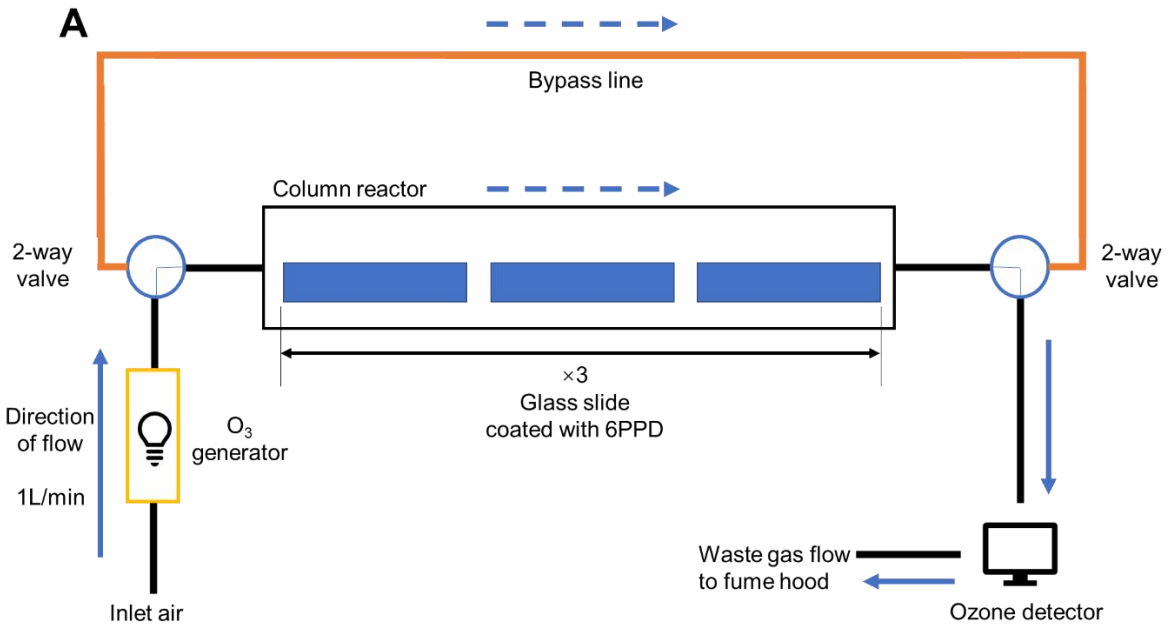


Figure S4.1. Proposed formation reactions of 6PPD to 6PPDQ. Adapted and updated from Seiwert et al. (Seiwert et al., 2021a). Presented structures are proposed based on observed fragment ions and plausibility as described in the original study.



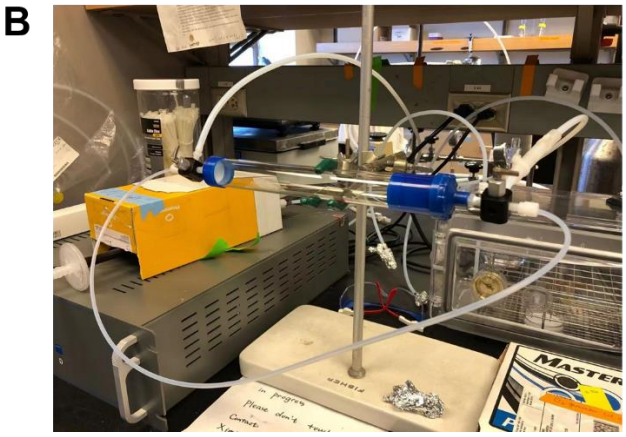


Figure S4.2. Schematic diagrams of experimental system setup, adapted from a previous study (Hu et al., 2022). (A) 6PPD or 6PPDQ coated slide ozonation setup (B) Photo of the ozonation column and associated valve and tubing connection. All tubing in direct contact with O_3 was made from PTFE. The chromatography columns were constructed of borosilicate glass with polypropylene inlets and outlets.

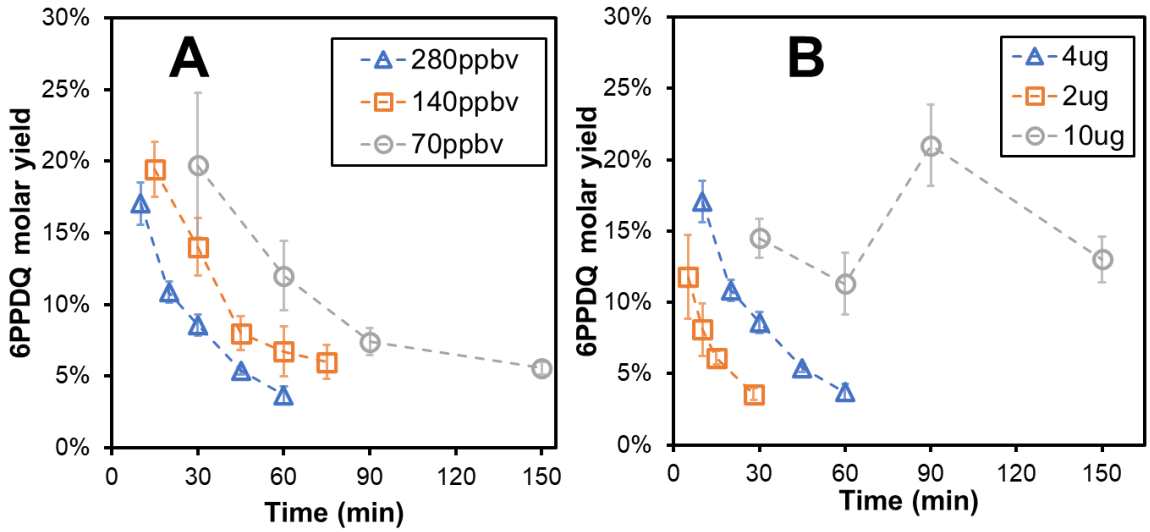


Figure S4.3. 6PPDQ molar yield vs. exposure time. (A) sample groups of 4 μg/plate 6PPD mass load tested with different inlet O_3 concentration and (B) different 6PPD mass load groups tested under 280 ppbv inlet O_3 concentration. The error bars in each panel represent the standard deviation from three replicate readings. The first point from panel B for the 10 μg/plate group was taken out as an outlier.

Appendix A 6QDI purity analysis with LC-MS instrumentations

The results and discussions were adapted as from Zhao et al. (Zhao et al. 2023). Notably, during the 6QDI sample analysis with LC-qTOF-HRMS and LC-MS/MS, large 6PPD signals (~1/3 peak area of 6QDI on LC-qTOF-HRMS; ~10 times peak area of 6QDI on LC-MS/MS; **Figure A1**) were observed (C18 column, 24.5 min LC gradient with 0.1% FA in both H2O and MeOH; ~6-7 min retention time). To inspect standard purity, we performed a series of "semi-direct LC-MS/MS injections" by connecting the autosampler needle seat directly into the MS nebulizer with a PEEK tubing (a syringe pump was not available at the facility for direct infusions). 6PPD and 6QDI standards (200 µg/L nominal concentrations; 6PPD prepared from solid standard, 6QDI prepared by diluting the commercial stock) were each prepared in duplicate in (a) ACN, (b) MeOH with 0.1% formic acid (FA), (c) 50/50 H2O/MeOH with 0.1% FA, and (d) 95/5 H2O/MeOH with 0.1% FA. The standards were injected (5 µL) with 0.5 mL/min pure acetonitrile as the mobile phase; peaks eluted within 0.2 minutes after injection (minimal reaction time on LC).

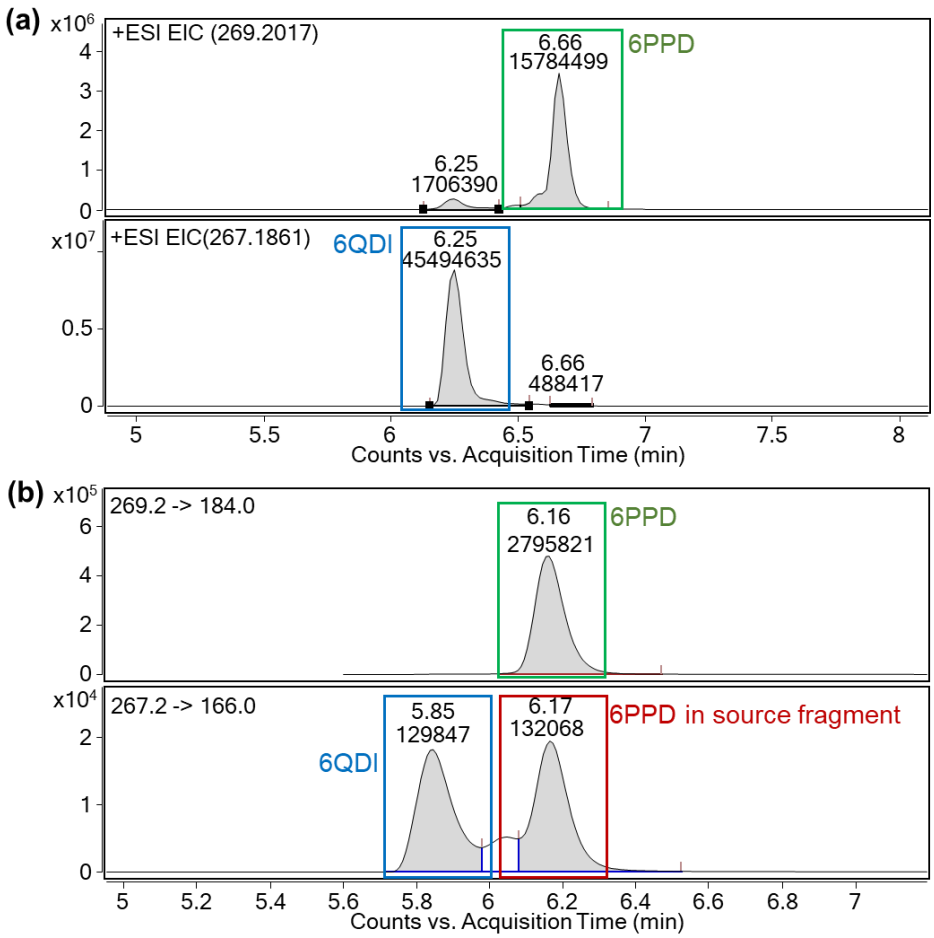


Figure A1. Peaks corresponding to 6PPD and 6QDI when analyzing 6QDI standards by (a) LC-qTOF-HRMS (full scan MS) and (b) LC-MS/MS. Peaks corresponding to 6PPD are highlighted with green boxes (see the peak at ~6.7 min on the extracted ion chromatogram (EIC) for *m/z* 269 on LC-HRMS in panel (a), and at ~6.2 min on the chromatogram for the 269 -> 184 transition on LC-MS/MS in panel (b), where each peak was confirmed as 6PPD by matching RT to an authentic 6PPD standard). The peak corresponding to the in-source fragment of 6PPD with MS spectrum matching 6QDI is highlighted with a red box (see the peak at ~6.2 min on the chromatogram for the 267 -> 166 transition on LC-MS/MS in panel (b)). Peaks corresponding to 6QDI are highlighted with blue boxes (see the peak at ~6.3 min on the EIC for *m/z* 267 on LC-HRMS in panel (a), and at ~5.9 min on the chromatogram for the 267 -> 166 transition on LC-MS/MS in panel (b)).

For 6QDI standards prepared in pure ACN, the peak area ratio of 6PPD transition (269.2 -> 184.0) over that of 6QDI transition (267.2 -> 166.0) was only 0.07 (Table S9). The peak area ratio was similar in 6QDI standards prepared in MeOH with 0.1% FA (peak area ratio of 0.09) but increased to 0.23 in 50/50 H2O/MeOH and to 0.43 in 95/5 H2O/MeOH (with 0.1% FA; Table S9). The results indicated that 6QDI converts to 6PPD in acidic aqueous conditions and were consistent with the large 6PPD response observed when analyzing 6QDI standards under regular LC-MS/MS analysis (with C18 column, 24.5 min LC gradient with 0.1% FA in both H2O and MeOH; ~6-7 min retention time). This is also consistent with our personal communication with HPC Inc (commercial vender of the 6QDI standard) where they also observed 6PPD when analyzing 6QDI stocks and indicated an apparently important role of water in this reaction. Furthermore, our analyses and HPC's communications are consistent with HPC specifications of ≥96% for the 6QDI stock purity.

Notably, samples obtained from 6QDI ozonation experiments were processed using only pure organic solvents prior to LC-MS analyses. Considering the high purity of 6QDI stock solutions (prepared in ACN) certified by HPC, and the apparent stability of 6QDI in ACN and MeOH noted above, we therefore hypothesize that 6PPD observed when analyzing samples from 6QDI ozonation experiments arose as an artifact during analyses, and 6PPD was not present at substantial concentrations (<5% based on HPC specifications; <~10% based on peak areas in our "semi-direct LC-MS/MS injections") in 6QDI stock solutions or ozonation samples themselves. Thus, the TPs observed in ozonated 6QDI can be taken to represent TP formation from 6QDI itself, with negligible contributions from 6PPD.

In “semi-direct injections” of 6PPD standards prepared in each of the four solvent matrixes described above, 6QDI signals were also observed, with peak area ratios of 6QDI transition (267.2 -> 166.0) over 6PPD transition (269.2 -> 184.0) observed to be stable at 0.29 ± 0.05. This 6QDI response observed when analyzing 6PPD standards could come from (1) impurities of the industrial 6PPD standards (95% purity), or (2) in-source fragmentation of 6PPD (**Figure A1**).

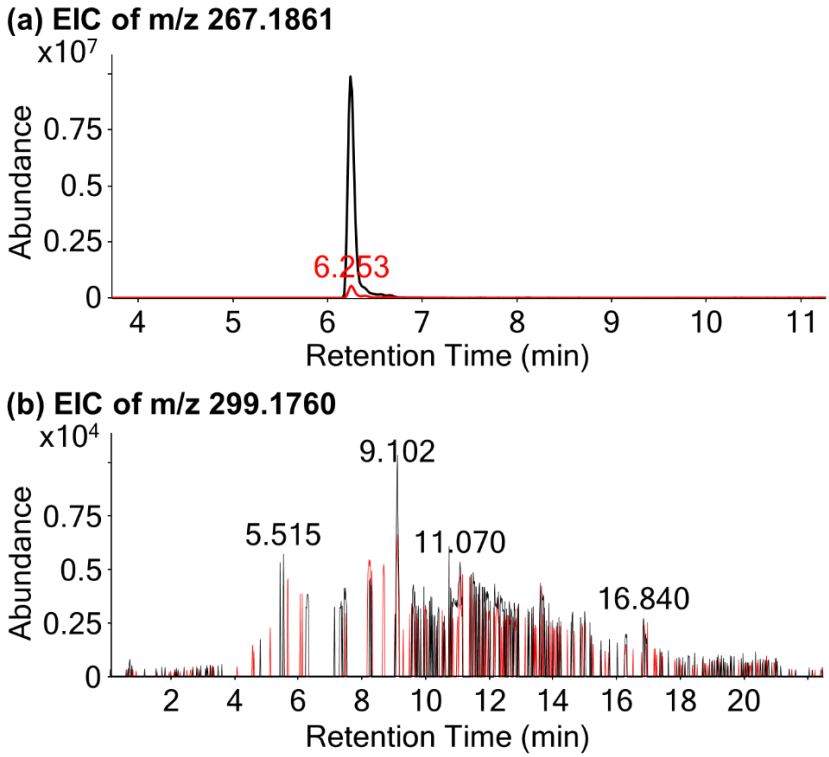


Figure A2. Extracted ion chromatograms of *m/z* values corresponding to (a) 6QDI (*m/z* 267.1861) and (b) 6PPDQ (*m/z* 299.1760) on LC-qTOF-HRMS in samples obtained during 6QDI ozonation (2 µg 6QDI mass, 0.25 µg 6QDI/cm^2, 289 ± 4 ppbv O_3 generated from ambient air with a photochemical O_3 generator). Black and red chromatograms represent 0-hr and 3-hr of O_3 exposure, respectively.

Chapter 5: Mass-Suite: A Novel Open-source Python Package for High Resolution Mass Spectrometry Data Analysis

5.0 Publication and Contribution statement

One publication was incorporated for this chapter as below, and XH led most of the research and development effort. For Hu et al. 2022 (in review), XH led the development team, designed and developed the software package major features and wrote the manuscript draft.

1. X. Hu, D. Mar, N. Suzuki, B. Zhang, K. T. Peter, D. A. C. Beck, E. P. Kolodziej, Mass-Suite: A Novel Open-source Python Package for High Resolution Mass Spectrometry Data Analysis, submitted to *Journal of Cheminformatics*, Nov. 2022.

5.1 Introduction

High-resolution mass spectrometry (HRMS) analyses provide especially comprehensive and open-ended screening capabilities to characterize complex samples containing mixtures of many unknown or unanticipated compounds. With increasing recognition that humans produce and discharge many thousands of potential new "emerging contaminants" to the environment (Hollender et al., 2014; Tian et al., 2020; Wang et al., 2020), these broad spectrum analytical methods are opening new frontiers in environmental chemistry, health, and engineering research. Specifically, non-targeted analysis (NTA) methods leverage the non-selective data collection capability of HRMS, with the resulting data supporting comprehensive characterization of chemical composition, identification of previously unknown contaminants, evaluation of compositional change across samples, and tracking of contaminant sources

(Dávila-Santiago et al., 2022; Gonzalez de Vega et al., 2021; Peter et al., 2019; Schollée et al., 2018; Tian et al., 2021; Wang et al., 2021). Such data uses are not unique to environmental analysis, with many applications relevant to multi-omics (Naz et al., 2014; Uppal et al., 2013; Yin and Xu, 2014), toxicology, and drug screening studies (Dom et al., 2018; Rosano et al., 2011; Tkalec et al., 2022; Wu et al., 2012).

Notably, pairing complex environmental samples with expansive HRMS data collection capacities results in generation of massive datasets; most such data remain under- or unused, in part due to limitations of existing data analysis workflows. HRMS data analysis workflows and software platforms incorporate data reduction, analysis, and interpretation elements, but significant opportunity remains for optimization and development of advanced data analysis capabilities, particularly for NTA data sets and for data interpretation endpoints beyond compound identification. Existing commercial software supports both basic data analysis (e.g., feature extraction, data alignment) and several advanced workflows (e.g., feature annotation, statistical analyses), but often are costly, limited to instrument-specific datafile formats, or provide outputs that struggle to interface with other platforms, databases, and tools. It is especially difficult for users to adapt existing workflows to integrate more complex approaches to feature prioritization, source tracking, or "-omics" analyses that require external functions or algorithms (e.g., machine learning, external database searching, or cloud computation). To address such needs, various open-source tools for handling HRMS data and implementing NTA workflows have been developed, including *MSDIAL* (Tsugawa et al., 2015), *openMS* (Röst et al., 2016), *XCMS* (Smith et al., 2006), *MZmine* (Pluskal et al., 2010), *PatRoon* (Helmus et al.,

2021), and *enviMass* (Schmitt, 2018), among others. Those tools provide flexible workflows with designated functionalities within specific intended fields (e.g., proteomics or metabolomics), but often implement a limited range of data analysis capabilities (**Table 5.1**) or are not useful for some types of environmental data analysis (e.g., source apportionment).

Table 5.1. Overview of commonly used open-source software tools and their data analysis features for HRMS workflows.

Features \ Tools	***MSS***	**TidyMS**	**MZmine2**	**XCMS***	**MSDIAL**	**PatRoon**
Language	Python	Python	Java	R	C#	R
Raw data preprocessing	√	√	√	√	√	√
QC-based batch correction	×	√	×	×	×	×
Quality reports	√	√	√	×	√	√
Normalization, imputation, scaling	√	√	√	×	√	√
Feature annotation	√	×	√	×	√	√
Isotope grouping	×	×	√	√	√	√
Interactive visualization plots	√	×	√	×	√	×
Clustering statistical analysis	√	×	×	×	×	×
Modeling tools	√	×	×	×	×	×

*XCMS is supported by various R packages and primarily acts as a starting point for subsequent analyses on other platforms.

Raw HRMS data often consists of many thousands of features, requiring substantial computational resources and potentially driving inaccuracy in subsequent analyses if used directly. Therefore, feature filtering and prioritization are critical to effectively reduce the size of the dataset and facilitate downstream analyses (Blum et al., 2017; Du et al., 2017; Schollée et al., 2018; Tian et al., 2021). To avoid inefficient or impractical manual operations (e.g., to

remove poorly integrated chromatogram peaks, to prioritize certain HRMS features) and facilitate data mining analysis, existing software platforms (e.g., Compound Discoverer, patRoon) commonly rely on descriptive statistics, data reduction, and data visualization (e.g., Principle Component Analysis (PCA), fold-change volcano plots) (Helmus et al., 2021). As a complementary automated approach, machine learning algorithms (e.g., supervised, unsupervised, or reinforcement learnings, etc.) can support more effective feature prioritization and predictive modeling workflows across several fields. For example, Nikolopoulou et al. developed a deep learning-based NTA workflow for environmental trend analysis to prioritize new emerging contraminants (Nikolopoulou et al., 2022). In metabolomics applications, machine learning algorithms can support clinical decisions, guide metabolic engineering, and facilitate biological studies (Chen et al., 2022; Iravani and Conrad, 2022; Lee and Durant, 2022; Liebal et al., 2020). However, existing workflows usually employ only one or a few algorithms concurrently, forcing users to jump back and forth between different platforms to achieve some analysis capabilities.

Currently, only a few software packages (e.g., *PatRoon*, *enviMass*) are specifically designed to address environmental NTA data analysis challenges. For example, identification and quantitative apportionment of complex chemical pollution sources remains a persistent challenge (Dávila-Santiago et al., 2022; Peter et al., 2019). Traditionally, contaminant source apportionment (i.e., estimating the presence and relative amount of a source in a mixed sample) has relied on the occurrence and quantification of a few pre-selected, targeted chemicals as unique source markers (Fauser et al., 1999; Rødland et al., 2022). However, source marker

chemicals are not always known or unique to individual sources. HRMS datasets provide a unique opportunity to establish source "fingerprints" comprised of hundreds to thousands of both identified and unknown chemical features (Peter et al., 2019), which are more likely to be source-specific and to contain marker chemicals that persist through dilution and transformation processes. Conceptually, this approach enables complex mixture quantitation and represents an important, cutting-edge analytical capability (Hu et al., 2021; Peter et al., 2018; Rager et al., 2016; Xue et al., 2019). However, few existing efforts have paired this concept with machine learning, indicating a clear opportunity for NTA workflows (Dávila-Santiago et al., 2022).

Finally, most open-source HRMS data tools were developed using R, C++, and Visual Basic programming languages, while relatively few software packages use Python (Bittremieux, 2020; Levitsky et al., 2019; Melnikov et al., 2020; Riquelme et al., 2020; Yunker et al., 2018). As one of the most popular and accessible programming languages, Python especially benefits from community contributions, including the well-known statistical analysis packages SciPy (Virtanen et al., 2020) and scikit-learn (Pedregosa et al., 2011). Additionally, Python is an interpreted programming language that is relatively easy to read, learn, and write for non-programmer researchers, providing much flexibility and convenience for users to optimize and adapt existing tools to their needs ("Python vs Java," n.d.).

Given these many data analysis needs and the limitations of existing software packages, we developed a Python package *Mass-Suite (MSS)* as an open-source data analysis toolbox with multiple HRMS data processing capabilities. The *MSS* package described here is

compatible with exported data from other commercial or open-source tools and includes basic functions like feature extraction, prioritization, and data visualization. Driven by machine learning algorithms and capabilities that are not currently available within other NTA workflows or tools (**Table 5.1**), *MSS* also provides advanced data analysis (e.g., unsupervised clustering analysis, source tracking modeling), heuristic data exploration, data mining, and predictive modeling capabilities within a user-friendly, automated, and full-stack platform. We anticipate *MSS* will enable researchers, especially those with limited programming expertise, to more efficiently and reliably extract meaningful information from NTA datasets.

5.2 Implementations

Development of *MSS* primarily depended on *Pandas* (McKinney, 2010) and *scikit-learn* (Pedregosa et al., 2011) packages for data processing and analysis, and *plotly* (Plotly Technologies Inc., 2015) and *matplotlib* ("Matplotlib: A 2D Graphics Environment | IEEE Journals & Magazine | IEEE Xplore," n.d.) packages for data visualization. To demonstrate major *MSS* functionalities, this Implementation section describes a representative NTA workflow using *MSS* for data import, feature extraction and alignment, data reduction, advanced data mining (statistical analyses, feature clustering, and a novel source tracking function), data visualization, feature annotation, and reporting. In the Results and Discussion section, we describe workflow performance validation to assess peak picking accuracy and feature detection consistency relative to other open-source data processing platforms. Three example case studies are then provided to illustrate application of *MSS* to analyze existing

experimental HRMS datasets. All related resources and an example workflow (in .ipynb format) are included in the demo file in the project GitHub repository: https://github.com/XiminHu/mass-suite. A README file accompanies *MSS* and most functions have individual documentation to ensure that *MSS* is readable and maintainable.

5.2.1 Workflow development

MSS uses a modularized layout to provide HRMS data analysis functions (**Figure 5.1**) and all *MSS* modules can be loaded in full or separately as needed.

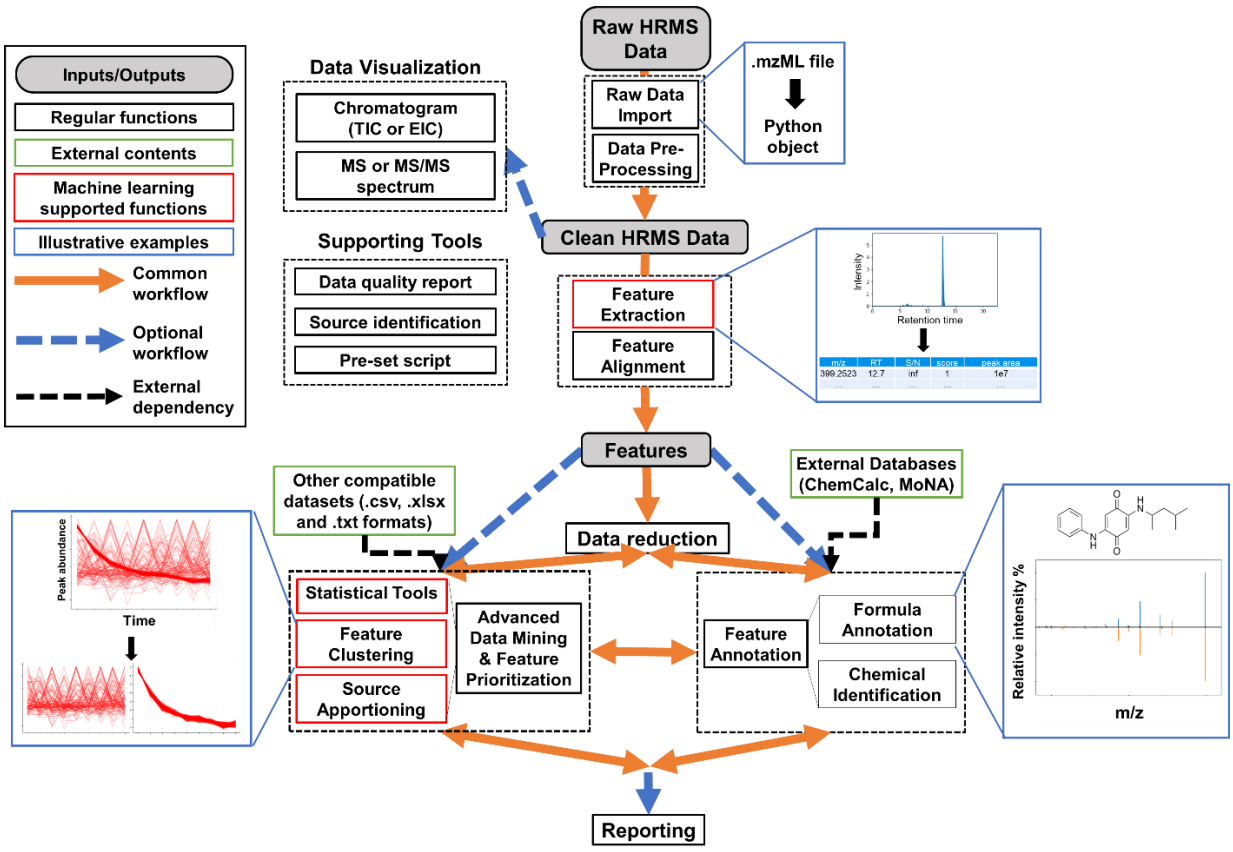


Figure 5.1. Overview of a typical *MSS* workflow for high-resolution mass spectrometry (HRMS) data analysis. The solid lines represent a typical workflow for typical HRMS non-targeted analysis (NTA) data processing; dashed lines represent additional optional workflows. All modules are optional.

Existing modules enable raw data import/pre-processing, feature extraction and alignment, data reduction, feature annotation, advanced data mining and feature prioritization, data visualization, and reporting. These functions and capabilities are summarized below and in **Table 5.1**. All package modules are optional, customizable, and compatible with various external data formats (e.g., .csv, .xlsx, .txt), enabling users to select and combine functions from different modules, external packages, and other platforms to create custom workflows.

Table 5.1. Major modules currently within *MSS* and related functions.

Module	Function	Description
mssmain	get_scans	Data import - HRMS data import
	batch_scans	Data import - Import multiple HRMS data
	noise_removal	Data reduction - HRMS data pre-processing, remove noise detections within each scan
	peak_pick	Feature extraction - peak picking from imported data using defined parameters
	peak_list	Feature extraction - generate a feature list based on user-defined settings
	formula_prediction	Data reduction - predicting selected feature's formula with isotopic pattern scoring
align	mss_align	Feature alignment - align the feature lists from different samples
	mss_process	Data import, feature extraction and alignment - automatic workflow from HRMS data import to MS peak list generation
visreader	tic_plot	Visualization - present the TIC plot for selected data
	ms_spectrum/frag_plot	Visualization - present the selected spectrum indexing by numbers
	ms_chromatogram	Visualization - generate a chromatogram plot for selected features
	integration_plot	Visualization - generate an integrated chromatogram plot for peak area checking
	iso_plot	Visualization - present the MS1 spectrum for selected features with the theoratical MS1 spectrum from input formula for comparison
	overview_scatter	Visualization - provide a scatter plot of all the features within the data

frag	frag_massbank	Annotation - search online database for the input fragment list
	frag_comp	Annotation - present the MS2 spectrum for selected precursors with optional online database search and comparison
dm	data_prep	Data mining - data cleaning and preliminary filters for subsequent analysis
	PCA_report	Data mining - perform a PCA on selected dataset
	eps_assess	Data mining - assisting tool for the DBSCAN clustering analysis
	ms_cluster	Data mining - unsupervised clustering on HRMS data, dimension reduction is optional
	trend_calc	Data mining - calculate the trend correlation of selected features
	source_label	Data mining - tracking the source from different samples
	feature_model	Data mining - modeling function towards experimental backgrounds
	cluster_pred	Data mining - prediction on new data based on trained models

For example, *MSS* capabilities can complement open-source software packages such as *OpenMS* (Röst et al., 2016), *XCMS* (Smith et al., 2006), *MetFrag* (Ruttkies et al., 2016), *MSDIAL* (Tsugawa et al., 2015), and *PatRoon* (Helmus et al., 2021) by processing exported compatible output files with *MSS* functions. Users can also import .mzML files with the support of external data conversion tools (e.g., ProteoWizard (Kessner et al., 2008), FragPipe (Kong et al., 2017)) to convert raw instrument-specific data formats (e.g., .d, .raw). *MSS* is able to interface with external Python functions or packages, including several popular packages like *SciPy* (Virtanen et al., 2020) or advanced machine learning packages involving neural networks such as *TensorFlow* (Abadi et al., 2016) and *PyTorch* (Paszke et al., 2019). New user-defined functions can easily be appended to existing modules to expand functionality and improve flexibility and data analysis capabilities. Statistical tools provided by *MSS* or *MSS*-interfaced external functions can process *MSS* data or external data in an Excel-compatible format.

5.2.2 Data import, feature extraction, and feature alignment

In a typical *MSS* workflow, raw data (.mzML format) is first imported and parsed with *pymzml* (Till, 2012) as Python-compatible metadata with the *mssmain.get_scans* function. Converted data are then available for optional baseline subtraction based on signal intensities prior to subsequent feature extraction (**Figure 5.2**).

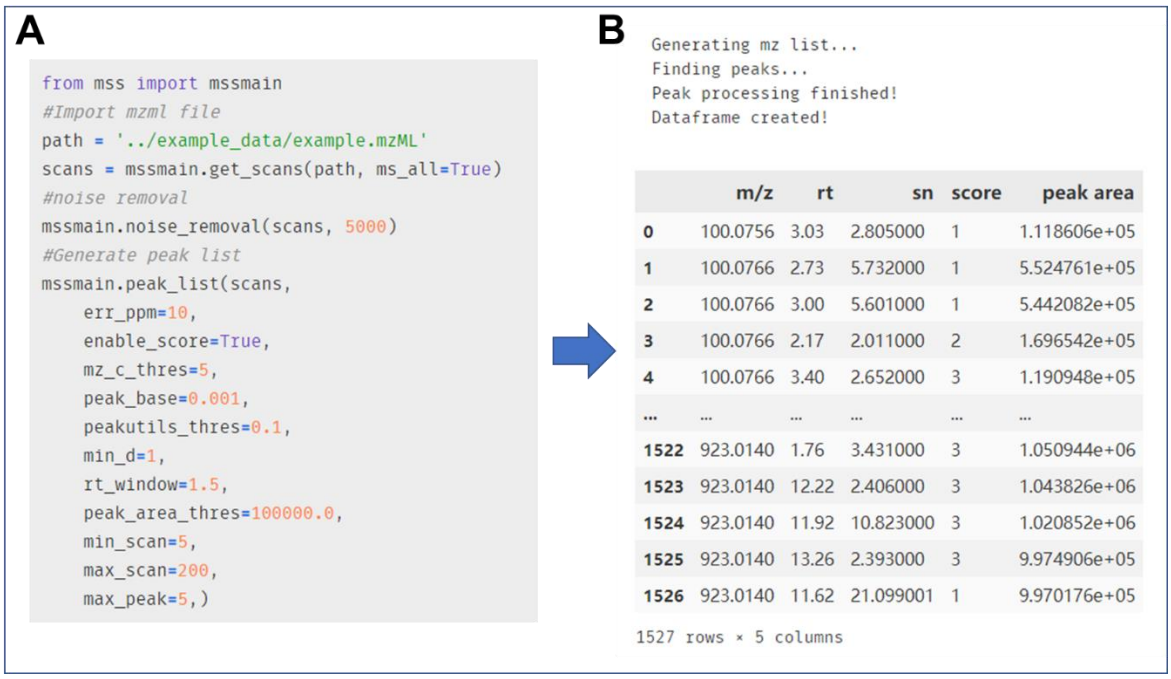


Figure 5.2. Example code and output for *MSS* data acquisition and parsing functions. (A) Example code for .mzML file reading, noise removal and peak list retrieval. All the parameters shown in the example code are default parameters. (B) Example output from the single file processing function *peak_list*. The score of the peak was calculated based on the built-in random forest model within *MSS*.

HRMS feature extraction (i.e., “peak picking”), where a feature is a single presumptive detection of a chemical or its adducts/isotopologues and is represented as an exact mass (*m/z* value) - retention time (RT) pair, usually involves extensive parameter tuning and quality assessment of any extracted peaks. The *MSS* feature extraction function (*mssmain.peak_pick*)

concatenates scans and finds peak indices (using the *PeakUtils* package (Negri and Vestri, 2017)), then performs post-processing (e.g., peak width filter, replicated peak filter, regression-based peak boundary determination) to reduce poor quality features (e.g., peak splitting, insufficient scan numbers, high baseline noise). To further exclude noisy peaks, optimize parameters, and improve the feature extraction accuracy, *MSS* also calculates 15 descriptive parameters for the extracted peaks (Baeza-Baeza et al., 2010) to provide an optional peak assessment score based on a pre-trained random forest model. An optional noise removal function is recommended to reduce the processing time and improve the feature extraction accuracy. The complete feature extraction process for a single .mzML data file as follows:

1. **Concatenate scans:** The closest neighbors (by *m/z*) of function-selected precursors (based on user defined mass ranges, error tolerances) from compiled scans are extracted with corresponding detection intensities. Extracted detections are aligned across all scans to compile a 1-D chromatogram intensity array for the proposed precursor mass. All precursor-associated 1-D arrays are subsequently merged as a chromatogram list.
2. **Find peak index:** Each precursor-associated 1-D chromatogram intensity array is processed individually. The peak indices within each array are located with the *PeakUtils* package, which finds the numerical index of the apex in the input data by taking the first-order difference of the intensity array (Negri and Vestri, 2017).
3. **Determine peak boundaries:** After locating peak apex indices, peak boundaries are determined iteratively. The function traverses in two directions from the estimated peak apex while the boundary data points are iteratively checked that: a) for both left and right

boundaries, signal intensity exceeds the user-defined relative baseline (defined as the ratio of boundary to apex intensities); and b) for the right boundary, every consecutive three points, if they didn't meet the criteria in a), then a linear regression is performed with the three points and the next three consecutive points are checked only if calculated R^2 value >0.6, otherwise, the third point is labeled a boundary point. Additionally, to exclude tails of prolonged peaks and reduce calculation time, peaks that surpass a user-defined maximum peak width are trimmed. After feature extraction, additional filters exclude duplicate detections (e.g., multiple noise peaks, overlapping detections). All numerical apex and peak boundary indices are then converted to scan time.

4. **Assess peak shape and remove noise:** Fifteen peak characterization parameters (Baeza-Baeza et al., 2010) are calculated during feature extraction. These parameters can be optionally screened by a pre-trained random-forest model for peak shape quality evaluation. The pre-trained model provided by *MSS* was trained with 5000 examples of manually labeled peak data and is available for re-training as needed. Peak shape was scored with values of 1-3 based on good, uncertain, and bad peak shapes, respectively, following published criteria (Melnikov et al., 2020).

To complete feature extraction, steps 2 to 4 are repeated for all possible precursors in the chromatogram list generated from raw HRMS data.

Because all parameters in this process are user-adjustable, the package provides options to trade off computational speed and feature extraction accuracy. For batch data processing of

multiple .mzML files, the same workflow is performed on each datafile. After feature extraction, feature alignment across datafiles is performed based on Euclidean distance:

$$D_{ij} = \sqrt{(mz_i - mz_j)^2 + (RT_i - RT_j)^2}$$

where D_{ij} is the Euclidean distance between each feature observed in datafiles *i* and *j*, mz is the *m/*z ratio, and RT is retention time. Feature pairs with the lowest calculated distance across different datafiles are aligned. Aligned batch data can be exported as different user-defined formats (.csv, .tsv, .txt, .hdf, etc.) for subsequent analysis with other tools.

5.2.3 Initial data reduction

Feature extraction and alignment usually yields datasets containing hundreds to thousands of HRMS features per sample. However, in NTA, more features do not necessarily indicate greater sample complexity and improved resolving power across samples, as internal (e.g., instrument/software artifacts) or external (e.g., background noise, impurities from sample processing) interferences may bias comparisons. Therefore, careful data reduction, supported by proper study design (e.g., experimental blanks, controls, replicates) to identify and exclude such interferences is important to ensure data quality and accuracy. Various customizable data reduction filters are available in *MSS* for HRMS feature lists. A representative data reduction process (Helmus et al., 2021; Kutlucinar et al., 2022) might include: a) background feature subtraction based on a peak area fold-change criteria between experimental and blank samples; b) replicate evaluation to remove features based on the calculated average and coefficient of variation for data from experimental or analytical replicates; and c) data trimming based on

selected *m/z* or retention time ranges. These data reduction steps often effectively reduce feature numbers by up to 10-fold, simplifying subsequent data analysis (*MSS* function example shown in **Figure 5.3**). Although some "real" data is inevitably lost upon data reduction, stringent criteria for noise reduction and interfering detections typically improve the accuracy of downstream analyses, conserve calculation resources, and prioritize smaller data subsets for subsequent analysis (Hollender et al., 2017).

```
from mss import dm
d_ms = pd.read_csv('../example_data/example.csv')
# Data cleaning -- blank removal, RSD check, retains features that meaningful for subsequent analysis
# User select key strings in the data label from the blank samples
keys=['solvent','MeOH','ISTD']
d_sample = dm.data_prep(d_ms,keys,rt_range = [2,20], mz_range = [200,800], area_thres=100000)
d_sample.head()
```

```
100%|■■■■■■■■■■| 7/7 [00:00<00:00,  8.61it/s]
```

	Average m/z	Average rt	Average sn	Average score	20210512_oz_15min_1	20210512_oz_15min_2	20210512_oz_15min_3
0	200.07184	6.269	3.636500	1.0	54678.79	44141.285	189940.880
1	200.07265	5.725	4.134429	1.0	105930.68	140324.440	61354.580
2	200.07127	4.848	9.874761	1.0	82868.14	94710.190	55623.477
4	200.08714	6.929	1267.909200	1.0	160914.33	221140.950	161338.190
6	200.09070	15.155	3.137500	1.0	68103.08	14056.476	82632.920

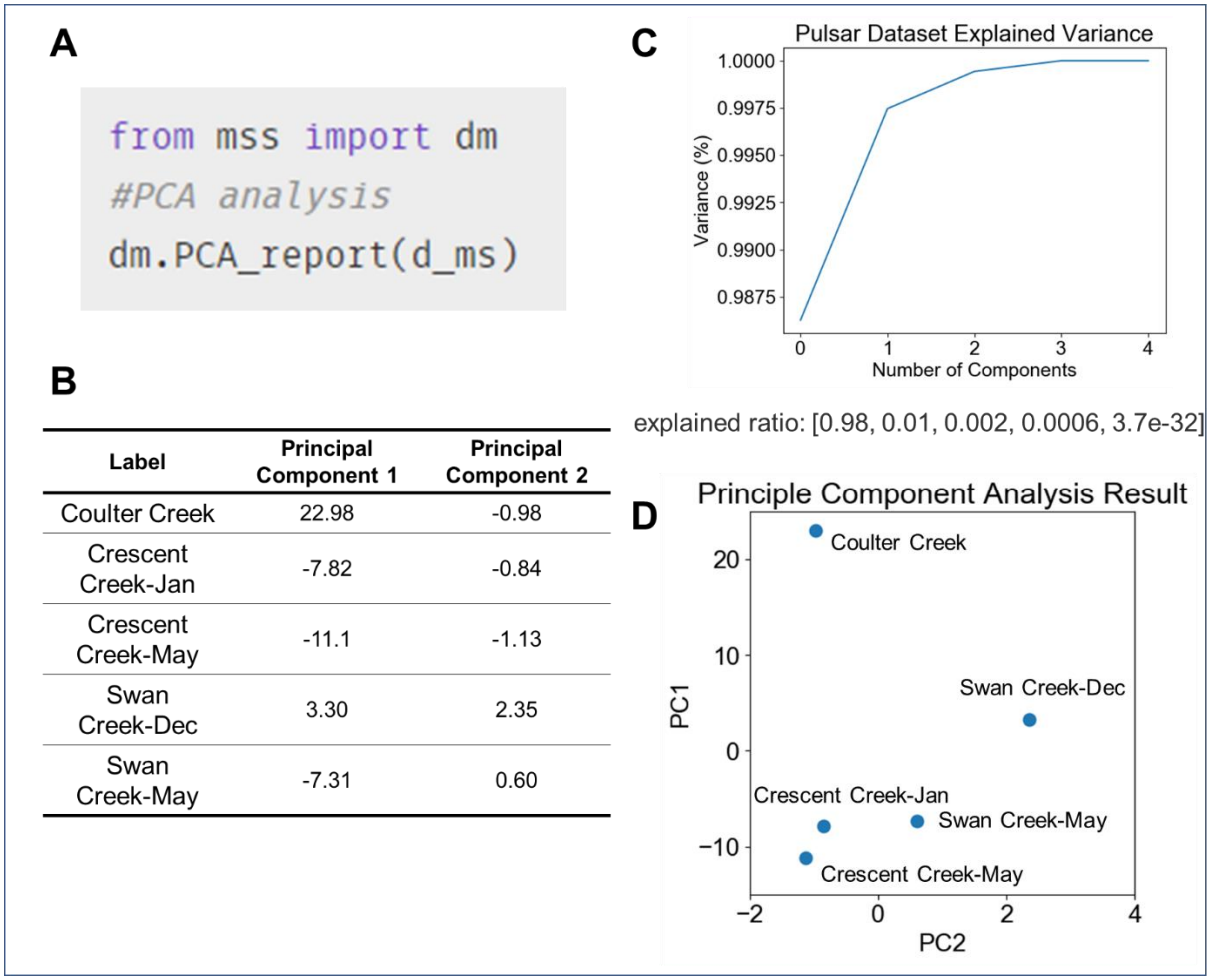
5 rows × 25 columns

Figure 5.3. Example code and output for data reduction functions. The example data were obtained from a previous study (Hu et al., 2022). For the data reduction, *MSS* relies on user-defined restrictions (*rt_range, mz_range* and *area_thres*) and performs blank subtracttion according to the *keys* statement. The *MSS* data reduction function also provides a consolidation option where replicates of the sample data will be automatically averaged and *MSS* will calculate the coefficient of variance for each averaged sample datapoint.

5.2.4 Advanced data mining

After initial workflow steps (e.g., feature extraction/alignment and data reduction), advanced analyses are often needed to extract meaningful information from NTA datasets (Helmus et al., 2021). When successful, these secondary data reduction processes also simplify the dataset and reduce the risk of incoherent classifications or predictions. Augmenting expected NTA workflow functionality, *MSS* provides basic statistical tools (e.g., hypothesis testing and trend comparison), as well as dimension reduction approaches (e.g., Principle Component Analysis (PCA) (Jolliffe and Cadima, 2016), t-distributed stochastic neighbor

embedding (T-SNE) (Maaten and Hinton, 2008)) to reduce the "curse of dimensionality" (Bellman and Lee, 1984) and provide simplified visualizations of complicated datasets. For example, PCA is easily performed using one-line commands in *MSS* (example provided in **Figure 5.4**).



Label	Principal Component 1	Principal Component 2
Coulter Creek	22.98	-0.98
Crescent Creek-Jan	-7.82	-0.84
Crescent Creek-May	-11.1	-1.13
Swan Creek-Dec	3.30	2.35
Swan Creek-May	-7.31	0.60

Figure 5.4. Example code and output for the PCA statistical tool, using archived experimental data from a previous study (Peter et al., 2019). Each datapoint was calculated as the mean of triplicate experimental data; labels for each datapoint represent the sampling site and time (if the sites were sampled more than once). (A) Example one-line code for executing PCA analysis. The d_ms is a dataframe pre-populated within previous codes (**Figure 5.3**). (B) Summary of principal components for the averaged datapoints from samples with different ozonation reaction time. (C) Cumulative explained variance vs. number of components. (D) PCA figure for the 1st and 2nd principal components for visualization. The PCA result evaluated chemical composition similarity between different creeks or the same creek at different sampling dates.

Beyond fundamental data mining tools, heuristic data exploration in *MSS* is supported by several machine learning-based approaches, including novel functionalities that are not offered by existing NTA workflows. These include: a) clustering tools to aggregate features with similar behavior patterns (i.e., similar trends of normalized abundances) across samples based on unsupervised machine learning algorithms (such as density-based spatial clustering of applications with noise [DBSCAN] (Ester et al., n.d.) and ordering points to identify the clustering structure [OPTICS] (Ankerst et al., 1999)); and b) a novel model-based source tracking tool. Detailed capabilities of these tools are described below.

5.2.5 Feature clustering

In *MSS*, feature prioritization is performed by the unsupervised clustering algorithm DBSCAN by default, with OPTICS as an alternative algorithm. DBSCAN finds core features that possess high density, then expands clusters from these cores with cluster boundaries delineated by user-defined tolerances ("DBSCAN," 2020). Compared to clustering algorithms commonly used in pattern recognition or temporal/spatial data grouping, such as KNN (Mucherino et al., 2009) and MeanShift (Comaniciu and Meer, 2002), DBSCAN can discover clusters with arbitrary shapes, is robust towards outlier detections, and has been successfully utilized in various areas including biochemical studies and text processing (Mustakim et al., 2019; Zhao et al., 2018). Using this approach, features with similar behavior patterns are automatically clustered with user-selected parameters, while outliers that diverge from recognized trends are excluded. Consequently, DBSCAN effectively prioritizes features that, for example, belong to a specific contamination source or are persistent or labile during a chemical reaction or treatment process. This can facilitate data processing and generate more

accurate results, while avoiding the need for laborious manual data processing in conjunction with user-defined or custom workflows.

Clustering analysis is performed with the *MSS* function *dm.ms_cluster*, which uses Z-score data normalization prior to clustering by default to eliminate data skewness and kurtosis:

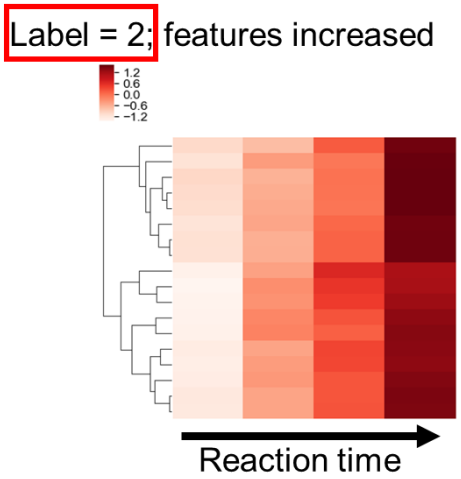
$$z = (x - \mu) / \sigma$$

where z is z-score, x is feature peak area in the sample, μ is the average peak area of the feature across all samples, and σ is the standard deviation of the peak areas. Other normalization algorithms are available from user settings (e.g., 0-1 scale normalization, log transformation). Normalized datasets are then processed with the DBSCAN (or OPTICS) algorithm for feature clustering. Optional dimension reduction methods (PCA or T-SNE) are available in the function according to user needs. Two tunable parameters for the DBSCAN algorithm, *min_samples* and *eps*, are determined via feature numbers (*min_samples*) and knee plot (*eps*; *MSS* provides a function *eps_assess* for this process). Clustered results can be optionally visualized for output evaluation, cluster selection for modeling analysis (**Figure S5.5**), and heuristic data exploration.

A

```
#Clustering analysis based primarily on DBSCAN algorithm
d_label = dm.ms_cluster(d_sample,
                        select_keyword=['exp'],
                        normalization='zscore',
                        visual=True,
                        d_reduce=False,
                        d_reduce_method='tsne',
                        perplexity=20,
                        cluster_method='dbscan',
                        eps=0.3,
                        min_samples=5)
```

B



C

m/z	RT	label	exp-0h	exp-4h	exp-10h	exp-29h
382.2164	7.2	2	0	18425.5	40000	104368
384.2324	7.6	2	0	87466	170762	312252
327.1849	5.7	2	23448	93887	168154	255889
309.1748	7.1	2	0	34545	65214	142342
…		…	…	…	…	…

Figure 5.5. Example code and output for the DBSCAN clustering tool applied to the synthetic progestin biotransformation dataset (Zhao et al., 2021). The data in .csv format was processed with following settings using DBSCAN algorithm (*eps=0.3, min_samples=5*). (A) Example code for deploying DBSCAN. (B) Example heatmaps for the DBSCAN grouped clusters for data evaluation and inspection during the NTA process: the labels represent the group number of the cluster and the color represents feature detected abundance (peak area). The feature clusters generated from the function were manually selected to prioritize the clusters that potentially consist of transformation products. For example, the label 2 cluster contains features that increased during the process, suggesting potential reaction products. (C) Example processed dataset with the cluster labels column (highlighted with red boxes), this data subset could be prioritized based on the clustering label for further analysis (annotation, source tracking, etc.).

5.2.6 Source tracking and apportionment

Leveraging the unsupervised clustering analysis in conjunction with predictive modeling approaches, *MSS* offers a novel source tracking functionality. In *MSS*, clustering functions described above are first used to isolate source fingerprints. Resulting fingerprint features are then aggregated to train and test a predictive source tracking model using user-selected algorithms (see **Section 5.3.3.3**). A complete workflow for source apportionment prediction from a pre-processed dataset using the *MSS* function (*dm.feature_model*) includes:

1. Prioritization of source fingerprint features by clustering analysis: Clustering analysis is performed on the dataset to designate features that cluster with source-associated patterns (e.g., decreasing abundance with source dilution) as source fingerprint candidates. Proper experimental design and sample preparation methods (e.g., a dilution series of a pollutant source sample, samples differentially impacted by the same pollutant source) are required to identify and prioritize source-representative features.
2. Data treatment for model training: A subset of the pre-processed original data is selected based on the prioritized fingerprint candidates, converted into a function-compatible format (e.g., renaming, data transposition, etc.), and split into training and test sets.
3. Model training: Using pairs of detected abundance and known source concentration for feature(s) or feature cluster(s) of interest (e.g., single feature, grouped features from

one or multiple clusters), the function trains the predictive model with user-selected algorithms. After training, the function optionally generates a performance report (e.g., coefficient of determination of the model (Pedregosa et al., 2011); visualized predicted vs. actual values) to support evaluation of the performance and importance of different feature clusters for accurate source apportionment.

4. Model validation and optimization: Trained models are validated using the testing data to assess model accuracy and avoid under- or overfitting. Based on the result of (3), users can tune model parameters or re-select feature cluster(s) to iteratively optimize results.
5. Source apportionment prediction: After model training and testing, users can deploy the model to evaluate source presence/concentration in unknown samples.

Currently, the *dm.feature_model* function incorporates several algorithms for multivariate regression, tree-based regression and support vector machine regression, providing flexibility for different datasets and user needs.

5.2.7 Feature annotation

Feature annotation (e.g., assigning a specific chemical identity to a detected feature) in *MSS* primarily exploits external databases with web Application Programming Interfaces (APIs). *MSS* functions interface to the ChemCalc online calculation tool (Patiny and Borel, 2013) for chemical formulae prediction and to the MassBank of North America database ("MassBank of North America," n.d.) for MS/MS fragment matching to facilitate identification (**Figure 5.6**).

A

```
# Formula annotation
msm.formula_prediction(scans, 299.1765, error=10, mfRange='C0-100H0-200N0-20O0-20P0-50')
```

Formula	Input Mass	Exact Mass	Unsat	Mass error (Da)	Mass error (ppm)	score
C18H23N2O2	299.175964	299.175953	8.5	-0.000011	-0.0380	71.729663
C8H27N7OP2	299.175964	299.175232	0.0	-0.000733	-2.4486	61.629385
CH17N17O2	299.175964	299.175113	2.0	-0.000851	-2.8461	56.447396
C12H24N6OP	299.175964	299.174921	4.5	-0.001043	-3.4873	51.891714
C16H21N5O	299.175964	299.174610	9.0	-0.001354	-4.5259	44.201249

B

```
# MS2 search
frag.frag_massbank(scan_frag,precursor = 301.2122,error=50)
```

```
Now showing index 0 of 30 total found scans
```

	name	formula	inchiKey	instrument	CE	ionization mode	precursor type	score	url
0	Eicosapentaenoic acid	C20H30O2	JAZBEHYOTPTENJ-JLNKQSITSA-N	LC-ESI-QTOF	20	negative	[M-H]-	0.980410	https://mona.fiehnlab.ucdavis.edu/spectra/display/RP030512
1	Phosphatidylethanolamine lyso 20:5	C25H42NO7P	UNILCTDPRMGSDM-JEBPEJKESA-N	LC-ESI-ITFT	30%	negative	[M-H]-	0.978217	https://mona.fiehnlab.ucdavis.edu/spectra/display/UT001941
2	Phosphatidylcholine lyso 20:5	C28H48NO7P	OYUDYTIZCLNYHF-NEUKSRIFSA-N	LC-ESI-ITFT	30%	negative	[M-CH3]-	0.976437	https://mona.fiehnlab.ucdavis.edu/spectra/display/UT002353

Figure 5.6. Example code and output for the feature annotation function. (A) Example code and output for the formula calculation function. The formula calculation is based on ChemCalc API (Luc, n.d.). (B) Example code and output for the MS2 online search annotation tool are based on the MassBank North America database ("MassBank of North America," n.d.).

For formula prediction in *MSS*, after candidate formulas are calculated from monoisotopic precursor mass by ChemCalc, prediction accuracy is evaluated with a dot-product based score via isotopic comparison between theoretical and observed spectra (Tsugawa et al., 2015). For compound identification, *MSS* supports individual or averaged spectra upload options and results retrieval, following MassBank database searching criteria and protocols. Processed data from *MSS* can be exported for further annotation using other platforms and databases, such as MetFrag (Ruttkies et al., 2016), SIRIUS (Dührkop et al., 2019), GNPS (Wang et al., 2016), and NIST databases (Daniel, 2017).

5.2.8 Visualization, reporting, and user interface

Several visualization functionalities are available in *MSS* for HRMS data inspection (within the *visreader* module), including an overview *m/z* & RT scatter plot, total ion chromatogram (TIC), extracted ion chromatogram (EIC), and selected MS or MS/MS spectra. Raw HRMS data is inspected as the parsed list object (**Figure 5.7**) or visualized using functions from the *visreader* module (**Figure 5.8**). Output figures are available in static or interactive formats. Some visualization functions (EICs, MS and MS/MS spectra) provide optional online database search options and comparison with theoretical results (e.g., isotopologue pattern, MS/MS fragmentation) to help users understand and communicate HRMS data. Beyond designated visualization functions within the *visreader* module, data output visualization options are also integrated into most *MSS* functions, including those for advanced data mining (e.g., PCA, feature clustering analysis), for users to immediately evaluate package results.

```
path = '../example_data/example.mzML'
scans = mssmain.get_scans(path, ms_all=True)
scan_index = 0
print(scans[scan_index].scan_time)
#Print the top 10 mz/rt pairs
pprint(list(zip(scans[scan_index].mz,scans[scan_index].i))[:10])
```

```
 (0.0654, 'minute')
 [(100.03542, 821.04535),
  (100.076225, 4184.5024),
  (100.387634, 242.57353),
  (100.93503, 4630.397),
  (101.0046, 864.1344),
  (101.06307, 559.711),
  (101.96105, 2411.1638),
  (102.01361, 106558.734),
  (103.009766, 4003.1045),
  (103.06762, 730.78284)]
```

Figure 5.7. Example code and output for raw data inspection. After data was acquired and converted by *MSS*, the .mzML data could be interfaced directly as a python-list object.

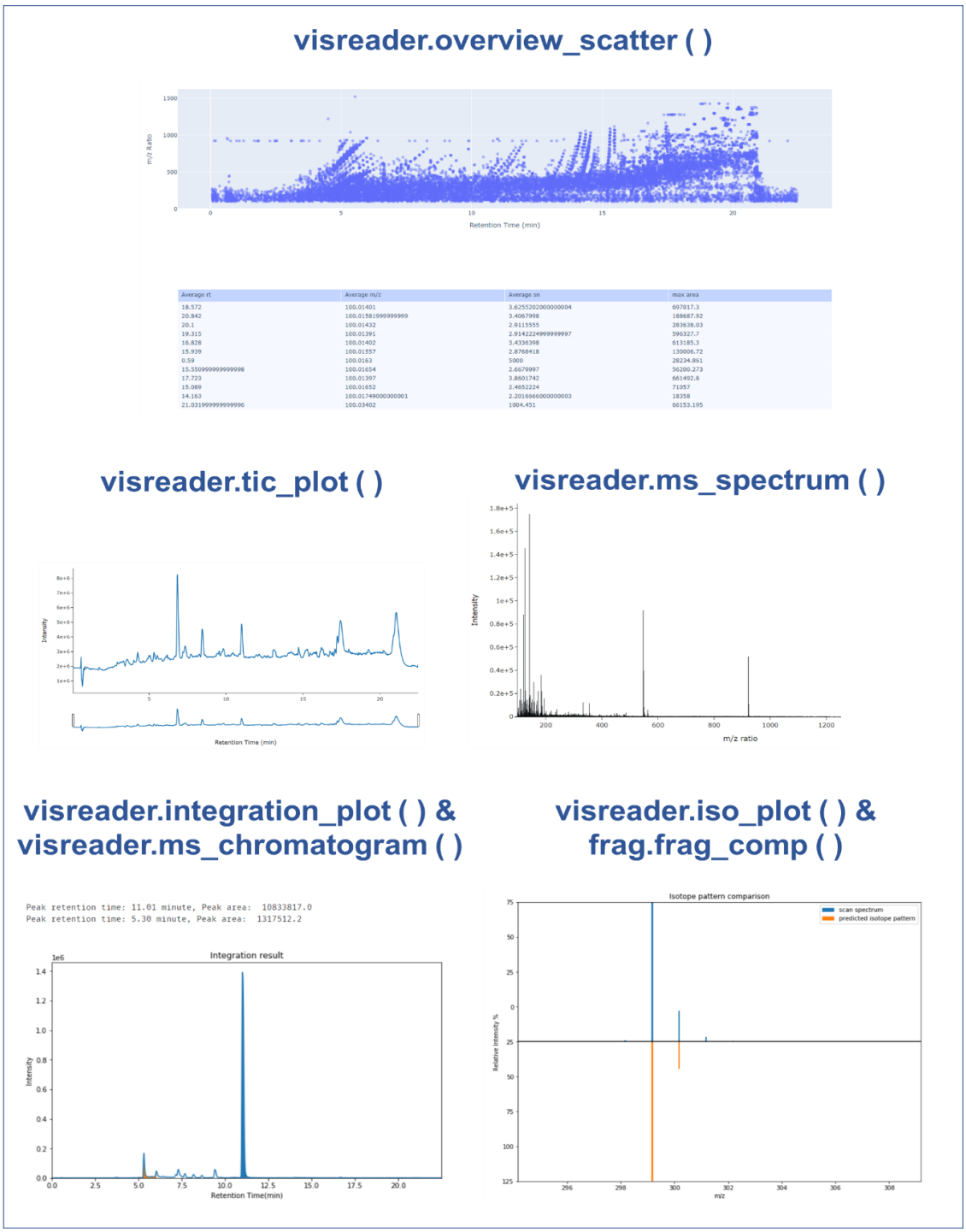


Figure 5.8. Example output for HRMS data visualization. The *overview_scatter* function is plotting all the extracted features from the *MSS* generated peak list and assists users in evaluating the feature extraction and sample data quality. The *tic_plot*, *ms_spectrum*, *integration_plot* and *ms_chromatogram* functions are used for raw data inspection, as assistance to existing software or workflow. The *iso_plot* and *frag_comp* functions are accompanied with feature annotation functions, users could visualize the MS spectrums for isotopic pattern comparison or MS/MS spectrums for chemical identification, using online or in-house databases.

MSS is designed to ensure easy interpretation and export of processed data. All processed data (as spreadsheets) can be saved with the *Pandas* function (McKinney, 2010). Visualization plots can be saved directly from the output window in user-defined formats (e.g., .png, .jpg). Trained models for feature extraction and quantitative source apportionment can be serialized using the *pickle* package (Van, 2020). Recommended interfaces for *MSS* are through notebook-style integrated development environments either locally (e.g., jupyter notebook) or remotely (e.g., Google Colab), while feature extraction and data alignment functions can be executed as a command line script to allow running the software on a high performance computing cluster or the cloud.

5.2.9 Software distribution and availability

MSS is distributed as a Python package with some external supporting packages developed with C++. The package currently supports *Microsoft Windows*, *Linux*, and *macOS* platforms. Documentation (https://github.com/XiminHu/mass-suite#readme) includes the latest patch notes, dependencies, tutorial examples, and example data for package testing. *MSS* was automatically tested during development with a continuous integration pipeline (GitHub Action). *MSS* distribution is generated with *dist* package and uploaded to PyPI server with *twine* package. Users can install the package via *pip install* command (https://pypi.org/project/mass-suite/), within Anaconda, or through the external command-line.

5.3 Results and Discussion

5.3.1 Feature extraction reliability

In HRMS data analysis, manual inspection of all extracted chromatographic peaks is typically impractical, so feature extraction accuracy impacts the quality of subsequent analyses. To assess reliability of *MSS* feature extraction, archived samples (mixed chemical standards) from the EPA ENTACT study (Ulrich et al., 2019a) (sample numbers #505, #506 and #508; 398 MS-amenable chemicals in total) were analyzed and processed through the *MSS* feature

extraction workflow. The instrument method was adapted from a previous study (Peter et al., 2019). The methanolic samples (N=3; batch code assigned with the samples: #505, #506 and #508; 413 chemicals spiked in total) from ENTACT study (Ulrich et al., 2019b) were provided by EPA (stored at -20 ℃). Details of the 398 MS-amenable chemicals are listed in **Table S5.1**. Constituents of all three samples were selected from ToxCast chemicals and verified to have >90% purity. #505 and #506 mixtures consist of chemicals that have no overlap in the monoisotopic mass to avoid erroneous identification. On the other hand, #508 mixture was developed with some isomeric and isobaric chemicals. All samples were analyzed using an Agilent 1290 Infinity ultrahigh performance liquid chromatograph (UHPLC) coupled to an Agilent 6530 quadrupole time-of-flight high-resolution mass spectrometer (QTOF-HRMS; Santa Clara, CA, USA). A reverse-phase C18 column (Agilent ZORBAX Eclipse Plus 2.1×100 mm, 1.8 µm) with a C18 guard column (2.1×5 mm, 1.8 µm) was used for the UHPLC separation at 45 ºC with 5 µL injection volume. LC separation used a gradient elution with mobile phases of 0.1% formic acid in each of deionized water (A) and methanol (B) as follows: 5% B at 0-1 min, 50% B at 4 min, 100% B at 17-20 min, 5% B at 20.1 min; stop time 22.5 min; post-time 2 min. The flow rate was 0.4 mL/min. Full scan data were acquired under 2 GHz Extended Dynamic Range mode at a range of 100-1700 m/z. Mass calibration was performed before the analysis.

After data collection, raw datasets were parsed and inspected for *m/z* and RT identifiers to confirm chemical presence. The datasets were also processed with *MSS* workflow with default settings (*noise_removal=2000, mz_error=10, peak_thres=0.001, peakutils_thres=0.1, min_d=1, peak_area_thres=1e5, min_scan=5, max_peak=5*; details in the function documentation). The retrieved peak list was compared with the manual confirmed *m/z* & RT list for the chemicals from the raw datasets. Among the 413 spiked chemicals, 398 were manually confirmed detection, while 25 chemicals excluded from analysis because of no

detection. The alignment was performed between two lists for comparison, the closest neighbors from the two lists were screened with filters of *m/z* difference <0.015 and RT difference <0.5 minutes. *MSS* extracted 99.5% of peaks known to be present in all three mixtures (Ulrich et al., 2019a). Among the 398 detected chemicals, only 2 peaks were failed during the peak retrieval process. One chemical is m-Cumenyl methylcarbamate, due to multiple chromatogram peaks that potentially surpassed the maximum peak filter within the *peak_pick* function. Another chemical is Trinexapac-ethyl, which has a bad peak shape that likely contributed to the detection failure.

The extracted feature list from *MSS* for all archived ENTACT mixture samples (#505, #506 and #508) was then compared with two other open-source platforms (*MSDIAL* (Tsugawa et al., 2015) and *XCMS* (Smith et al., 2006)) to evaluate feature extraction performance for total reported features. The data processing parameters were set as similar to those in *MSS* to the extent possible. For MS-DIAL: MS1 tolerance = 0.003 Da; Min Peak Height = 2000; Mass Slice Width = 0.01; Min Peak Width = 7 scan. For XCMS: MS1 tolerance = 10 ppm; Min Peak Width = 10 sec; Max Peak Width = 90 sec; S/N threshold = 10; MZdiff = 0.01; Prefilter Peaks = 1; Prefilter Intensity = 2000 (**Table 5.2**). All other default parameter values were retained. The feature peak lists generated with *MSS* were then compared to the peak lists from these two software platforms, using the same criteria for peak retrieval validation (*m/z* difference <0.015 and RT difference <0.5 minutes).

Table 5.2. Key parameters for peak extraction for *MSS, MSDIAL and XCMS*

Tools / Features	MSS	XCMS	MSDIAL
Noise removal	2000	2000	2000
Signal-to-noise threshold	10	10	NA
m/z error tolerance	10 ppm	10 ppm	0.003 Da

Peak area threshold	1e5	1e5	1e5
Minimum peak width	5 scans	10 sec	7 scans

From comparison, most features extracted by *MSS* overlapped with those reported by other software (**Figure 5.9**; on average 52 ± 5% and 52 ± 6%, for *MSDIAL* and *XCMS* respectively), validating *MSS* performance in comparison to other well-accepted feature extraction tools. RT & *m/z* differences between the overlapped features also suggested similar data processing outcomes across these three packages (**Figure S5.1**).

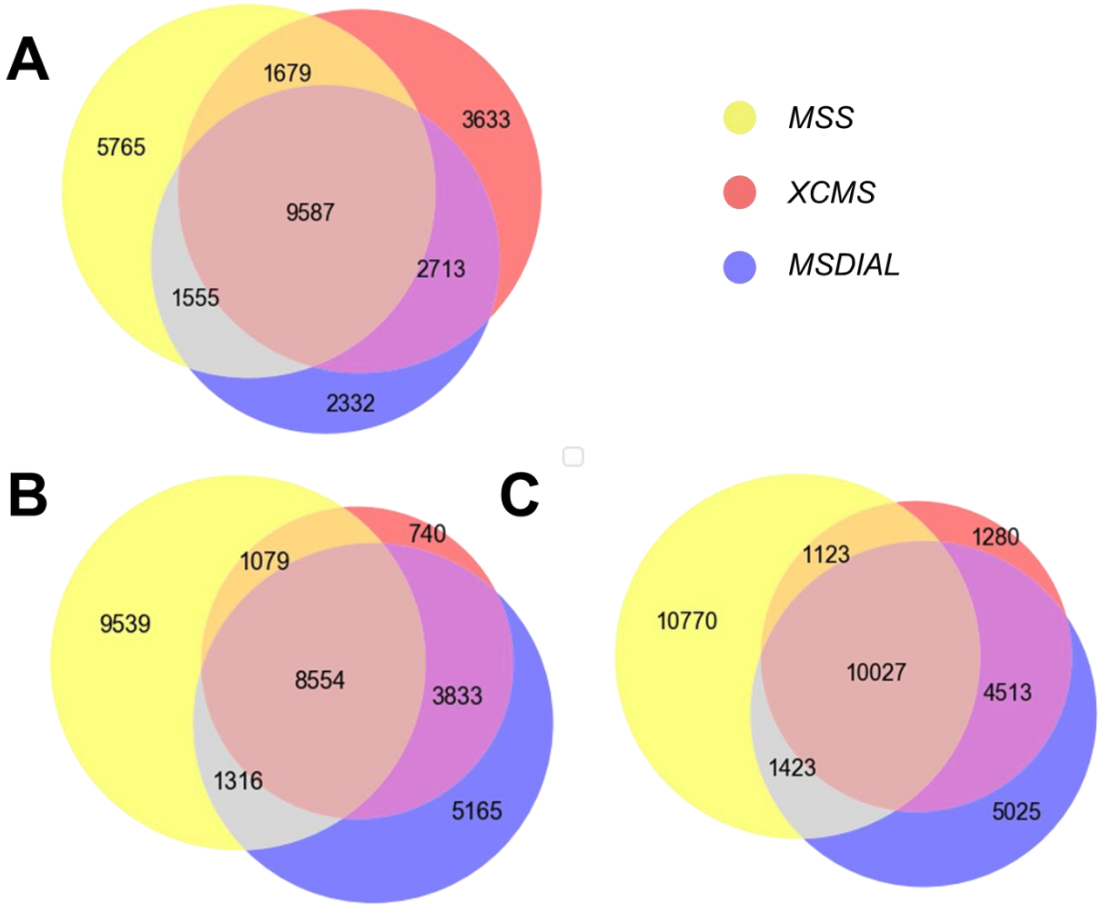


Figure 5.9. Comparisons of feature extraction outcomes for identical input samples. Samples numbered A) #505, B) #506 and C) #508 from the ENTACT study (Ulrich et al., 2019a) with *MSS*, *XCMS* and *MSDIAL* software processing. Venn diagrams report extracted features overlap between different platforms. The feature extractions were performed with parameters matched as closely as possible across the different platforms. Key parameters for peak extraction for different platforms are reported in **Table 5.2**.

5.3.2 Multiprocessing benchmarks

To minimize computational runtime, multiprocessing is optionally available for the most calculation-intensive functions (*peak_pick* and *peak_list*) that handle single or batch-file peak extraction. Multiprocessing occupies all available cores for calculation by default but is user-customizable. The data files used for benchmarking were from the same samples (ENTACT #505, #506 and #508) as the feature extraction validation. The multiprocessing benchmarks were performed on a personal laptop with pre-complied scrtips, which equipped with an 11th Gen Intel(R) Core (TM) i7-1165G7 @ 2.80GHz, 1.69 GHz processor (4 cores, 8 threads), 12 gigabyte RAM, SATA SSD storage, and Windows 10 Home operating system. The scripts were executed within PyCharm (version 2020.2.1) and the runtimes for peak extraction, with multiprocessing enabled and disabled using default parameters, were recorded in triplicate. Compared to single core processing, with all cores working, the processing time decreased from 201 ± 1.1 s to 58 ± 1.4 s (87 ± 3% of the theoretical maximum for 4 cores of computational power) for single file feature extraction and 897 ± 9.1 s to 350 ± 37 s (65 ± 7% of theoretical maximum) for multiple file (batch) feature extraction. Thus, parallel processing scripts did provide optional high-efficiency processing allowing for some optimization of computational resources.

5.3.3 Demonstration of MSS applications

The sections above introduced *MSS* functionalities and described validation of the software package reliability. Here, three applications of *MSS* to analyze lab-generated datasets are provided, focusing on: I-II) automated feature prioritization and III) source tracking analysis. We note that *MSS* was not solely used for all HRMS data processing in examples I (feature extraction/alignment) and II (feature extraction/alignment, blank subtraction) to maintain consistency with other analyses and studies.

Example I: Clustering analysis to prioritize 6PPD transformation products

This example demonstrated use of the *MSS* feature prioritization workflow to facilitate non-target screening of transformation products from a reaction process. An early (pre-release) version of *MSS* was used to aid prioritization (by clustering analysis) and identification (by formula annotation) of potential transformation products of 6PPD (a tire rubber antioxidant) during laboratory ozonation studies, fully described in Hu et al. (Hu et al., 2022). Initial feature extraction and data alignment in Hu et al. (Hu et al., 2022) was accomplished by *MSDIAL* (version 3.46) (Tsugawa et al., 2015), with all subsequent data cleaning, formula annotation, and statistical analysis performed in *MSS*. Key outcomes were that the *MSS* data pre-processing workflow effectively reduced the total feature count across 61 unique samples (excluding blanks) from 41,808 to 936 by blank subtraction, replicate filtering, and intensity filters within desired *m/z* and RT ranges (*m/z* 100-900; RT 3-18 min). Clustering analysis in *MSS* with the DBSCAN algorithm (DBSCAN parameters: *eps=0.4*, *min_samples=3*) prioritized 297 features with trends of increasing peak area abundance during ozone exposure based on chemical clusters (processing time < 30min). Ninety-eight features were retained after filtering based on detected abundance and predicted chemical formula. Critically, the unique workflow provided by *MSS* allows users to discover clustered behavior patterns of HRMS features, select features with relevant patterns (e.g., increasing over time, as expected for stable transformation products), and reduce analysis time (compared to manual operation, typically ~15-20 h for a dataset of this size), thus facilitating feature prioritization.

5.3.3.2 Example II: Clustering analysis for biotransformation product discovery

To further demonstrate and validate *MSS* workflow capabilities for accurately prioritizing features of interest by clustering analysis, archived HRMS data obtained from a previous biotransformation study (Zhao et al., 2021) was re-analyzed with *MSS* (version 1.1.2). Briefly, the synthetic progestins dienogest and drospirenone were incubated in batch reactors, with samples collected over time (0, 4, 10, and 29 h) to measure biotransformation kinetics and

identify transformation products (Zhao et al., 2021). Initial feature extraction, data alignment, and blank subtraction used Agilent software (MassHunter Profinder (B.08.00) and Mass Profiler Professional (B.13.00). Originally, features were manually prioritized as potential transformation products based on molecular formula and diagnostic MS/MS fragments (~30 h manual time). Here, as an illustrative case, the data exported from the Agilent software (.csv format) was processed in *MSS* using clustering analysis to prioritize potential transformation products. *MSS* efficiently clustered features with similar trends (**Figure 5.5**; DBSCAN parameters: *eps=0.3*, *min_samples=5*; total processing time < 30 min), with 18 features identified as potential transformation products from the input list (after pre-processing for blank and control subtraction) of 136 features. Among those, nine *MSS*-prioritized candidates matched products reported originally (dienogest: TP311, TP 309, TP327b; drospirenone: TP 384, TP 380, TP 370a, TP 370b, TP382c and TP 368), representing 82% of the 11 "major biotransformation products" reported in Zhao et al. (Zhao et al., 2021). The function was primarily tuned to prioritize potential TPs that were resistant to further reactions (i.e., monotonically increasing abundance). Thus, the manually-identified intermediate TPs (2 TPs, dienogest: TP 313; drospirenone: TP 364), which degraded after initial formation, were not reported in the *MSS* prioritization results. Overall, the *MSS* data reduction and clustering analysis workflow yielded accurate results and significantly reduced data processing time, with improved performance anticipated with further parameter optimization, additional feature information (e.g., MS/MS spectra), or additional data processing to reduce false positive and false negative results.

5.3.3.3 Example III: Source apportionment modeling

The source tracking approach within *MSS* builds on our previous laboratory study on this topic (Peter et al., 2019); preliminary testing of *MSS* was conducted by re-analyzing archived sample data from that same study. Detailed sample composition and data acquisition methods

are provided elsewhere (Peter et al., 2019). Briefly, two complex roadway runoff samples were diluted and mixed with other water samples to mimic downstream mixing behaviors of multiple potential contaminant sources. In the original work, after HRMS analysis and data extraction using Agilent software (MassHunter Profinder (B.08.00) and Mass Profiler Professional (B.13.00)), fingerprint features were manually isolated and used to quantitatively apportion the amount of contaminant source in the mixed samples (Peter et al., 2019). Using *MSS* (version 1.1.2), we replicated this conceptual approach while incorporating machine learning approaches. The archived processed MS peak list in .csv format from a previous study (Peter et al., 2019) was used for the example application case. Initial data reduction of the imported data was done via *MSS*, to isolate features with maximum peak area > 100,000, *m/z* between 200-800 Da, retention time between 2-20 min, and present at peak area 5-fold greater than peak area in any of the solvent blanks or the ISTD controls. Features satisfying all these criteria were retained. After data reduction, DBSCAN clustering analysis was conducted to prioritize features as the fingerprint of the source (denoted "SR520"). All the peak areas were normalized as follows to eliminate data skewness and kurtosis:

$$z = (x - \mu) / \sigma$$

where z is z-score, x is the peak area of the feature, μ is the average peak area of the feature across all the samples, and σ is the standard deviation of the peak areas. After normalization, the DBSCAN clustering was conducted on the SR520 dilution curve samples (*eps*=0.6, *min_samples*=10) to prioritize the features representative of the SR520 source. Three clusters were automatically populated (Cluster 0, 1 and 2; N=630 in total) and used for model training. The data were randomly split as training and testing datasets and an ensemble random forest model was trained and validated using the 630 features prioritized. After modeling, the model performance report was generated based on the coefficient of determination. Based on the model score, cluster 0 (score=0.993; N=587) was used to

estimate the source concentration of the mixture samples. Subsequent model training, output summary, and source apportionment predictions are shown in **Figure 5.10**.

A

```
# Data import
d_ms = pd.read_csv('../example_data/clustering/sample1114.csv')
d_test = pd.read_csv('../example_data/clustering/sample0815.csv')

# Data cleaning
keys=['solvent','MeOH','ISTD']
d_sample = dm.data_prep(d_ms,keys,rt_range = [2,20], mz_range = [200,800], area_thres=100000)

# Clustering analysis
d_label = dm.ms_cluster(d_sample, ['SR520-Cal'], 'zscore', d_reduce=False, visual=False, cluster_method='dbscan', eps=0.6, min_samples=10)

# Model training
d_model = d_label[d_label['label']!=-1].copy()
d_merge = dm.batch_alignment(d_model,d_test)
dilu_col=['SR520-Cal','SR520_Cal']
d_transpose = dm.transpose(d_merge, dilu_col)
reg = dm.feature_model(d_transpose, cluster_algorithm=True, model_method='ensemble_rf', report=True)

# Prediction on selected new samples based on the model generated above
dm.cluster_pred(reg, d_merge, '20181114_SR520-Creek_Mix1_1', selected_cluster=[0], model_merge='weighted_size')
```

B

```
100%|■■■■■■■■■■| 37/37 [00:35<00:00,  1.05it/s]
data split finished!
modeling finished!
Report starting...
RandomForestRegressor(max_features=5, n_estimators=1000, random_state=10)
The model score for cluster 0 is: 0.993
The model score for cluster 1 is: 0.734
The model score for cluster 2 is: 0.767
The model score is the coefficient of determination.
Report finished!
```

C

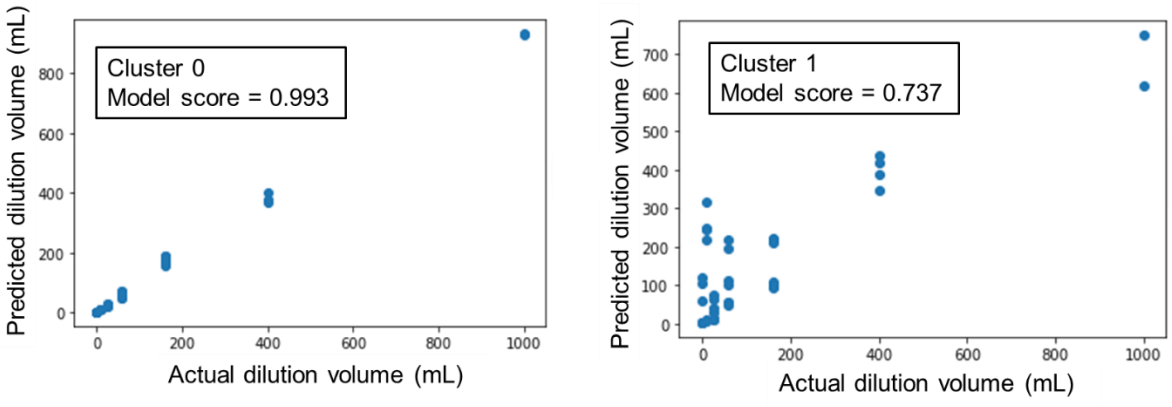


Figure 5.10. Illustrative example code and output for the modeling function. Archived sample data from a previous study were re-analyzed for method development and testing (Peter et al., 2019). A) example code for the data reduction, feature clustering, model training, model evaluation and source concentration prediction. Data reduction was performed with following settings: *RT*=2-20 min, *m/z*=200-800, *Peak Area*>100000. DBSCAN algorithm (*eps*=0.6, *min_samples*=10) was used to cluster features for subsequent modeling functions. B) Example output for model performance evaluation and dilution volume prediction. The ensemble random forest model was trained with data subsets of selected features (#0). The model performance report is available for users upon request to assist modeling algorithm and feature cluster selection. The model score is the coefficient of determination (i.e., higher score represents better model fitting) and C) example model performance inspection for selected clusters. Each datapoint in panel C represents a single sample data.

MSS predictions, using an ensemble random forest regression model, were compared with original prediction results (**Figure 5.11**). Note that the *MSS* estimates were derived from an initial clustering analysis and model without further optimization, so accuracy could presumably be improved with iterative optimization. Challenges remain for improving predictions when a) limited chemical features are available at lower pollutant source concentrations (e.g., *MSS* prediction error ranged from 40% to 400% for Mixes 4A, 5A and 6A, which contained 4%, 1% and 0.6% pollutant source by volume, respectively, compared to Mixes 1-3 containing >10% source); and b) co-occurring sources and/or the background matrix introduce features that overlap with source fingerprint features and bias predictions (e.g., *MSS* prediction errors were higher in mixes 4B, 5B, and 6B, which contain 10%, 2.5% and 0.4% by volume, respectively, of a second roadway runoff source). Nevertheless, prediction accuracy for mixtures with higher source concentrations (Mixes 1, 2 and 3; 30%, 18% and 10% pollutant source by volume, respectively) were similarly accurate (~ 5% differences in predicted source concentrations) as the original results, validating the utility of the source apportionment modeling function in *MSS*.

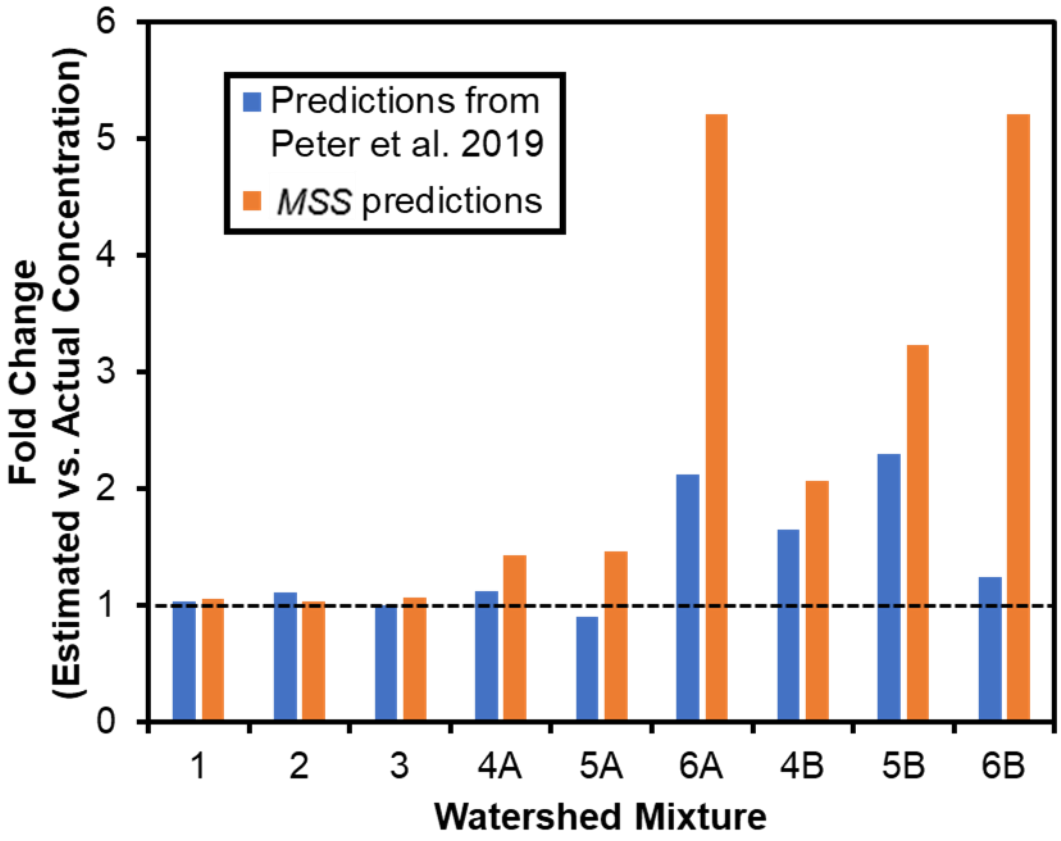


Figure 5.11. Estimates of fold change (estimated vs. actual concentration) of source (roadway runoff) concentration from a previous study (Peter et al., 2019) and *MSS* model predictions. *MSS* predictions were built from an ensemble random forest model that was trained with roadway runoff source sample dilution. One cluster of compounds (Cluster label=0, N=587) was prioritized from DBSCAN clustering analysis and used to derive estimates. The dashed line (fold change=1) indicates predicted concentration equal to actual concentration.

5.4 Conclusions

We here communicated the structure and research capabilities of *MSS* as an open source and customizable HRMS data analysis software package developed with Python. *MSS* provides numerous default and user-defined modules (data import, feature extraction, data reduction, data visualization, feature annotation, and advanced data mining), that are accessible, flexible, and optimizable for custom study designs and data analysis scenarios, ensuring reproducible and accurate HRMS data analysis. Complementing traditional NTA data analysis approaches that focus on prioritization and identification of a small group of chemicals, core *MSS* functions provide a workflow for feature extraction, clustering analyses, and source tracking approaches that are supported by machine learning algorithms, allowing users to better leverage all relevant

HRMS features for prioritization and modeling. These novel functions replace manual data reduction efforts and facilitate exploratory studies intended to utilize HRMS data as "big data". Functional documentation and examples provide a quick and easy training guide for users with basic computational expertise. With respect to integrated software performance, the reliability tests and benchmarks also demonstrate the accuracy, efficiency, and power of *MSS* data analysis for various NTA and HRMS studies.

Because the *MSS* package is actively maintained and updated, several innovative functions and tools are in development for further NTA applications, including optimization of the chemical fingerprint-based source apportionment tool and a tool for matrix effect assessment and correction that leverages feature network analysis approaches. *MSS* is published on pypi.org, is fully open-source, and is available to anyone interested in using the default settings, adapting the code to their specific needs, or making contributions. Feedback and real-world case studies from interested users within the NTA community are especially welcome. We anticipate that the comprehensive, integrated functionalities provided by the *MSS* software package, together with its strengths of open availability, easy use, and external calculation resource compatibility will be especially useful to the HRMS and data science communities to assist with fully exploiting the rich datasets generated with HRMS instruments.

5.5 Availability and Requirements

Project name: Mass-Suite (*MSS*).
Project home page: https://github.com/XiminHu/mass-suite
Operating system(s): Platform independent (tested on Microsoft Windows and Linux).
Programming language(s): Python.
Other requirements: none.
License: MIT License.
Any restrictions to use by non-academics: none

References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G.S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mane, D., Monga, R., Moore, S., Murray, D., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viegas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y., Zheng, X., 2016. TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems. arXiv:1603.04467 [cs].

Ankerst, M., Breunig, M.M., Kriegel, H.-P., Sander, J., 1999. OPTICS: ordering points to identify the clustering structure, in: Proceedings of the 1999 ACM SIGMOD International Conference on Management of Data, SIGMOD '99. Association for Computing Machinery, New York, NY, USA, pp. 49–60. https://doi.org/10.1145/304182.304187

Baeza-Baeza, J.J., Pous-Torres, S., Torres-Lapasió, J.R., García-Álvarez-Coque, M.C., 2010. Approaches to characterise chromatographic column performance based on global parameters accounting for peak broadening and skewness. J. Chromatogr. A 1217, 2147–2157. https://doi.org/10.1016/j.chroma.2010.02.010

Bellman, R., Lee, E., 1984. History and development of dynamic programming. IEEE Control Syst. 4, 24–28. https://doi.org/10.1109/MCS.1984.1104824

Bittremieux, W., 2020. spectrum_utils: A Python Package for Mass Spectrometry Data Processing and Visualization. Anal. Chem. 92, 659–661. https://doi.org/10.1021/acs.analchem.9b04884

Blum, K.M., Andersson, P.L., Renman, G., Ahrens, L., Gros, M., Wiberg, K., Haglund, P., 2017. Non-target screening and prioritization of potentially persistent, bioaccumulating and toxic domestic wastewater contaminants and their removal in on-site and large-scale sewage treatment plants. Sci. Total Environ. 575, 265–275. https://doi.org/10.1016/j.scitotenv.2016.09.135

Chen, C.-J., Lee, D.-Y., Yu, J., Lin, Y.-N., Lin, T.-M., 2022. Recent advances in LC-MS-based metabolomics for clinical biomarker discovery. Mass Spectrom Rev e21785. https://doi.org/10.1002/mas.21785

Comaniciu, D., Meer, P., 2002. Mean shift: a robust approach toward feature space analysis. IEEE Trans. Pattern Anal. Mach. Intell. 24, 603–619. https://doi.org/10.1109/34.1000236

Daniel, S., 2017. NIST Standard Reference Simulation Website [WWW Document]. URL https://chemdata.nist.gov/ (accessed 9.29.21).

Dávila-Santiago, E., Shi, C., Mahadwar, G., Medeghini, B., Insinga, L., Hutchinson, R., Good, S., Jones, G.D., 2022. Machine Learning Applications for Chemical Fingerprinting and Environmental Source Tracking Using Non-target Chemical Data. Environ. Sci. Technol. 56, 4080–4090. https://doi.org/10.1021/acs.est.1c06655

DBSCAN, 2020. . Wikipedia.

Dom, I., Biré, R., Hort, V., Lavison-Bompard, G., Nicolas, M., Guérin, T., 2018. Extended Targeted and Non-Targeted Strategies for the Analysis of Marine Toxins in Mussels and Oysters by (LC-HRMS). Toxins 10, 375. https://doi.org/10.3390/toxins10090375

Du, B., Lofton, J.M., Peter, K.T., Gipe, A.D., James, C.A., McIntyre, J.K., Scholz, N.L., Baker, J.E., Kolodziej, E.P., 2017. Development of suspect and non-target screening methods for detection of organic contaminants in highway runoff and fish tissue with high-resolution time-of-flight mass spectrometry. Environ. Sci.: Processes Impacts 19, 1185–1196. https://doi.org/10.1039/C7EM00243B

Dührkop, K., Fleischauer, M., Ludwig, M., Aksenov, A.A., Melnik, A.V., Meusel, M., Dorrestein, P.C., Rousu, J., Böcker, S., 2019. SIRIUS 4: a rapid tool for turning tandem mass spectra into metabolite structure information. Nat. Methods 16, 299–302. https://doi.org/10.1038/s41592-019-0344-8

Ester, M., Kriegel, H.-P., Xu, X., n.d. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise 6.

Fauser, P., Tjell, J.C., Mosbaek, H., Pilegaard, K., 1999. Quantification of Tire-Tread Particles Using Extractable Organic Zinc as Tracer. Rubber Chemistry and Technology 72, 969–977. https://doi.org/10.5254/1.3538846

Gonzalez de Vega, R., Cameron, A., Clases, D., Dodgen, T.M., Doble, P.A., Bishop, D.P., 2021. "Simultaneous targeted and non-targeted analysis of per- and polyfluoroalkyl substances in environmental samples by liquid chromatography-ion mobility-quadrupole time of flight-mass spectrometry and mass defect analysis." J. Chromatogr. A 1653, 462423. https://doi.org/10.1016/j.chroma.2021.462423

Helmus, R., ter Laak, T.L., van Wezel, A.P., de Voogt, P., Schymanski, E.L., 2021. patRoon: open source software platform for environmental mass spectrometry based non-target screening. J. Cheminf. 13, 1. https://doi.org/10.1186/s13321-020-00477-w

Hollender, J., Bourgin, M., Fenner, K.B., Longrée, P., Mcardell, C.S., Moschet, C., Ruff, M., Schymanski, E.L., Singer, H.P., 2014. Exploring the Behaviour of Emerging Contaminants in the Water Cycle using the Capabilities of High Resolution Mass Spectrometry. CHIMIA International Journal for Chemistry 68, 793–798. https://doi.org/10.2533/chimia.2014.793

Hollender, J., Schymanski, E.L., Singer, H.P., Ferguson, P.L., 2017. Nontarget Screening with High Resolution Mass Spectrometry in the Environment: Ready to Go? Environ. Sci. Technol. 51, 11505–11512. https://doi.org/10.1021/acs.est.7b02184

Hu, X., Walker, D.I., Liang, Y., Smith, M.R., Orr, M.L., Juran, B.D., Ma, C., Uppal, K., Koval, M., Martin, G.S., Neujahr, D.C., Marsit, C.J., Go, Y.-M., Pennell, K.D., Miller, G.W., Lazaridis, K.N., Jones, D.P., 2021. A scalable workflow to characterize the human exposome. Nat Commun 12, 5575. https://doi.org/10.1038/s41467-021-25840-9

Hu, X., Zhao, H.N., Tian, Z., Peter, K.T., Dodd, M.C., Kolodziej, E.P., 2022. Transformation Product Formation upon Heterogeneous Ozonation of the Tire Rubber Antioxidant 6PPD (N-(1,3-dimethylbutyl)-N′-phenyl-p-phenylenediamine). Environ. Sci. Technol. Lett. 9, 413–419. https://doi.org/10.1021/acs.estlett.2c00187

Iravani, S., Conrad, T.O.F., 2022. An Interpretable Deep Learning Approach for Biomarker Detection in LC-MS Proteomics Data. IEEE/ACM Transactions on Computational Biology and Bioinformatics 1–1. https://doi.org/10.1109/TCBB.2022.3141656

Jolliffe, I.T., Cadima, J., 2016. Principal component analysis: a review and recent developments. Philos. Trans. R. Soc. A 374, 20150202. https://doi.org/10.1098/rsta.2015.0202

Kessner, D., Chambers, M., Burke, R., Agus, D., Mallick, P., 2008. ProteoWizard: open source software for rapid proteomics tools development. Bioinformatics 24, 2534–2536. https://doi.org/10.1093/bioinformatics/btn323

Kong, A.T., Leprevost, F.V., Avtonomov, D.M., Mellacheruvu, D., Nesvizhskii, A.I., 2017. MSFragger: ultrafast and comprehensive peptide identification in mass spectrometry-based proteomics. Nat. Methods 14, 513–520. https://doi.org/10.1038/nmeth.4256

Kutlucinar, K.G., Handl, S., Allabashi, R., Causon, T., Troyer, C., Mayr, E., Perfler, R., Hann, S., 2022. Non-targeted analysis with high-resolution mass spectrometry for investigation of riverbank filtration processes. Environ. Sci. Pollut. Res. https://doi.org/10.1007/s11356-022-20301-2

Lee, E.S., Durant, T.J.S., 2022. Supervised machine learning in the mass spectrometry laboratory: A tutorial. Journal of Mass Spectrometry and Advances in the Clinical Lab 23, 1–6. https://doi.org/10.1016/j.jmsacl.2021.12.001

Levitsky, L.I., Klein, J.A., Ivanov, M.V., Gorshkov, M.V., 2019. Pyteomics 4.0: Five Years of Development of a Python Proteomics Framework. J. Proteome Res. 18, 709–714. https://doi.org/10.1021/acs.jproteome.8b00717

Liebal, U.W., Phan, A.N.T., Sudhakar, M., Raman, K., Blank, L.M., 2020. Machine Learning Applications for Mass Spectrometry-Based Metabolomics. Metabolites 10, 243. https://doi.org/10.3390/metabo10060243

Luc, P., n.d. ChemCalc: A Building Block for Tomorrow's Chemical Infrastructure | Journal of Chemical Information and Modeling [WWW Document]. URL https://pubs.acs.org/doi/10.1021/ci300563h (accessed 4.30.21).

Maaten, L. van der, Hinton, G., 2008. Visualizing Data using t-SNE. J. Mach. Learn. Res. 9, 2579–2605.

MassBank of North America [WWW Document], n.d. URL https://mona.fiehnlab.ucdavis.edu/ (accessed 9.29.21).

Matplotlib: A 2D Graphics Environment | IEEE Journals & Magazine | IEEE Xplore [WWW Document], n.d. URL https://ieeexplore.ieee.org/document/4160265 (accessed 11.8.22).

McKinney, W., 2010. Data Structures for Statistical Computing in Python. Presented at the Python in Science Conference, Austin, Texas, pp. 56–61. https://doi.org/10.25080/Majora-92bf1922-00a

Melnikov, A.D., Tsentalovich, Y.P., Yanshole, V.V., 2020. Deep Learning for the Precise Peak Detection in High-Resolution LC–MS Data. Anal. Chem. 92, 588–592. https://doi.org/10.1021/acs.analchem.9b04811

Mucherino, A., Papajorgji, P.J., Pardalos, P.M., 2009. k-Nearest Neighbor Classification, in: Mucherino, A., Papajorgji, P.J., Pardalos, P.M. (Eds.), Data Mining in Agriculture, Springer Optimization and Its Applications. Springer, New York, NY, pp. 83–106. https://doi.org/10.1007/978-0-387-88615-2_4

Mustakim, Indah, R.N.G., Novita, R., Kharisma, O.B., Vebrianto, R., Sanjaya, S., Hasbullah, Andriani, T., Sari, W.P., Novita, Y., Rahim, R., 2019. DBSCAN algorithm: twitter text clustering of trend topic pilkada pekanbaru. J. Phys.: Conf. Ser. 1363, 012001. https://doi.org/10.1088/1742-6596/1363/1/012001

Naz, S., Moreira dos Santos, D.C., García, A., Barbas, C., 2014. Analytical protocols based on LC–MS, GC–MS and CE–MS for nontargeted metabolomics of biological tissues. Bioanalysis 6, 1657–1677. https://doi.org/10.4155/bio.14.119

Negri, L.H., Vestri, C., 2017. lucashn/peakutils: v1.1.0. https://doi.org/10.5281/zenodo.887917

Nikolopoulou, V., Aalizadeh, R., Nika, M.-C., Thomaidis, N.S., 2022. TrendProbe: Time profile analysis of emerging contaminants by LC-HRMS non-target screening and deep learning convolutional neural network. Journal of Hazardous Materials 428, 128194. https://doi.org/10.1016/j.jhazmat.2021.128194

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Köpf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., Chintala, S., 2019. PyTorch: An Imperative Style, High-Performance Deep Learning Library. arXiv:1912.01703 [cs, stat].

Patiny, L., Borel, A., 2013. ChemCalc: A Building Block for Tomorrow's Chemical Infrastructure. J. Chem. Inf. Model. 53, 1223–1228. https://doi.org/10.1021/ci300563h

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, É., 2011. Scikit-learn: Machine Learning in Python. J. Mach. Learn. Res. 12, 2825–2830.

Peter, K.T., Tian, Z., Wu, C., Lin, P., White, S., Du, B., McIntyre, J.K., Scholz, N.L., Kolodziej, E.P., 2018. Using High-Resolution Mass Spectrometry to Identify Organic Contaminants Linked to Urban Stormwater Mortality Syndrome in Coho Salmon. Environ. Sci. Technol. 52, 10317–10327. https://doi.org/10.1021/acs.est.8b03287

Peter, K.T., Wu, C., Tian, Z., Kolodziej, E.P., 2019. Application of Nontarget High Resolution Mass Spectrometry Data to Quantitative Source Apportionment. Environ. Sci. Technol. 53, 12257–12268. https://doi.org/10.1021/acs.est.9b04481

Plotly Technologies Inc., 2015. Collaborative data science.

Pluskal, T., Castillo, S., Villar-Briones, A., Orešič, M., 2010. MZmine 2: Modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. BMC Bioinf. 11, 395. https://doi.org/10.1186/1471-2105-11-395

Python vs Java: What's The Difference? [WWW Document], n.d. . BMC Blogs. URL https://www.bmc.com/blogs/python-vs-java/ (accessed 10.31.22).

Rager, J.E., Strynar, M.J., Liang, S., McMahen, R.L., Richard, A.M., Grulke, C.M., Wambaugh, J.F., Isaacs, K.K., Judson, R., Williams, A.J., Sobus, J.R., 2016. Linking high resolution mass spectrometry data with exposure and toxicity forecasts to advance high-throughput environmental monitoring. Environment International 88, 269–280. https://doi.org/10.1016/j.envint.2015.12.008

Riquelme, G., Zabalegui, N., Marchi, P., Jones, C.M., Monge, M.E., 2020. A Python-Based Pipeline for Preprocessing LC-MS Data for Untargeted Metabolomics Workflows. Metabolites 10, E416. https://doi.org/10.3390/metabo10100416

Rødland, E.S., Samanipour, S., Rauert, C., Okoffo, E.D., Reid, M.J., Heier, L.S., Lind, O.C., Thomas, K.V., Meland, S., 2022. A novel method for the quantification of tire and polymer-modified bitumen particles in environmental samples by pyrolysis gas

chromatography mass spectroscopy. Journal of Hazardous Materials 423, 127092. https://doi.org/10.1016/j.jhazmat.2021.127092

Rosano, T.G., Wood, M., Swift, T.A., 2011. Postmortem Drug Screening by Non-Targeted and Targeted Ultra-Performance Liquid Chromatography-Mass Spectrometry Technology. J Anal Toxicol 35, 411–423. https://doi.org/10.1093/anatox/35.7.411

Röst, H.L., Sachsenberg, T., Aiche, S., Bielow, C., Weisser, H., Aicheler, F., Andreotti, S., Ehrlich, H.-C., Gutenbrunner, P., Kenar, E., Liang, X., Nahnsen, S., Nilse, L., Pfeuffer, J., Rosenberger, G., Rurik, M., Schmitt, U., Veit, J., Walzer, M., Wojnar, D., Wolski, W.E., Schilling, O., Choudhary, J.S., Malmström, L., Aebersold, R., Reinert, K., Kohlbacher, O., 2016. OpenMS: a flexible open-source software platform for mass spectrometry data analysis. Nat. Methods 13, 741–748. https://doi.org/10.1038/nmeth.3959

Ruttkies, C., Schymanski, E.L., Wolf, S., Hollender, J., Neumann, S., 2016. MetFrag relaunched: incorporating strategies beyond in silico fragmentation. J. Cheminf. 8, 3. https://doi.org/10.1186/s13321-016-0115-9

Schmitt, U., 2018. blosloos/enviMass: enviMass version 3.5. https://doi.org/10.5281/zenodo.1213098

Schollée, J.E., Bourgin, M., von Gunten, U., McArdell, C.S., Hollender, J., 2018. Non-target screening to trace ozonation transformation products in a wastewater treatment train including different post-treatments. Water Res. 142, 267–278. https://doi.org/10.1016/j.watres.2018.05.045

Smith, C.A., Want, E.J., O'Maille, G., Abagyan, R., Siuzdak, G., 2006. XCMS: Processing Mass Spectrometry Data for Metabolite Profiling Using Nonlinear Peak Alignment, Matching, and Identification. Anal. Chem. 78, 779 – 787. https://doi.org/10.1021/ac051437y

Tian, Z., Peter, K.T., Gipe, A.D., Zhao, H., Hou, F., Wark, D.A., Khangaonkar, T., Kolodziej, E.P., James, C.A., 2020. Suspect and Nontarget Screening for Contaminants of Emerging Concern in an Urban Estuary. Environ. Sci. Technol. 54, 889–901. https://doi.org/10.1021/acs.est.9b06126

Tian, Z., Zhao, H., Peter, K.T., Gonzalez, M., Wetzel, J., Wu, C., Hu, X., Prat, J., Mudrock, E., Hettinger, R., Cortina, A.E., Biswas, R.G., Kock, F.V.C., Soong, R., Jenne, A., Du, B., Hou, F., He, H., Lundeen, R., Gilbreath, A., Sutton, R., Scholz, N.L., Davis, J.W., Dodd, M.C., Simpson, A., McIntyre, J.K., Kolodziej, E.P., 2021. A ubiquitous tire rubber–derived chemical induces acute mortality in coho salmon. Science 371, 185–189. https://doi.org/10.1126/science.abd6951

Till, B., 2012. pymzML—Python module for high-throughput bioinformatics on mass spectrometry data [WWW Document]. Oxford Academic. URL https://academic.oup.com/bioinformatics/article/28/7/1052/209917 (accessed 8.2.22).

Tkalec, Ž., Codling, G., Klánová, J., Horvat, M., Kosjek, T., 2022. LC-HRMS based method for suspect/non-targeted screening for biomarkers of chemical exposure in human urine. Chemosphere 300, 134550. https://doi.org/10.1016/j.chemosphere.2022.134550

Tsugawa, H., Cajka, T., Kind, T., Ma, Y., Higgins, B., Ikeda, K., Kanazawa, M., VanderGheynst, J., Fiehn, O., Arita, M., 2015. MS-DIAL: Data Independent MS/MS

Deconvolution for Comprehensive Metabolome Analysis. Nat. Methods 12, 523–526. https://doi.org/10.1038/nmeth.3393

Ulrich, E.M., Sobus, J.R., Grulke, C.M., Richard, A.M., Newton, S.R., Strynar, M.J., Mansouri, K., Williams, A.J., 2019a. EPA's non-targeted analysis collaborative trial (ENTACT): genesis, design, and initial findings. Anal Bioanal Chem 411, 853–866. https://doi.org/10.1007/s00216-018-1435-6

Ulrich, E.M., Sobus, J.R., Grulke, C.M., Richard, A.M., Newton, S.R., Strynar, M.J., Mansouri, K., Williams, A.J., 2019b. EPA's non-targeted analysis collaborative trial (ENTACT): genesis, design, and initial findings. Anal. Bioanal. Chem. 411, 853–866. https://doi.org/10.1007/s00216-018-1435-6

Uppal, K., Soltow, Q.A., Strobel, F.H., Pittard, W.S., Gernert, K.M., Yu, T., Jones, D.P., 2013. xMSanalyzer: automated pipeline for improved feature detection and downstream analysis of large-scale, non-targeted metabolomics data. BMC Bioinf 14, 15. https://doi.org/10.1186/1471-2105-14-15

Van, R., G., 2020. pickle — Python object serialization — Python 3.9.7 documentation [WWW Document]. URL https://docs.python.org/3/library/pickle.html (accessed 9.30.21).

Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S.J., Brett, M., Wilson, J., Millman, K.J., Mayorov, N., Nelson, A.R.J., Jones, E., Kern, R., Larson, E., Carey, C.J., Polat, İ., Feng, Y., Moore, E.W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E.A., Harris, C.R., Archibald, A.M., Ribeiro, A.H., Pedregosa, F., van Mulbregt, P., 2020. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat. Methods 17, 261–272. https://doi.org/10.1038/s41592-019-0686-2

Wang, M., Carver, J.J., Phelan, V.V., Sanchez, L.M., Garg, N., Peng, Y., Nguyen, D.D., Watrous, J., Kapono, C.A., Luzzatto-Knaan, T., Porto, C., Bouslimani, A., Melnik, A.V., Meehan, M.J., Liu, W.-T., Crüsemann, M., Boudreau, P.D., Esquenazi, E., Sandoval-Calderón, M., Kersten, R.D., Pace, L.A., Quinn, R.A., Duncan, K.R., Hsu, C.-C., Floros, D.J., Gavilan, R.G., Kleigrewe, K., Northen, T., Dutton, R.J., Parrot, D., Carlson, E.E., Aigle, B., Michelsen, C.F., Jelsbak, L., Sohlenkamp, C., Pevzner, P., Edlund, A., McLean, J., Piel, J., Murphy, B.T., Gerwick, L., Liaw, C.-C., Yang, Y.-L., Humpf, H.-U., Maansson, M., Keyzers, R.A., Sims, A.C., Johnson, A.R., Sidebottom, A.M., Sedio, B.E., Klitgaard, A., Larson, C.B., P., C.A.B., Torres-Mendoza, D., Gonzalez, D.J., Silva, D.B., Marques, L.M., Demarque, D.P., Pociute, E., O'Neill, E.C., Briand, E., Helfrich, E.J.N., Granatosky, E.A., Glukhov, E., Ryffel, F., Houson, H., Mohimani, H., Kharbush, J.J., Zeng, Y., Vorholt, J.A., Kurita, K.L., Charusanti, P., McPhail, K.L., Nielsen, K.F., Vuong, L., Elfeki, M., Traxler, M.F., Engene, N., Koyama, N., Vining, O.B., Baric, R., Silva, R.R., Mascuch, S.J., Tomasi, S., Jenkins, S., Macherla, V., Hoffman, T., Agarwal, V., Williams, P.G., Dai, J., Neupane, R., Gurr, J., Rodríguez, A.M.C., Lamsa, A., Zhang, C., Dorrestein, K., Duggan, B.M., Almaliti, J., Allard, P.-M., Phapale, P., Nothias, L.-F., Alexandrov, T., Litaudon, M., Wolfender, J.-L., Kyle, J.E., Metz, T.O., Peryea, T., Nguyen, D.-T., VanLeer, D., Shinn, P., Jadhav, A., Müller, R., Waters, K.M., Shi, W., Liu, X., Zhang, L., Knight, R., Jensen, P.R., Palsson, B.O., Pogliano, K., Linington, R.G., Gutiérrez, M., Lopes, N.P., Gerwick,

W.H., Moore, B.S., Dorrestein, P.C., Bandeira, N., 2016. Sharing and community curation of mass spectrometry data with GNPS. Nat. Biotechnol. 34, 828–837. https://doi.org/10.1038/nbt.3597

Wang, T., Duedahl-Olesen, L., Lauritz Frandsen, H., 2021. Targeted and non-targeted unexpected food contaminants analysis by LC/HRMS: Feasibility study on rice. Food Chem. 338, 127957. https://doi.org/10.1016/j.foodchem.2020.127957

Wang, Z., Walker, G.W., Muir, D.C.G., Nagatani-Yoshida, K., 2020. Toward a Global Understanding of Chemical Pollution: A First Comprehensive Analysis of National and Regional Chemical Inventories. Environ. Sci. Technol. 54, 2575–2584. https://doi.org/10.1021/acs.est.9b06379

Wu, A.H., Gerona, R., Armenian, P., French, D., Petrie, M., Lynch, K.L., 2012. Role of liquid chromatography–high-resolution mass spectrometry (LC-HR/MS) in clinical toxicology. Clin Toxicol 50, 733–742. https://doi.org/10.3109/15563650.2012.713108

Xue, J., Lai, Y., Liu, C.-W., Ru, H., 2019. Towards Mass Spectrometry-Based Chemical Exposome: Current Approaches, Challenges, and Future Directions. Toxics 7, 41. https://doi.org/10.3390/toxics7030041

Yin, P., Xu, G., 2014. Current state-of-the-art of nontargeted metabolomics based on liquid chromatography–mass spectrometry with special emphasis in clinical applications. J Chromatogr A 1374, 1–13. https://doi.org/10.1016/j.chroma.2014.11.050

Yunker, L., Yeung, D., McIndoe, J.S., 2018. PythoMS: A Python Framework to Simplify and Assist in the Processing and Interpretation of Mass Spectrometric Data. https://doi.org/10.26434/chemrxiv.7264175.v1

Zhao, H.N., Tian, Z., Kim, K.E., Wang, R., Lam, K., Kolodziej, E.P., 2021. Biotransformation of Current-Use Progestin Dienogest and Drospirenone in Laboratory-Scale Activated Sludge Systems Forms High-Yield Products with Altered Endocrine Activity. Environ. Sci. Technol. 55, 13869–13880. https://doi.org/10.1021/acs.est.1c03805

Zhao, Y., Liu, X., Li, X., 2018. An improved DBSCAN algorithm based on cell-like P systems with promoters and inhibitors. PLoS One 13, e0200751. https://doi.org/10.1371/journal.pone.0200751

Supplementary Materials

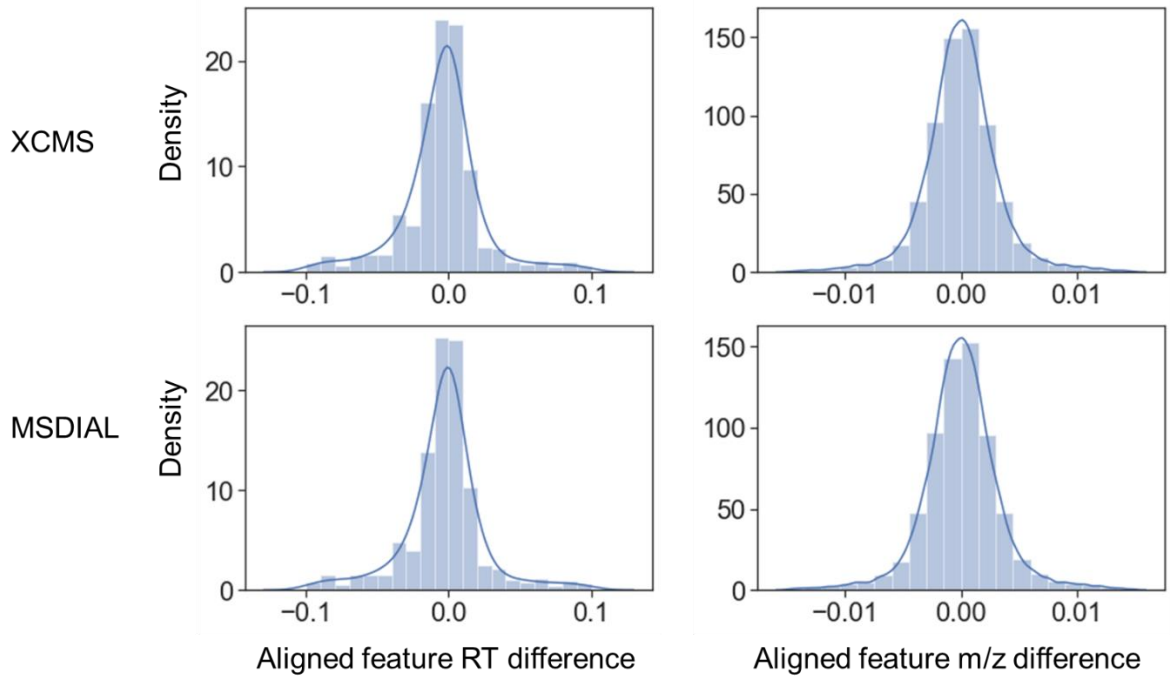


Figure S5.1. RT and *m/z* difference distribution of the aligned features between different software. The comparisons were separately conducted for *XCMS* or *MSDIAL*, each relative to *MSS*. All the RT and *m/z* values from MSS-extracted features were subtracted from the corresponding values from its closest neighbor (closest *m/z* and RT pair, determined by Euclidian distance) in *XCMS* or *MSDIAL* for comparison. The comparison results for RT and *m/z* differences from all 3 testing samples (ENTACT #505, #506 and #508) were integrated. The hist bins represent the original data, while the curves are kernel density estimation curves that represent probability densities calculated from the original data.

Table 5.2.A. List of spiked chemicals for feature extraction validation samples (detected; EPA ENTACT study) (Ulrich et al., 2019b)

Sample #	*Name*	*Formula*	*Mass*	*RT*
505	2-Methyl-4-amino-6-methoxy-s-triazine	C5 H8 N4 O	140.07	3.3
505	1-Naphthol	C10 H8 O	144.06	6.5
505	1,3-Dipropan-2-ylurea	C7 H16 N2 O	144.13	4.4
505	1,2-Benzisothiazolin-3-one	C7 H5 N O S	151.01	4.7
505	Acetaminophen	C8 H9 N O2	151.06	3.1
505	2-Isopropyl-6-methyl-4-pyrimidone	C8 H12 N2 O	152.09	4.1
505	2-Amino-5-nitrophenol	C6 H6 N2 O3	154.04	11.2
505	Ethionamide	C8 H10 N2 S	166.06	4.7
505	6-Pentyl-2H-pyran-2-one	C10 H14 O2	166.10	9.4
505	2-(Ethyl(3-methylphenyl)amino)acetonitrile	C11 H14 N2	174.12	7.7
505	Propham	C10 H13 N O2	179.09	5.3
505	Hexamethylphosphoramide	C6 H18 N3 O P	179.12	4.9
505	N-(3-Chloro-4-methylphenyl)acetamide	C9 H10 Cl N O	183.05	7.0
505	Phenazone	C11 H12 N2 O	188.09	4.5
505	Propamocarb	C9 H20 N2 O2	188.15	3.2
505	Aldicarb	C7 H14 N2 O2 S	190.08	5.4
505	Isoprocarb	C11 H15 N O2	193.11	7.2
505	Methiuron	C10 H14 N2 S	194.09	5.4
505	Methyl 2,4-dihydroxy-3,6-dimethylbenzoate	C10 H12 O4	196.07	5.3
505	Dibenzylamine	C14 H15 N	197.12	4.8
505	Azacyclotridecan-2-one	C12 H23 N O	197.18	8.6
505	Pyrimethanil	C12 H13 N3	199.11	8.5
505	Carbaryl	C12 H11 N O2	201.08	6.5
505	Metamitron	C10 H10 N4 O	202.08	4.7
505	7-(Dimethylamino)-4-methylcoumarin	C12 H13 N O2	203.09	7.2
505	Panthenol	C9 H19 N O4	205.13	3.1
505	5,7-Dimethoxy-2H-chromen-2-one	C11 H10 O4	206.06	5.9
505	Ethylphenylacetylurea	C11 H14 N2 O2	206.11	6.5
505	N-(4-Methoxyphenyl)-3-oxobutanamide	C11 H13 N O3	207.09	4.7
505	Propoxur	C11 H15 N O3	209.11	6.1
505	Ethirimol	C11 H19 N3 O	209.15	7.2
505	4-Methoxy-2-methyl-N-phenylaniline	C14 H15 N O	213.11	10.1
505	Pymetrozine	C10 H11 N5 O	217.09	3.9
505	Primidone	C12 H14 N2 O2	218.11	6.9
505	Quinmerac	C11 H8 Cl N O2	221.02	4.2
505	Chloridazon	C10 H8 Cl N3 O	221.04	4.8
505	Diethyl phthalate	C12 H14 O4	222.09	7.6
505	Stavudine	C10 H12 N2 O4	224.08	11.6
505	Secbumeton	C10 H19 N5 O	225.16	8.0

505	Ametryn	C9 H17 N5 S	227.12	8.4
505	7-Diethylamino-4-methylcoumarin	C14 H17 N O2	231.13	9.2
505	Carbetamide	C12 H16 N2 O3	236.12	5.7
505	1-Amino-2-methylanthraquinone	C15 H11 N O2	237.08	9.7
505	2-(Thiocyanomethylthio)benzothiazole	C9 H6 N2 S3	237.97	8.4
505	Pirimicarb	C11 H18 N4 O2	238.14	6.9
505	Dimetilan	C10 H16 N4 O3	240.12	4.9
505	N,N'-Dicyclohexylthiourea	C13 H24 N2 S	240.17	9.5
505	Linuron	C9 H10 Cl2 N2 O2	248.01	8.4
505	C.I. Solvent Yellow 14	C16 H12 N2 O	248.09	13.8
505	2,4,6-Tris(allyloxy)-1,3,5-triazine	C12 H15 N3 O3	249.11	9.4
505	Diisopropyl phthalate	C14 H18 O4	250.12	10.1
505	2,2'-(Oxydimethanediyl)bis(2-ethylpropane-1,3-diol)	C12 H26 O5	250.18	7.2
505	2,6-Di-tert-butyl-4-nitrophenol	C14 H21 N O3	251.15	8.1
505	Imidacloprid	C9 H10 Cl N5 O2	255.05	4.5
505	Dimethametryn	C11 H21 N5 S	255.15	10.6
505	Ifosfamide	C7 H15 Cl2 N2 O2 P	260.02	5.2
505	Formononetin	C16 H12 O4	268.07	8.0
505	Michler's ketone	C17 H20 N2 O	268.16	10.2
505	Acetochlor	C14 H20 Cl N O2	269.12	10.0
505	Lauryldiethanolamine	C16 H35 N O2	273.27	10.3
505	2,7-Acetylaminofluorene	C17 H16 N2 O2	280.12	5.7
505	Flufenamic acid	C14 H10 F3 N O2	281.07	12.9
505	Pendimethalin	C13 H19 N3 O4	281.14	13.6
505	Di(2-methoxyethyl) phthalate	C14 H18 O6	282.11	6.0
505	Penconazole	C13 H15 Cl2 N3	283.06	10.8
505	PharmaGSID_47263	C15 H16 N4 O2	284.13	6.0
505	Piperine	C17 H19 N O3	285.14	9.7
505	Chloroxuron	C15 H15 Cl N2 O2	290.08	9.6
505	Crufomate	C12 H19 Cl N O3 P	291.08	10.5
505	Butralin	C14 H21 N3 O4	295.15	14.1
505	Sulfaquinoxaline	C14 H12 N4 O2 S	300.07	5.2
505	17-Methyltestosterone	C20 H30 O2	302.22	9.9
505	Imazamox	C15 H19 N3 O4	305.14	4.3
505	Buprofezin	C16 H23 N3 O S	305.16	13.1
505	Tebuconazole	C16 H22 Cl N3 O	307.15	11.0
505	Tolazamide	C14 H21 N3 O3 S	311.13	7.3
505	Benzyl butyl phthalate	C19 H20 O4	312.14	12.3
505	Norgestrel	C21 H28 O2	312.21	9.8
505	Amoxapine	C17 H16 Cl N3 O	313.10	7.0
505	Propane-1,3-diyl bis(4-aminobenzoate)	C17 H18 N2 O4	314.13	6.1

505	Ranitidine	C13 H22 N4 O3 S	314.14	3.1
505	Oxfendazole	C15 H13 N3 O3 S	315.07	5.9
505	Bupirimate	C13 H24 N4 O3 S	316.16	10.5
505	Zearalenone	C18 H22 O5	318.15	9.4
505	Tomelukast	C16 H22 N4 O3	318.17	9.1
505	Pirinixic acid	C14 H14 Cl N3 O2 S	323.05	8.4
505	Flutolanil	C17 H16 F3 N O2	323.11	9.1
505	5HPP-33	C20 H21 N O3	323.15	11.1
505	Octabenzone	C21 H26 O3	326.19	12.2
505	Epoxiconazole	C17 H13 Cl F N3 O	329.07	10.1
505	Fenarimol	C17 H12 Cl2 N2 O	330.03	9.9
505	Halofenozide	C18 H19 Cl N2 O2	330.11	8.7
505	Dicyclohexyl phthalate	C20 H26 O4	330.18	14.2
505	17alpha-Hydroxyprogesterone	C21 H30 O3	330.22	9.6
505	Ipconazole	C18 H24 Cl N3 O	333.16	12.4
505	Cloquintocet-mexyl	C18 H22 Cl N O3	335.13	13.2
505	Famotidine	C8 H15 N7 O2 S3	337.04	3.1
505	Danazol	C22 H27 N O2	337.20	11.4
505	Topiramate	C12 H21 N O8 S	339.10	5.4
505	Thiophanate-methyl	C12 H14 N4 O4 S2	342.05	6.1
505	Triflumizole	C15 H15 Cl F3 N3 O	345.09	12.4
505	AM580	C22 H25 N O3	351.18	12.2
505	Carabersat	C20 H20 F N O4	357.14	6.7
505	Chlorfenvinphos	C12 H14 Cl3 O4 P	357.97	11.6
505	Triethylene glycol dibenzoate	C20 H22 O6	358.14	10.2
505	Florasulam	C12 H8 F3 N5 O3 S	359.03	5.1
505	Haloxyfop	C15 H11 Cl F3 N O4	361.03	9.0
505	Flufenacet	C14 H13 F4 N3 O2 S	363.07	9.9
505	Flamprop-isopropyl	C19 H19 Cl F N O3	363.10	11.2
505	Pyridaben	C19 H25 Cl N2 O S	364.14	14.8
505	Picoxystrobin	C18 H16 F3 N O4	367.10	10.5
505	Methoxyfenozide	C22 H28 N2 O3	368.21	9.3
505	Tamoxifen	C26 H29 N O	371.22	11.2
505	Quizalofop-ethyl	C19 H17 Cl N2 O4	372.09	12.7
505	FR167356	C19 H17 Cl2 N O3	377.06	8.3
505	CP-457677	C22 H21 F N2 O3	380.15	8.2
505	Fluazifop-P-butyl	C19 H20 F3 N O4	383.13	12.9
505	Pyraclostrobin	C19 H18 Cl N3 O4	387.10	11.5
505	PD-0333941	C23 H22 N2 O4	390.15	12.0
505	Dialifor	C14 H17 Cl N O4 P S2	393.00	11.8

505	Triamcinolone	C21 H27 F O6	394.18	5.7
505	Melengestrol acetate	C25 H32 O4	396.23	11.1
505	Sulfasalazine	C18 H14 N4 O5 S	398.07	5.6
505	Colchicine	C22 H25 N O6	399.17	5.8
505	4-[3-(4-Acetyl-3-hydroxy-2-propylphenoxy)propoxy]phenoxy-acetic acid	C22 H26 O7	402.17	10.4
505	CP-114271	C17 H19 F3 N2 O4 S	404.10	5.0
505	Forskolin	C22 H34 O7	410.23	9.4
505	Pyraflufen-ethyl	C15 H13 Cl2 F3 N2 O4	412.02	11.2
505	Cyproterone acetate	C24 H29 Cl O4	416.18	10.5
505	Fenpyroximate (Z,E)	C24 H27 N3 O4	421.20	14.3
505	Flumiclorac-pentyl	C21 H23 Cl F N O5	423.13	13.0
505	Flurandrenolide	C24 H33 F O6	436.23	10.9
505	Ingliforib	C23 H24 Cl N3 O5	457.14	9.1
505	CI-1029	C28 H37 N O4 S	483.24	6.8
505	Hydramethylnon	C25 H24 F6 N4	494.19	12.4
505	Surinabant	C23 H23 Br Cl2 N4 O	520.04	14.6
505	PharmaGSID_48511	C27 H35 F2 N7 O3	543.28	7.2
505	PharmaGSID_48521	C31 H49 N5 O4	555.38	5.2
505	Etoposide	C29 H32 O13	588.18	5.9
505	Nelivaptan	C30 H32 Cl N3 O8 S	629.16	10.9
505	Flubendiamide	C23 H22 F7 I N2 O4 S	682.02	11.0
506	N,N'-Methylenebisacrylamide	C7 H10 N2 O2	154.07	2.7
506	6-Methyl coumarin	C10 H8 O2	160.05	6.1
506	1-Cyclohexylpyrrolidin-2-one	C10 H17 N O	167.13	6.3
506	Metronidazole	C6 H9 N3 O3	171.06	3.2
506	Phenacetin	C10 H13 N O2	179.09	5.3
506	N-(3-Amino-4-methoxyphenyl)acetamide	C9 H12 N2 O2	180.09	3.8
506	Carbimazole	C7 H10 N2 O2 S	186.05	4.2
506	1-(4-Methoxyphenyl)-1-pentene-3-one	C12 H14 O2	190.10	12.7
506	N-Octyl-2-pyrrolidone	C12 H23 N O	197.18	10.6
506	Monuron	C9 H11 Cl N2 O	198.06	5.9
506	Tacrine	C13 H14 N2	198.12	4.6
506	Ronidazole	C6 H8 N4 O4	200.05	3.1
506	Carbaryl	C12 H11 N O2	201.08	6.5
506	N,N-Dimethyldecylamine oxide	C12 H27 N O	201.21	9.1
506	Indole-3-butyric acid	C12 H13 N O2	203.09	5.3
506	3,4-Dichlorophenylurea	C7 H6 Cl2 N2 O	203.99	6.8
506	Isoproturon	C12 H18 N2 O	206.14	7.4
506	Promecarb	C12 H17 N O2	207.13	11.1

506	1,1,2-Trimethyl-1H-benzo[e]indole	C15 H15 N	209.12	9.6
506	Atraton	C9 H17 N5 O	211.14	6.9
506	Simetryn	C8 H15 N5 S	213.10	7.2
506	Metribuzin	C8 H14 N4 O S	214.09	6.1
506	Nootkatone	C15 H22 O	218.17	11.4
506	Thidiazuron	C9 H8 N4 O S	220.04	6.2
506	Methabenzthiazuron	C10 H11 N3 O S	221.06	7.2
506	Flavone	C15 H10 O2	222.07	9.2
506	Diethyl phthalate	C12 H14 O4	222.09	7.6
506	Tebuthiuron	C9 H16 N4 O S	228.10	6.4
506	Icaridin	C12 H23 N O3	229.17	8.1
506	4-Dimethylaminoantipyrine	C13 H17 N3 O	231.14	4.9
506	2-(Phenylsulfonyl)aniline	C12 H11 N O2 S	233.05	6.0
506	Butam	C15 H23 N O	233.18	10.1
506	Carboxin	C12 H13 N O2 S	235.07	6.5
506	1,4-Diaminoanthraquinone	C14 H10 N2 O2	238.07	6.7
506	Clomazone	C12 H14 Cl N O2	239.07	7.9
506	Diphenamid	C16 H17 N O	239.13	7.8
506	Quinclorac	C10 H5 Cl2 N O2	240.97	4.3
506	Phenethyl anthranilate	C15 H15 N O2	241.11	10.7
506	1,1'-Diazene-1,2-diyldicyclohexanecarbonitrile	C14 H20 N4	244.17	9.7
506	Clothianidin	C6 H8 Cl N5 O2 S	249.01	4.5
506	Hexazinone	C12 H20 N4 O2	252.16	6.2
506	Triamterene	C12 H11 N7	253.11	5.1
506	Thiobencarb	C12 H16 Cl N O S	257.06	11.6
506	Metobromuron	C9 H11 Br N2 O2	258.00	7.1
506	Naphthalen-2-yl 2-aminobenzoate	C17 H13 N O2	263.09	11.8
506	PharmaGSID_47261	C13 H15 N O3 S	265.08	5.9
506	Albendazole	C12 H15 N3 O2 S	265.09	8.7
506	Sulfisoxazole	C11 H13 N3 O3 S	267.07	4.4
506	Alachlor	C14 H20 Cl N O2	269.12	10.0
506	Methyl red	C15 H15 N3 O2	269.12	10.0
506	Tolbutamide	C12 H18 N2 O3 S	270.10	7.0
506	Carbutamide	C11 H17 N3 O3 S	271.10	5.3
506	Napropamide	C17 H21 N O2	271.16	10.0
506	N-Dodecanoyl-N-methylglycine	C15 H29 N O3	271.21	11.6
506	Paraoxon	C10 H14 N O6 P	275.06	7.1
506	Dimethenamid	C12 H18 Cl N O2 S	275.07	8.7
506	Imazapic	C14 H17 N3 O3	275.13	4.5
506	Ethyl 2-cyano-3,3-diphenylacrylate	C18 H15 N O2	277.11	9.9
506	Oxadixyl	C14 H18 N2 O4	278.13	5.6
506	Metalaxyl	C15 H21 N O4	279.15	7.5
506	Karbutilate	C14 H21 N3 O3	279.16	6.2

506	Ethyl phthalyl ethyl glycolate	C14 H16 O6	280.10	7.6
506	Fosthiazate	C9 H18 N O3 P S2	283.05	6.9
506	N-[3-(Dimethylamino)propyl]dodecanamide	C17 H36 N2 O	284.28	10.7
506	HC Blue 2	C12 H19 N3 O5	285.13	4.5
506	Myclobutanil	C15 H17 Cl N4	288.11	9.4
506	PHA-00568487	C16 H20 N2 O3	288.15	3.9
506	3-((Ethylphenylamino)methyl)benzenesulfonic acid	C15 H17 N O3 S	291.09	6.3
506	4-(2,5-Diethoxy-4-nitrophenyl)morpholine	C14 H20 N2 O5	296.14	8.6
506	CP-409092	C17 H19 N3 O2	297.15	4.7
506	Tolclofos-methyl	C9 H11 Cl2 O3 P S	299.95	9.5
506	Flutriafol	C16 H13 F2 N3 O	301.10	7.3
506	Fluconazole	C13 H12 F2 N6 O	306.10	5.9
506	Dipentyl phthalate	C18 H26 O4	306.18	14.1
506	Quinoxyfen	C15 H8 Cl2 F N O	307.00	13.4
506	Tolnaftate	C19 H17 N O S	307.10	12.7
506	Diflubenzuron	C14 H9 Cl F2 N2 O2	310.03	10.4
506	Butachlor	C17 H26 Cl N O2	311.16	13.2
506	Neopentyl glycol dibenzoate	C19 H20 O4	312.14	12.3
506	Diethylene glycol dibenzoate	C18 H18 O5	314.12	10.0
506	Progesterone	C21 H30 O2	314.22	11.1
506	Oxyphenisatin	C20 H15 N O3	317.11	6.1
506	Metconazole	C17 H22 Cl N3 O	319.14	11.5
506	Iprovalicarb	C18 H28 N2 O3	320.21	9.8
506	Sulprofos	C12 H19 O2 P S3	322.03	13.7
506	Octrizole	C20 H25 N3 O	323.20	16.8
506	Acetohexamide	C15 H20 N2 O4 S	324.11	6.7
506	Benalaxyl	C20 H23 N O3	325.17	11.1
506	Triphenyl phosphate	C18 H15 O4 P	326.07	11.3
506	Fluridone	C19 H14 F3 N O	329.10	8.3
506	Dicyclohexyl phthalate	C20 H26 O4	330.18	14.2
506	Fenbuconazole	C19 H17 Cl N4	336.11	10.3
506	Butylphthalyl butylglycolate	C18 H24 O6	336.16	12.1
506	CP-863187	C18 H14 F2 N4 O	340.11	7.3
506	Oxadiazon	C15 H18 Cl2 N2 O3	344.07	13.2
506	Testosterone propionate	C22 H32 O3	344.24	13.3
506	Azinphos-ethyl	C12 H16 N3 O3 P S2	345.04	9.8
506	Meloxicam	C14 H13 N3 O4 S2	351.03	6.4
506	PK 11195	C21 H21 Cl N2 O	352.13	11.3
506	PharmaGSID_48505	C17 H12 Cl N5 O2	353.07	9.2
506	2,2,4-Trimethylpentane-1,3-diyl dibenzoate	C22 H26 O4	354.18	14.2

506	Cyhalofop-butyl	C20 H20 F N O4	357.13	7.6
506	Triflumuron	C15 H10 Cl F3 N2 O3	358.03	11.6
506	FR140423	C18 H16 F2 N2 O2 S	362.09	7.4
506	Z-Tetrachlorvinphos	C10 H9 Cl4 O4 P	363.90	10.6
506	Indapamide	C16 H16 Cl N3 O3 S	365.06	6.0
506	Finasteride	C23 H36 N2 O2	372.28	9.7
506	Pyrazophos	C14 H20 N3 O5 P S	373.09	11.7
506	Spirotetramat	C21 H27 N O5	373.19	9.9
506	Prochloraz	C15 H16 Cl3 N3 O2	375.03	11.5
506	CP-401387	C23 H29 N3 O2	379.23	11.7
506	Ethion	C9 H22 O4 P2 S4	383.99	13.4
506	UK-373911	C17 H18 Cl2 N2 O2 S	384.05	6.1
506	Megestrol acetate	C24 H32 O4	384.23	10.8
506	Buspirone	C21 H31 N5 O2	385.25	5.4
506	Betamethasone	C22 H29 F O5	392.20	7.8
506	Diflufenican	C19 H11 F5 N2 O2	394.07	12.2
506	Bensulide	C14 H24 N O4 P S3	397.06	10.7
506	Tris(2-butoxyethyl) phosphate	C18 H39 O7 P	398.24	13.0
506	Fenchlorazole-ethyl	C12 H8 Cl5 N3 O2	400.91	11.5
506	SAR115740	C24 H17 F2 N3 O	401.13	11.0
506	AVE5638	C25 H23 F N2 O2	402.17	8.7
506	Acetyl tributyl citrate	C20 H34 O8	402.23	13.7
506	Diclosulam	C13 H10 Cl2 F N5 O3 S	404.99	6.7
506	Difenoconazole	C19 H17 Cl2 N3 O3	405.06	12.0
506	SSR161421	C25 H20 N4 O2	408.16	9.2
506	Carfentrazone-ethyl	C15 H14 Cl2 F3 N3 O3	411.04	10.8
506	Cyclopamine	C27 H41 N O2	411.31	7.4
506	SR271425	C22 H27 N3 O3 S	413.18	5.9
506	Zamifenacin	C27 H29 N O3	415.21	8.8
506	CP-728663	C26 H33 N3 O2	419.26	6.0
506	Triamcinolone acetonide	C24 H31 F O6	434.21	8.0
506	Glipizide	C21 H27 N5 O4 S	445.18	7.7
506	Pravastatin sodium	C23 H35 Na O7	446.23	7.5
506	Fluocinolone acetonide	C24 H30 F2 O6	452.20	8.0
506	2,2'-Dibenzoylaminodiphenyl disulfide	C26 H20 N2 O2 S2	456.10	12.3
506	Butafenacil	C20 H18 Cl F3 N2 O6	474.08	9.9
506	Betamethasone valerate	C27 H37 F O6	476.26	11.3

506	PharmaGSID_48519	C22 H20 Br N5 O4	497.07	8.6
506	AVE9423	C23 H12 Cl2 F2 N2 O5	504.01	8.8
506	Farglitazar	C34 H30 N2 O5	546.22	13.5
506	SSR69071	C27 H32 N4 O7 S	556.20	7.6
506	Atorvastatin	C33 H35 F N2 O5	558.25	10.1
506	Ouabain	C29 H44 O12	584.28	4.3
506	CP-532623	C27 H27 F9 N2 O3	598.19	14.2
506	Sucrose octaacetate	C28 H38 O19	678.20	7.6
508	sec-Butylurea	C5 H12 N2 O	116.09	3.5
508	Benzimidazole	C7 H6 N2	118.05	2.8
508	N,N-Dimethylacetoacetamide	C6 H11 N O2	129.08	3.9
508	2-Aminobenzimidazole	C7 H7 N3	133.06	3.3
508	N-Hydroxybenzamide	C7 H7 N O2	137.05	6.1
508	1-Naphthol	C10 H8 O	144.06	6.5
508	Methyl trans-styryl ketone	C10 H10 O	146.07	13.3
508	2-Acetamidophenol	C8 H9 N O2	151.06	3.1
508	N-(Butoxymethyl)acrylamide	C8 H15 N O2	157.11	5.1
508	6-Nitrobenzimidazole	C7 H5 N3 O2	163.04	4.5
508	2,2-Dimethyl-2,3-dihydro-1-benzofuran-7-amine	C10 H13 N O	163.10	5.4
508	N-(3-Aminophenyl)propanamide	C9 H12 N2 O	164.09	4.6
508	Benzocaine	C9 H11 N O2	165.08	12.1
508	Terephthalic acid	C8 H6 O4	166.03	5.8
508	Ethionamide	C8 H10 N2 S	166.06	4.7
508	2-(Ethyl(3-methylphenyl)amino)acetonitrile	C11 H14 N2	174.12	7.2
508	4-(Dimethylamino)phenylthiocyanate	C9 H10 N2 S	178.06	8.2
508	Propham	C10 H13 N O2	179.09	5.3
508	Theobromine	C7 H8 N4 O2	180.06	3.1
508	Fuberidazole	C11 H8 N2 O	184.06	5.7
508	N,N'-Dibutylthiourea	C9 H20 N2 S	188.13	7.2
508	Caffeine	C8 H10 N4 O2	194.08	4.1
508	Methiuron	C10 H14 N2 S	194.09	5.4
508	4-Methylbenzophenone	C14 H12 O	196.09	10.0
508	N-Octyl-2-pyrrolidone	C12 H23 N O	197.18	10.6
508	4-Nitrosodiphenylamine	C12 H10 N2 O	198.08	7.7
508	2,2'-[Benzene-1,3-diylbis(oxy)]diethanol	C10 H14 O4	198.09	4.8
508	Ronidazole	C6 H8 N4 O4	200.05	3.1
508	Simazine	C7 H12 Cl N5	201.08	6.1
508	Carbaryl	C12 H11 N O2	201.08	6.5
508	Metamitron	C10 H10 N4 O	202.08	4.7
508	2-Butyl-1H-isoindole-1,3(2H)-dione	C12 H13 N O2	203.09	7.2
508	Panthenol	C9 H19 N O4	205.13	3.1
508	5,7-Dimethoxy-2H-chromen-2-one	C11 H10 O4	206.06	6.7

508	Milrinone	C12 H9 N3 O	211.07	3.8
508	Salicylanilide	C13 H11 N O2	213.08	7.8
508	N-Butylbenzenesulfonamide	C10 H15 N O2 S	213.08	6.8
508	Simetryn	C8 H15 N5 S	213.10	7.2
508	Monolinuron	C9 H11 Cl N2 O2	214.05	6.7
508	8-Methoxypsoralen	C12 H8 O4	216.04	6.8
508	Primidone	C12 H14 N2 O2	218.11	4.8
508	Acetamiprid	C10 H11 Cl N4	222.07	4.8
508	2-Methylanthraquinone	C15 H10 O2	222.07	9.2
508	Vanillin isobutyrate	C12 H14 O4	222.09	7.6
508	Aziprotryne	C7 H11 N7 S	225.08	9.0
508	Ethiofencarb	C11 H15 N O2 S	225.08	6.7
508	Lamivudine	C8 H11 N3 O3 S	229.05	2.7
508	Propazine	C9 H16 Cl N5	229.11	8.7
508	Methfuroxam	C14 H15 N O2	229.11	8.0
508	Fluometuron	C10 H11 F3 N2 O	232.08	6.8
508	Aminoglutethimide	C13 H16 N2 O2	232.12	5.0
508	Cyanazine	C9 H13 Cl N6	240.09	5.8
508	N,N'-Dicyclohexylthiourea	C13 H24 N2 S	240.17	9.5
508	Forchlorfenuron	C12 H10 Cl N3 O	247.05	7.7
508	Dapsone	C12 H12 N2 O2 S	248.06	4.5
508	Cinchophen	C16 H11 N O2	249.08	6.4
508	Dipropyl phthalate	C14 H18 O4	250.12	10.1
508	Thiacloprid	C10 H9 Cl N4 S	252.02	5.1
508	Daidzein	C15 H10 O4	254.06	9.1
508	Propyzamide	C12 H11 Cl2 N O	255.02	9.1
508	Ifosfamide	C7 H15 Cl2 N2 O2 P	260.02	5.2
508	PharmaGSID_47261	C13 H15 N O3 S	265.08	5.9
508	2-Amino-N-cyclohexyl-N-methylbenzenesulfonamide	C13 H20 N2 O2 S	268.12	8.9
508	Acetochlor	C14 H20 Cl N O2	269.12	10.0
508	4,4'-(4-Methylpentane-2,2-diyl)diphenol	C18 H22 O2	270.16	8.1
508	5,6-Benzoflavone	C19 H12 O2	272.08	12.3
508	2-Ethylhexyl 4-(dimethylamino)benzoate	C17 H27 N O2	277.20	14.9
508	Sulfamethazine	C12 H14 N4 O2 S	278.08	4.1
508	Metolachlor	C15 H22 Cl N O2	283.13	10.1
508	PharmaGSID_47263	C15 H16 N4 O2	284.13	6.0
508	Oxazepam	C15 H11 Cl N2 O2	286.05	7.5
508	Chloroxuron	C15 H15 Cl N2 O2	290.08	9.6
508	Imazalil	C14 H14 Cl2 N2 O	296.05	8.8
508	Fenoxycarb	C17 H19 N O4	301.13	8.5
508	Tolnaftate	C19 H17 N O S	307.10	12.7

508	Tebuconazole	C16 H22 Cl N3 O	307.15	11.0
508	Warfarin	C19 H16 O4	308.10	8.6
508	N-Benzyl-9-(tetrahydro-2H-pyran-2-yl)adenine	C17 H19 N5 O	309.16	8.0
508	Isopropalin	C15 H23 N3 O4	309.17	14.5
508	Benzyl butyl phthalate	C19 H20 O4	312.14	12.3
508	dl-Norgestrel	C21 H28 O2	312.21	9.8
508	Isoxathion	C13 H16 N O4 P S	313.05	11.5
508	Triazophos	C12 H16 N3 O3 P S	313.07	9.6
508	Hexaconazole	C14 H17 Cl2 N3 O	313.07	11.3
508	Amoxapine	C17 H16 Cl N3 O	313.10	7.0
508	Diethylene glycol dibenzoate	C18 H18 O5	314.12	10.0
508	Diphenyl isophthalate	C20 H14 O4	318.09	6.2
508	Tomelukast	C16 H22 N4 O3	318.17	9.1
508	Flutolanil	C17 H16 F3 N O2	323.11	9.1
508	Acetohexamide	C15 H20 N2 O4 S	324.11	6.7
508	Diniconazole	C15 H17 Cl2 N3 O	325.07	11.8
508	Parinol	C18 H13 Cl2 N O	329.04	11.5
508	Triallyl trimellitate	C18 H18 O6	330.11	11.4
508	Halofenozide	C18 H19 Cl N2 O2	330.11	8.7
508	Dicyclohexyl phthalate	C20 H26 O4	330.18	14.2
508	17alpha-Hydroxyprogesterone	C21 H30 O3	330.22	9.6
508	Fluorescein	C20 H12 O5	332.07	6.8
508	Ipconazole	C18 H24 Cl N3 O	333.16	13.0
508	Thiophanate-methyl	C12 H14 N4 O4 S2	342.05	6.1
508	Prednisolone	C21 H28 O5	360.19	6.9
508	Fluroxypyr-meptyl	C15 H21 Cl2 F N2 O3	366.09	13.8
508	Picoxystrobin	C18 H16 F3 N O4	367.10	10.5
508	Volinanserin	C22 H28 F N O3	373.21	5.4
508	Methylprednisolone	C22 H30 O5	374.21	8.0
508	Fluorometholone	C22 H29 F O4	376.20	8.3
508	Bis(2-butoxyethyl) decanedioate	C22 H42 O6	402.30	14.7
508	Suxibuzone	C24 H26 N2 O6	438.18	9.3
508	Fluoxastrobin	C21 H16 Cl F N4 O5	458.08	9.9

Table 5.2.B. List of spiked chemicals for feature extraction validation samples (undetected; EPA ENTACT study) (Ulrich et al., 2019b)

Name	***Formula***	***Mass***	***RT***	***Sample #***	***Comment***
m-Cumenyl methylcarbamate	C11 H15 N O2	193.11	7.2	506	Too many chromatogram peaks
Trinexapac-ethyl	C13 H16 O5	252.09	7.4	508	Bad chromatogram peak shape

Chapter 6: Development and optimization of HRMS data processing workflow for contamination source identification and apportionment

6.1 Introduction

Surface water pollution is a global issue that presents a significant threat to both human health (e.g., waterborne diseases (Cissé, 2019; "Drinking-water," n.d.; Qamar et al., 2022) and ecosystems (Fent, 2015; Hiki et al., 2021; Scholz et al., 2011; Tian et al., 2021)). Accompanying the increasing environmental and human health risks is the challenge of identifying those few chemical sources and specific chemicals that are driving adverse effects. For example, surface waters are complex chemical mixtures and contain many uncharacterized anthropogenic chemicals and transformation products in addition to substantial numbers of naturally generated chemical constituents. For toxic chemicals, identifying and apportioning their contamination sources is crucial for developing effective water management and remediation strategies. Such information helps to identify and prioritize toxicants, enables risk assessment and determines responsible parties for pollution events (e.g., illegal waste discharge) (Mazza et al., 2015; Triassi et al., 2015). Given the hydraulic and chemical complexity of most environmental waters, there remains a pressing need for development of accurate source identification and apportionment methodologies.

Traditionally, source identification and apportionment have relied on highly accurate tracking and quantification of one or a few marker/surrogate/indicator chemicals that are assumed to be source specific and conserved through environmental and hydraulic processes (Buerge et al., 2003; Chae et al., 2021; Fauser et al., 1999; Klöckner et al., 2021; Rødland et al., 2022). Based on their relative concentrations, marker chemicals can be leveraged to identify the presence of

contamination sources (e.g., tire wear particles (TWP)) and extrapolate the source concentrations, but quantitative accuracies are prone to matrix effects. Also, marker chemicals must be both known and unique to individual sources for effective apportionment.

The introduction of computational data processing workflows with high-resolution mass spectrometry (HRMS) (Díaz et al., 2011; Masiá et al., 2014) provides a set of potential alternative strategies for this challenge. The richness of chemical information in HRMS data allows researchers to track and semi-quantify novel chemical contaminants without prior knowledge of their identity or presence (Aalizadeh et al., 2022; Du et al., 2020; Hug et al., 2014; Kiefer et al., 2021). Also, it enables non-target data processing where users can exploit larger quantities of HRMS features that provide capabilities for source identification or quantification. Especially, HRMS datasets allow the identification and extraction of "source fingerprint features" (FP features) (Barbosa et al., 2020; Campmajó et al., 2022; Lübeck et al., 2020) that are comprised of hundreds to thousands of chemical features that are representative of the contamination source (Peter et al., 2018). Conceptually, multivariate and machine learning based modeling could be used to leverage those FP features for complex mixture quantification. However, despite this potential, few research efforts have yet effectively developed and exploited this data processing strategy with HRMS data (Brack et al., 2019; Dávila-Santiago et al., 2022; Kiefer et al., 2021; Peter et al., 2019). The major challenge and barrier to such approaches is that without feature identification and use of commercial standards to determine peak area responses, relating observed feature signal intensity (e.g., peak height/area) to feature concentration remains uncertain.

We hypothesized that contamination sources in surface water samples can be isolated and quantified as unique entities using unidentified FP features generated from novel HRMS data processing workflows. Here, a workflow for source identification and apportionment was proposed

with a series of method development and optimization studies using tire leachate as a representative contamination source with lab-scale simulation experiments. Our objective was to evaluate the performance of different models and algorithms for source tracking studies, including assessing the accuracy, flexibility and generalizability of different models that are challenged by samples with different source concentrations and background chemical matrices. We anticipate the workflow will enable researchers to extract meaningful information more efficiently and reliably from HRMS datasets for source tracking studies.

6.2 Materials and Methods

6.2.1 Chemicals and reagents

Methanol, formic acid and water (all Optima LC/MS grade, ≥99%) were purchased from Fisher Scientific (Waltham, MA, USA) and used as received. A water purification system (Thermo Barnstead Nanopure Diamond UV; Dubuque, IA, USA) provided ultrapure deionized (DI) water. SPE cartridges (Oasis HLB, 200mg) were purchased from Waters (Milford, MA) and used as received. The information for internal standard (ISTD) chemicals for instrument performance evaluation is summarized in **Table 6.1**.

Table 6.1. ISTD Chemicals for instrument performance evaluation.

Compound Name	RT	ESI	CAS	Formula	Vendor
2,4-Di-tert-butylphenol-d19	12.7	-	96-76-4	C14H3D19O	**SCBT**
5-methyl-1H-benzotriazole-d6	5.05	+/-	1246820-65-4	C7HD6N3	**Toronto Research Chemicals**
Atrazine-d5	7.14	+/-	163165-75-1	C8H9D5ClN5	**Sigma Aldrich**
Bis(2-ethylhexyl)phthalate-d4	16.9	+	93951-87-2	C24H34D4O4	**CDN isotopes**
Caffeine-13C3	4.5	+	78072-66-9	C5[13C]3H10N4O2	**Sigma Aldrich**
Carbamazapine-d10	6.53	+	132183-78-9	C15H2D10N2O	**Sigma Aldrich**
Cotinine-d3	3.46	+	110952-70-0	C10H9D3N2O	**Sigma Aldrich**

DEET-d7	7.27	+/-	1219799-37-7	C12H10D7NO	**CDN isotopes**
Docosahexaenoic acid-d5	15	-	25167-62-8	C22H27D5O2	**Cayman**
Ethylparaben-d4	5.82	+/-	1219795-53-5	C9H6D4O3	**Toronto Research Chemicals**
Ibuprofen-d3	9.08	-	121662-14-4	C13H15D3O2	**Sigma Aldrich**
Linolenic Acid-d4	13.3	+	463-40-1	C18H26D4O2	**Cayman**
Lithocholic Acid-d4	15	+	434-13-9	C24H36D4O3	**Cayman**
Metolachlor-d6	10	+	1219803-97-0	C15H16D6ClNO2	**Toronto Research Chemicals**
Nicotine-d3	1.99	+	69980-24-1	C10H11D3N2	**Sigma-Aldrich**
Prometon-d3	8.05	+	1219803-43-6	C10H16D3N5O	**CDN isotopes**
Propylparaben-d4	6.99	+/-	1219802-67-1	C10H8D4O3	**Toronto Research Chemicals**
Sulfadimethoxine-d6	4.99	+/-	73068-02-7	C12H8D6N4O4S	**Sigma Aldrich**
Sulfamethoxazole-d4	4.27	+/-	1020719-86-1	C10H7D4N3O3S	**Toronto Research Chemicals**
Theobromine-d6	3.1	-	117490-40-1	C7H2D6N4O2	**Sigma Aldrich**
Vanillin-d3	4.23	+/-	74495-74-2	C8H5D3O3	**Sigma-Aldrich**

6.2.2 Tire Leachate Preparation and sampling site location

The preparation of TWP (focused on new and used passenger car and light truck tires) for leaching studies was described elsewhere (Tian et al., 2021). Tire tread wear particles (TWP) were made by physical abrasion of the treads of nine tires (brand, vehicle type, and use history in **Table 2.3**) using an angle grinder with a steel carbide disk (Parkes MCM90; Miller Tire, Wauseon, OH, U.S.A.) and mixed in equal weights. The TWPs were stored in a zip-lock bag and isolated from ambient air prior to experiments. To generate the TWP leachate (TL) stock, TWP was dispersed with glass beads (3mm OD, pre-cleaned with methanol) in 1 L stainless steel columns and leached by RO (Reverse Osmosis) water (24 h) from a recirculating stainless-steel reservoir through the columns via upward flow with a peristaltic pump (130 rpm; Masterflex L/S Tygon® tubing, 3/8" ID). TWP leachate stock was always generated at 1 g TWP/L water (35 g TWP/ 35 L RO water in

the reservoir). After being collected, the TL was stored in 4 L amber glass jars at 4 °C until extraction within 24 hrs. Leachates were split into 35 aliquots (1 L each) and each aliquot was spiked with QTOF ISTD mixture (50 μL) and extracted using preconditioned (10 mL methanol, 25 mL water) SPE cartridges (Oasis HLB) at 5-10 mL/min. SPE cartridges were then rinsed with 10 mL DI water, air-dried (30 min), then eluted with methanol (2 × 5 mL). The eluents were combined (350 mL) and stored under -20℃ for later dilution or spiking experiments.

Field samples were collected from different locations to represent different background chemical matrices. The sites located in rural areas that are away from major highways or heavy traffic were selected to avoid the presence of interfering TWP contamination. Water from eight different sites and an ultrapure water control (see **Table 6.2**) were used to prepare the TL source dilution series. The relative contamination level was estimated for reference based on the surrounding environment, traffic, and sample extract color (the samples with heavier matrices were a darker color) by visual inspection. The samples (6-8 L per site) were collected with a polypropylene plastic scoop, stored and transported in 4 L pre-cleaned amber glass jars, and stored at 4℃ until spiking and extraction (within 24 h).

Table 6.2. Sampling site information for field samples.

Sample Site	Abbrev.	Est. matrix level	Sampling date	Coordinate
Kobayashi Park	KP	Low	12/6/2022	47°11'53.0"N 122°31'48.5"W
LCMS water	MQ	Low	1/4/2023	NA
Brunett Rd river	BR	Low	1/4/2023	47°08'13.9"N 122°03'45.5"W
Lake Louise	LL	Mid	12/6/2022	47°09'43.1"N 122°34'03.2"W
Nisqually River	NR	Mid	12/6/2022	47°03'28.6"N 122°41'29.3"W
Lake Tapps	LT	Mid	1/4/2023	47°11'30.8"N 122°09'53.7"W
Point Defiance	PD	High	12/6/2022	47°18'10.8"N 122°30'60.0"W

Local Pond	HG	High	12/6/2022	46°53'54.1"N 122°38'05.9"W
Tacoma downtown park	DTP	High	1/4/2023	47°15'41.0"N 122°26'59.8"W

6.2.3 Sample processing

The TL dilution series were prepared at 7 different levels, covering a range of 1-100% source concentration (0.01, 0.02, 0.05, 0.1, 0.2, 0.5, 1 g/L achieved by spiking different volumes of TL methanolic stock into field samples; **Figure S6.1**). Additionally, a methanolic dilution series of the TL stock that represents the same dilution levels was prepared to represent a 'pure source' dilution series for comparison. All spike concentrations were prepared in duplicate, with each field sample split into 200 mL aliquots for spiking (n=8*2 for each site, including duplicate unspiked samples). The TL stock was spiked into field samples at different volumes, with pure methanol added to ensure an equal volume of spiked organic solvent in each sample (2 mL total; **Table S6.1**).

To simulate actual sampling and perform model evaluation, 12 'blinded spikes' were prepared by another person (labeled as 'blind') as a validation dataset independent of the model training process. To make the blind samples, arbitrary volumes of TL stock solution were spiked into the samples (compensated to a total spiking volume of 2 mL with methanol; final concentration within dilution series concentration range). Due to the lack of prior knowledge of the data distribution, the blinded spike samples only covered TL concentrations above 10%. Therefore, resampling was done (i.e., randomly moving samples with TL concentrations between 1-10% from the spike sample dataset into the blind sample dataset) to compensate for the tight concentration distribution of the blind samples.

Methanol dilution samples were analyzed by LC-HRMS after dilution, while all spiked samples were extracted through SPE cartridges (**Section 6.2.1)**, and then analyzed. Most samples

were injected in triplicate as analytical triplicates (n=495 for known spiked samples injections; n=54 for blinded spiked samples injections; after resampling, n=478 for known spiked samples injections; n=71 for blinded spiked samples injections). This number of replications would allow for subsequent statistical analysis.

6.2.4 LC-QTOF analysis and data analysis

All samples were analyzed using an Agilent 1290 Infinity II ultrahigh performance liquid chromatograph (UHPLC) coupled to an Agilent 6546 quadrupole time-of-flight high-resolution mass spectrometer (QTOF-HRMS; Santa Clara, CA, USA). A reverse-phase C18 column (Agilent ZORBAX Eclipse Plus 2.1×100 mm, 1.8 µm) with a C18 guard column (2.1×5 mm, 1.8 µm) was used for the UHPLC separation at 45 ºC with 5 µL injection volume. Separation employed a gradient elution with mobile phases of 0.1% formic acid in each of deionized water (A) and methanol (B) as follows: 5% B at 0-1 min, 50% B at 4 min, 100% B at 17-20 min, 5% B at 20.1 min; stop time 22.5 min; post-time 2 min. The flow rate was 0.4 mL/min. Full scan data were acquired under 10 GHz Extended Dynamic Range mode at a range of 100-1700 m/z; Quality assurance and quality control (QA/QC) included mass calibration before each analytical batch and mass accuracy correction by continuous infusion of purine and HP-921 calibrants. Solvent blanks and internal standards (details available in previous studies (Du et al., 2017; Peter et al., 2018; Tian et al., 2020) and **Table 6.1**) were each injected every 12 samples (~4.5 h) to assess carry-over (none observed) and variation in instrumental response throughout analysis batches (one analytical batch per experimental system; <30% relative standard deviation of internal standard peak areas).

After data acquisition, raw data files (Agilent .d format) were converted to .abf format (Reifycs Abf Converter); MS-DIAL (version 3.46) (Tsugawa et al., 2015) was used for primary data processing (i.e., non-target feature extraction and alignment). Parameter settings were as

follows: *m/z* tolerance for feature extraction 0.005; minimum peak height 2000; minimum peak width 5 scans; *m/z* slice 0.01 Da; alignment *m/z* tolerance 0.015 Da; alignment retention time tolerance 0.1 minutes. Initial data reduction used Python programming language (version 3.9.12) with Pandas (version 1.5.1) to isolate features with maximum peak area (across all samples) > 50000, *m/z* between 100 and 900 Da, and retention time between 2-18 min; and present at peak area 5-fold greater than peak area in any of the solvent blanks or the ISTD samples. Features satisfying all these criteria (n = 14304) were retained.

For source tracking model training and optimization in current study, TL was the representative pollutant source. Therefore, samples from the methanolic dilution series of TL stock solution ('pure source dilution') were used to prioritize fingerprint features for further analysis, assuming these samples represent the minimal possible background matrix effect that could impact feature abundances. Based on this assumption, FP features that are exclusively derived from the TL stock solution should exhibit similar decreasing trends (linear or non-linear) as the TL stock solution is diluted. These features were prioritized as potential TL fingerprint features by the Density-Based Spatial Clustering of Applications with Noise (DBSCAN; **Figure 6.1**) algorithm (Ester et al., n.d.) using Python programming language with scikit-learn package (version 1.1.3).

Z-score normalization of data was conducted to eliminate the data skewness and kurtosis before the analysis as below:

$$z = (x - \mu) / \sigma$$

where z is z-score, x is the peak area of the feature, μ is the average peak area of the feature across all the samples, and σ is the standard deviation of the peak areas. Algorithm parameters, including *minPts* and *eps*, were tuned iteratively by manually checking the cluster quality to optimize the

clustering outcome. A typical clustering heatmap of features with increasing relative abundance is shown in **Figure 6.1**. The determined parameters (*minPts* = 10, *eps* = 2) were used to produce final clustering results (2695 features; from 2 clusters). The features were further analyzed for model development and optimization.

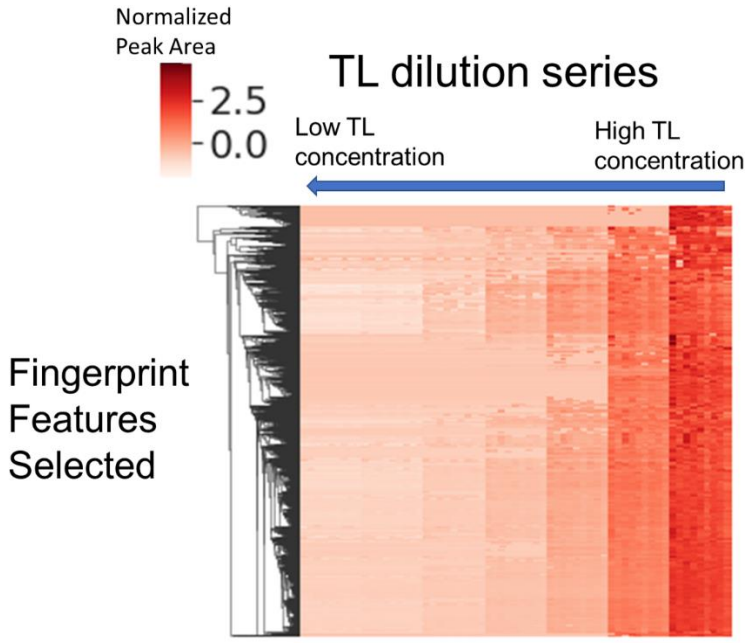


Figure 6.1. Example clustering heatmap from the DBSCAN analysis. Each column represents an analytical result (n=63), and each row represents a unique chemical feature (total features n=2695) that is a potential FP feature. The values presented in the plot are z-score normalized peak area values.

After FP feature prioritization, similarity scores were calculated between each spiked field sample (labeled 'spike') and L7 (1g/L) TL methanol samples (labeled 'source') using peak abundance of the FP features. Three scores were calculated for comparison, including dot product (Tsugawa et al., 2015), jaccard scores ("Jaccard index," 2023) and direct comparison with numpy (version 1.23.5) package. Additionally, different regressions and modeling approaches were used to develop the source tracking workflow, including regularized regression, XGBoost regression and classification, dimension reduced modeling (linear regression, SVM regression and random

forest regression). The model development and optimization were based on scikit-learn (version 1.1.3) and xgboost (version 1.7.3) packages.

6.3 Results and discussion

Using prioritized FP features, various methods were tested for qualitatively identifying the presence of a known source in a complex sample and quantifying the amount or concentration of that source in the sample. Firstly, the similarity scores (between spiked samples and highest concentration TL source sample) were first examined to evaluate possible trends with respect to source presence (i.e., qualitative indication of source presence) and different source concentrations (i.e., quantitative source apportionment). Second, qualitative methods to indicate TL source presence and background chemical matrices, respectively, were developed based on logistic regression and the XGBoost classification model. Finally, different quantification models were developed to predict TL source concentrations, using the spiked samples as training and testing data. The developed models were validated with blind samples.

6.3.1 Similarity scores

Four scores were calculated with FP feature abundance between spiked field samples (*Spike*) and triplicate averaged L7 (1g/L) TL methanol samples (*Source*). Direct comparison scores (Simp) were calculated as follows:

$$Simp = Mean(\frac{PA_{Spike}}{PA_{Source}})$$

where *PA* represents peak abundance of the FP features.

Two Jaccard scores were calculated as follows:

$$J(Spike, Source) = \frac{|Spike \cap Source|}{|Spike \cup Source|}$$

Jaccard scores were calculated as both a weighted average (Jaccard raw) using log(peak area) in spike and source samples and a binary average (Jaccard bin) using 0 and 1 binary values to indicate presence or absence of the peak, respectively, in spike and source samples.

Dot product scores were calculated as follows (Tsugawa et al., 2015):

$$Dot\ Score = \frac{(U_{Spike} \cdot U_{Source})}{\left\|U_{Spike} \cdot U_{Source}\right\|} / \frac{(I_{Spike} \cdot I_{Source})}{\left\|I_{Spike} \cdot I_{Source}\right\|}$$

where *U* represents the peak area of features belonging to the union FP features (present in either spike and source samples) and *I* represent the intensity of features belonging to the intersection of FP features (present in both spike and source samples).

By examining the calculated scores relative to known source concentrations in spiked samples (**Figure 6.2**), the dot product effectively discriminated between the samples containing TL at concentrations >0 g/L vs. samples spiked only with pure methanol, regardless of the background matrix, suggesting value for indicating presence-absence. The other three scoring approaches exhibited apparent relationships (linear or non-linear) relationships with the source concentration, albiet with large variations that are potentially driven by different background matrices.

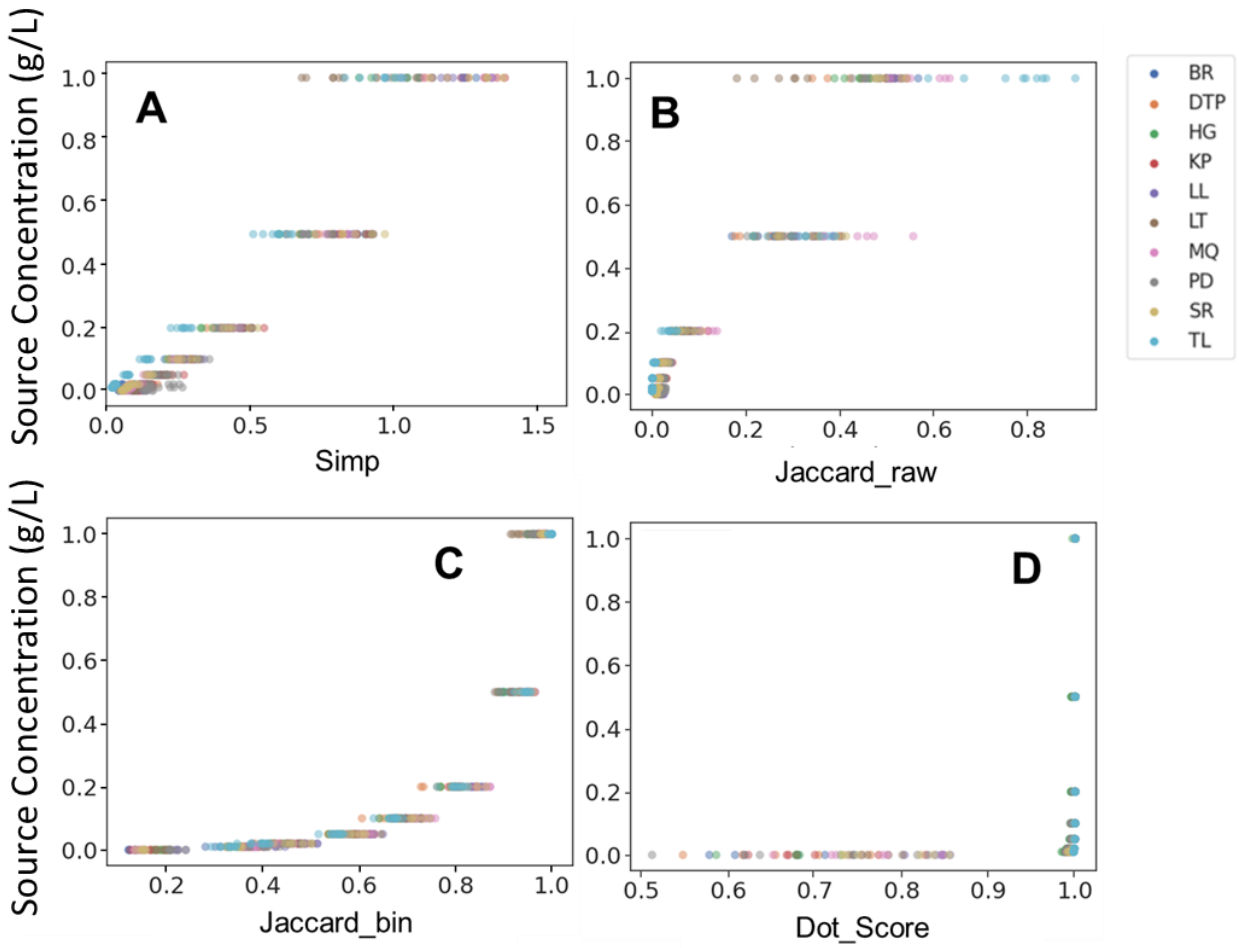


Figure 6.2. Calculated similarity scores between spiked samples and source samples vs. spiked TL concentrations. (A) direct comparison score; (B) weighted average Jaccard score; (C) binary average Jaccard score and; (D) dot score. Each dot represents an individual sample, the legend represents the sample location.

6.3.2 Qualitative analysis for source identification

As shown in **Figure 6.2D**, dot product score can distinguish the presence and absence of the TL source. The implementation of such discriminative algorithms and statistical analysis to distinguish presence/absence of sources would provide valuable insights for source tracking method development, especially supporting selection of control samples without exhaustive data processing. In addition to source identification, background matrix identification is also valuable,

providing critical information to assist model optimization. Therefore, two models were developed for qualitative analysis, including source presence evaluation based on dot scores and background matrix identification with XGBoost classification model.

Logistic regression was employed to analyze the spike sample dot scores. The L2 penalty was applied to the model (regularization strength = 70.0), with liblinear solver. The fitted regression model was:

$$P = \frac{e^{-27.84+32.40x}}{1+e^{-27.84+32.40x}}$$

where P is the probability of predicting the presence of source (i.e., prediction of 1; while prediction of 0 indicates source absence), x is the dot score calculated from the FP features. After models were trained, blind samples were used to test model performance, with confusion matrices for both spike and blind sample datasets used to examine performance results. Under the current conditions, the model accurately predicted the presence of the source.(**Figure 6.3**).

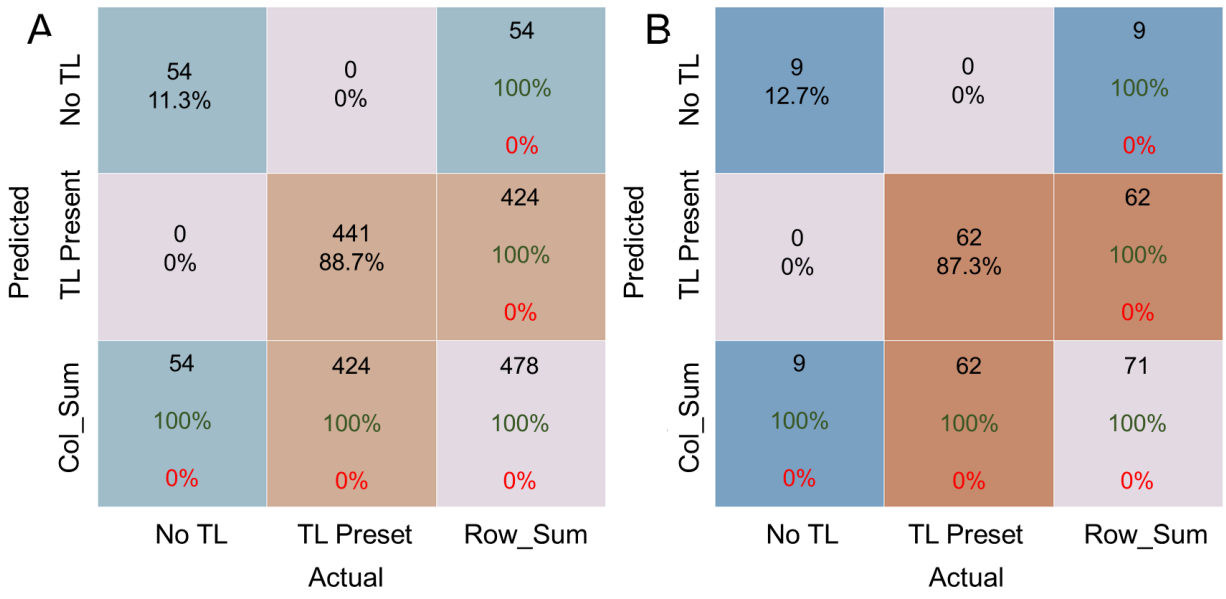
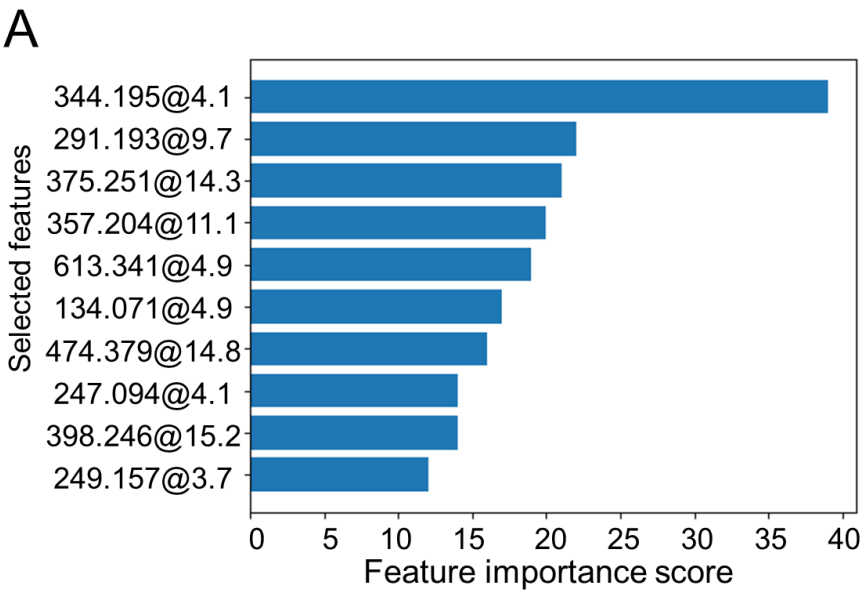
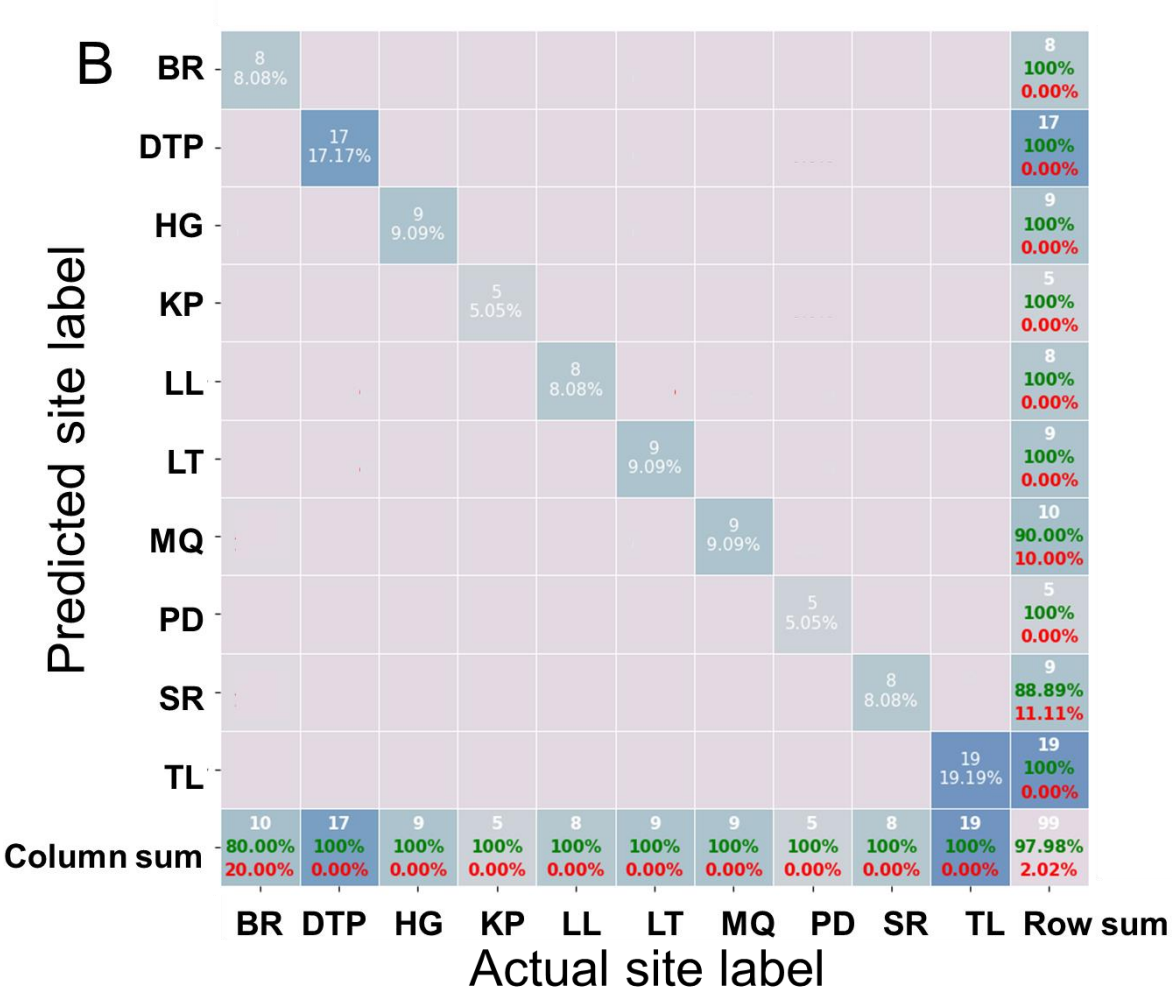


Figure 6.3. Confusion matrix of classification results from on dot score logistic regression for (A) spike samples used to train and optimize the model and (B) blind samples that served as a model validation dataset. The black number annotations represent the number of samples falls into the actual-predicted cell. The black percentage annotations represent the percentage of the samples from all the samples predicted. The green and red percentage annotations represent correct and wrong prediction portions, respectively.

Classification models based on XGBoost were also trained to identify the background matrices from different sampling sites. The FP features were standardized (see **section 6.2.4**) prior to modeling. Spike sample dataset was split into 80% training and 20% testing subsets for model development. After training, the top 10 features (determined by feature importance gain) were selected and used to retrain the model. After model training, the blind samples were used to evaluate model performance. The top 10 features selected for training and associated confusion matrices for both spike and blind sample datasets are shown in **Figure 6.4**.




B
Predicted site label
BR
DTP
HG
KP
LL
LT
MQ
PD
SR
TL
Column sum
8 8.08%
8 100% 0.00%
17 17.17%
17 100% 0.00%
9 9.09%
9 100% 0.00%
5 5.05%
5 100% 0.00%
8 8.08%
8 100% 0.00%
9 9.09%
9 100% 0.00%
9 9.09%
10 90.00% 10.00%
5 5.05%
5 100% 0.00%
8 8.08%
9 88.89% 11.11%
19 19.19%
19 100% 0.00%
10 80.00% 20.00%
17 100% 0.00%
9 100% 0.00%
5 100% 0.00%
8 100% 0.00%
9 100% 0.00%
9 100% 0.00%
5 100% 0.00%
8 100% 0.00%
19 100% 0.00%
99 97.98% 2.02%
BR DTP HG KP LL LT MQ PD SR TL Row sum
Actual site label

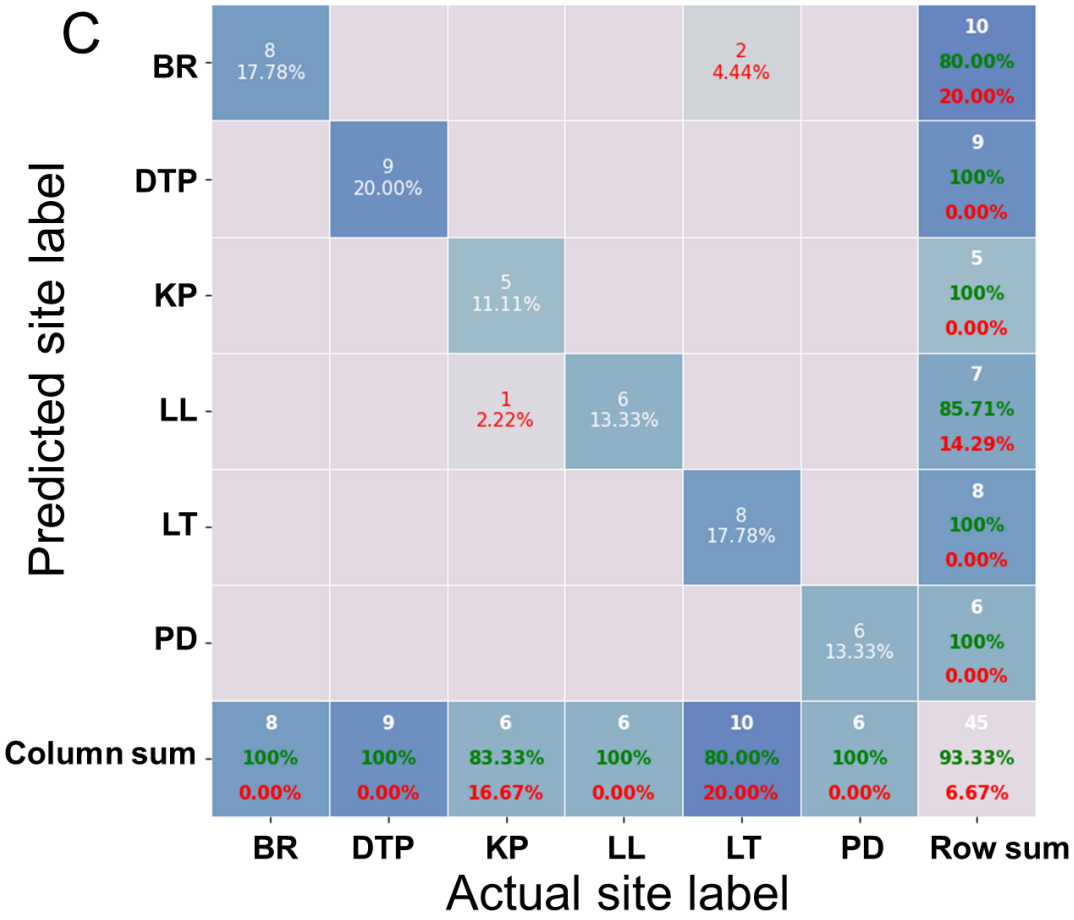


Figure 6.4. Top 10 important features and confusion matrix from XGBoost classification model. (A) top 10 important features, prioritized with average gain of splits that uses the feature; (B) confusion matrix of spike sample testing dataset and; (C) confusion matrix of blind sample dataset. The bottom right cell (Col_Sum, Row_Sum) represents the model prediction accuracy. The colors of the cells in panels B and C represent the numbers of predictions made, darker (blue) color represents more predictions, lighter (pink) color represents for less predictions.

The classification model provided decent accuracy during testing with the current datasets (97 and 76% for spike and blind datasets, respectively) with respect to source presence prediction, but due to the practical limitation of sampling efforts, further model optimization and evaluation based on larger dataset are necessary. In addition, the current dataset is imbalanced (e.g., more source spiked samples vs. blank samples) and small (<100 data points), so further data collection and processing focused on balancing the samples from different matrices by collecting more

sample. Nevertheless, current results suggested dot score logistic regression and XGBoost classification models are powerful tools for TL source (at 0-1g/L range of source concentrations) and background matrix identification. These algorithms improve the data processing efficiency and source apportionment accuracy by reducing the noise and background detections and labeling the background matrix.

6.3.3 Quantitative analysis for source apportionment

In the current study, ~550 data files (representing analytical detections from samples and replicates) were collected for source tracking method development using ~3000 FP features. However, when working with machine learning algorithms, feature numbers that far exceed the number of observations (e.g., samples) can lead to overfitting. The model may become too complex and prioritize noise in the data rather than any underlying patterns. Consequently, if FP features were used for model development without pre-processing, the model may perform well on training data but poorly on new, unseen data. Therefore, in this study, several strategies were leveraged and compared for their performance to avoid overfitting and improve the model's performance and generalizability. These included: a) feature selection, only selecting the most important features that determine model performance for model development; b) feature extraction, by translating the features into descriptive hyper-features (e.g., similarity scores or principal components); and c) use of a penalty term within the loss function (which used to optimize the model accuracy) during model training.

The spike sample dataset was randomly split into 80% training and 20% testing and used in all the subsequent model training processes. The testing subset was used to optimize the model hyperparameters and evaluate model performance. After the model was developed, the blind sample dataset was used for model validation and to evaluate model accuracy for TL source

concentration predictions. The model accuracy was evaluated with R-squared (R^2) score, MSE, and RMSE, calculated as follows:

$$R^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2}$$

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2$$

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}$$

where y_i is the actual value of source concentration, $\hat{y}_i$ is the predicted source concentration, and $\bar{y}$ is the mean value for actual source concentrations. Note that all the similarity score evaluation matrices were calculated with all spike samples. Other model evaluation matrices were calculated with spike sample testing subset only. The source apportionment performance of different models for the testing subset of spike sample dataset was summarized in **Table 6.3** and **Figure 6.6**.

XGBoost regression model was trained. Proper hyperparameters (*n_estimators*=50, *max_depth*=6) were determined by grid search approach with 5-fold cross validation. The model automatically selected FP features with the highest importance (n = 10) for modeling.

To leverage similarity scores for source quantification, different regression models (see **Figure 6.5**) were used to fit the calculated similarity scores (Jaccard raw, Jaccard bin, and simp). Polynomial regression (degree=2) was used for Jaccard raw scoring as follows:

$$y = -1.5x^2 + 2.5x + 0.0037$$

Logistic regression was used for Jaccard bin scoring:

$$y = \frac{576.08}{1 + e^{-9.20\times(x-1.68)} + 0.015}$$

For simp score, linear regression was leveraged to fit the data points:

$$y = 0.882x - 0.094$$

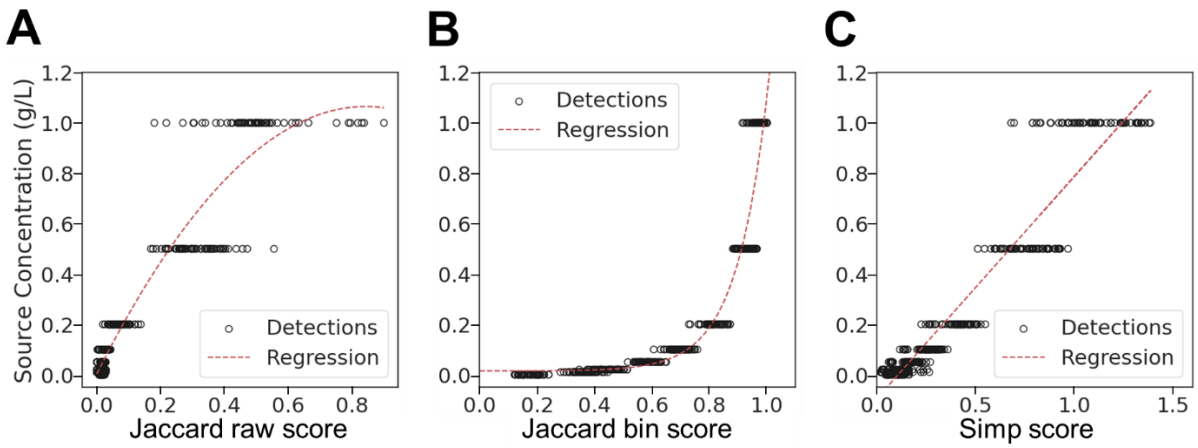


Figure 6.5. Regression models for similarity score predictions. (A) polynomial regression (degree=2) that fits jaccard raw score; (B) logistic regression that fits jaccard bin score and; (C) linear regression that fits simp score. Black markers represent individual data points, the red dashed line represents fitted regression curves.

In addition to similarity score that abstracts FP features into a single variable, principal component analysis (PCA) was also deployed as another feature extraction algorithm on the dataset for feature dimension reduction. Similar to XGBoost regression, all the FP features were standardized before PCA (**Section 6.2.4**). The FP features were reduced to 25 principal components that explained >95% of total variance. The pulsar dataset explained variance was shown in **Figure S6.1**. Three different regression models were trained for TL source concentration prediction using the 25 principal components, including linear regression, random forest regression (*max_depth*=5, *n_estimators*=500) and SVM regression (*C*=1, *epsilon*=0.01, *kernel*=linear). The optimal hyperparameters were determined by grid search and cross validation approaches.

Regularized regression models were also trained using all FP features for TL source concentration prediction. All 3 regularization models, Lasso regression (*alpha*=0.01, *tol*=0.01), Ridge regression (*alpha*=0.01) and Elastic Net regression (*alpha*=0.01, *L1_ratio*=0.5, *tol*=0.01, were trained and tested. The optimal hyperparameters were determined by a cross validation approach. After all the models were trained and optimized, the blind sample dataset was used to evaluate model performance; results are summarized in **Table 6.4** and **Figure 6.7**.

Table 6.3. Summary of model performance for TL source concentration predictions on spike sample datasets.

	Feature selection	**Feature extraction – similarity score***			**Feature extraction – PCA dimension reduction**			**Regression with regularization**		
Regression Model	XGBoost	Jaccard raw – polynomial	Jaccard bin – logistic	Simp - linear	Linear	Random forest	SVM	Lasso	Ridge	Elastic net
R^2 score	0.99	0.91	0.92	0.91	0.97	0.94	0.96	0.94	0.98	0.94
MSE	0.001	0.0098	0.0087	0.01	0.003	0.0072	0.0041	0.0084	0.0032	0.0076
RMSE	0.032	0.099	0.093	0.01	0.055	0.085	0.064	0.092	0.057	0.087

* Note that all the similarity score evaluation matrices were calculated with all spike samples. Other model evaluation matrices were calculated with spike sample testing subset only.

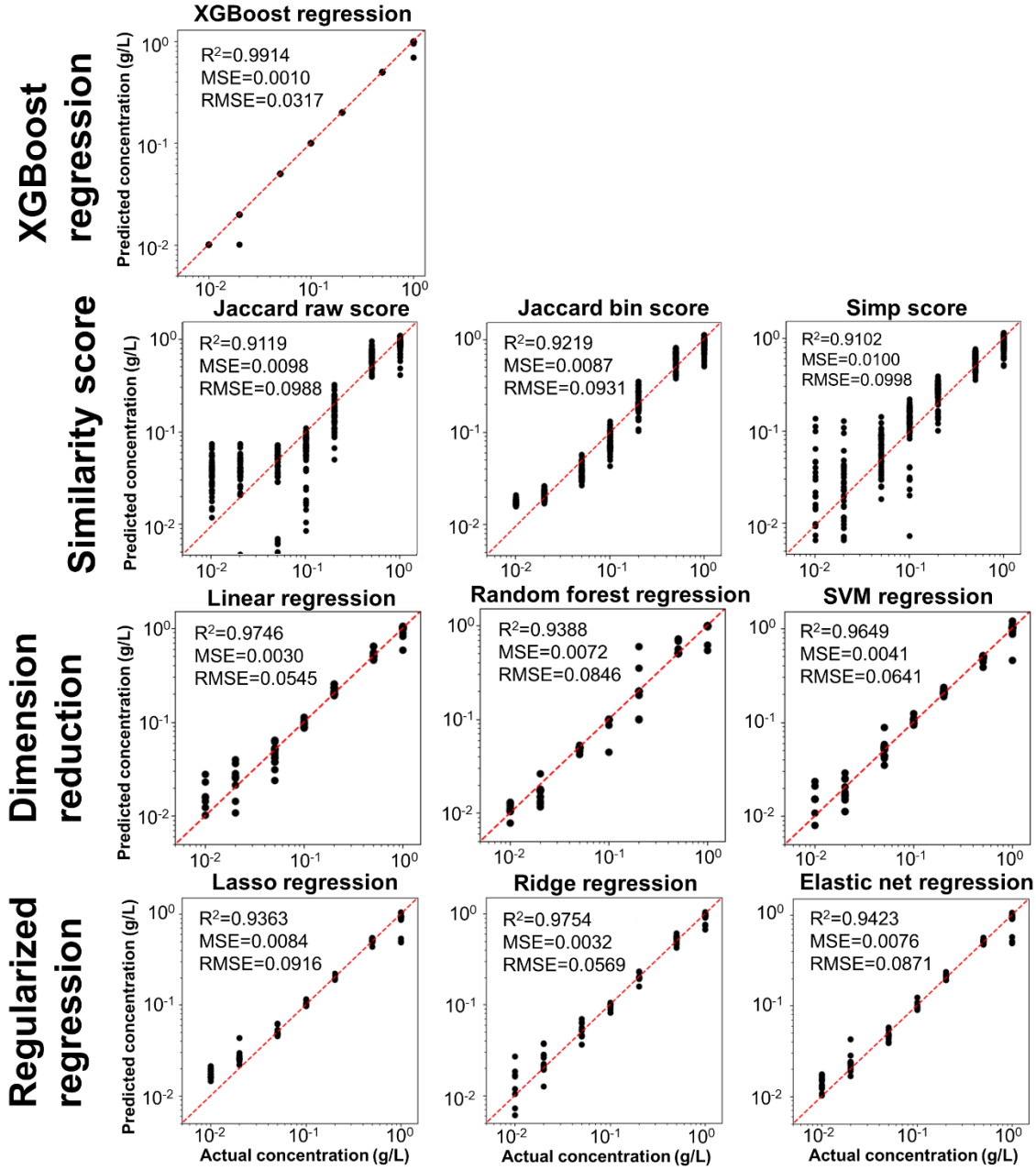


Figure 6.6. Summary of model performance for TL source concentration predictions on spiked sample datasets. All the data were plotted in log-scale for visualization. Each data point represents individual instrumental injections. The red dashed line represents the ideal model performance (y_pred = y_actual).

Table 6.4. Summary of model performance for TL source concentration predictions with blind sample dataset.

	Feature selection	**Feature extraction - similarity score**			**Feature extraction – PCA dimension reduction**			**Regression with regularization**		
Regression Model	XGBoost	Jaccard raw - polynomial	Jaccard bin - logistic	Simp - linear	Linear	Random forest	SVM	Lasso	Ridge	Elastic net
R^2 score	0.79	0.81	0.78	0.77	0.85	0.78	0.87	0.89	0.88	0.89
MSE	0.028	0.026	0.03	0.031	0.021	0.03	0.017	0.015	0.016	0.015
RMSE	0.17	0.16	0.17	0.18	0.14	0.17	0.13	0.12	0.13	0.12

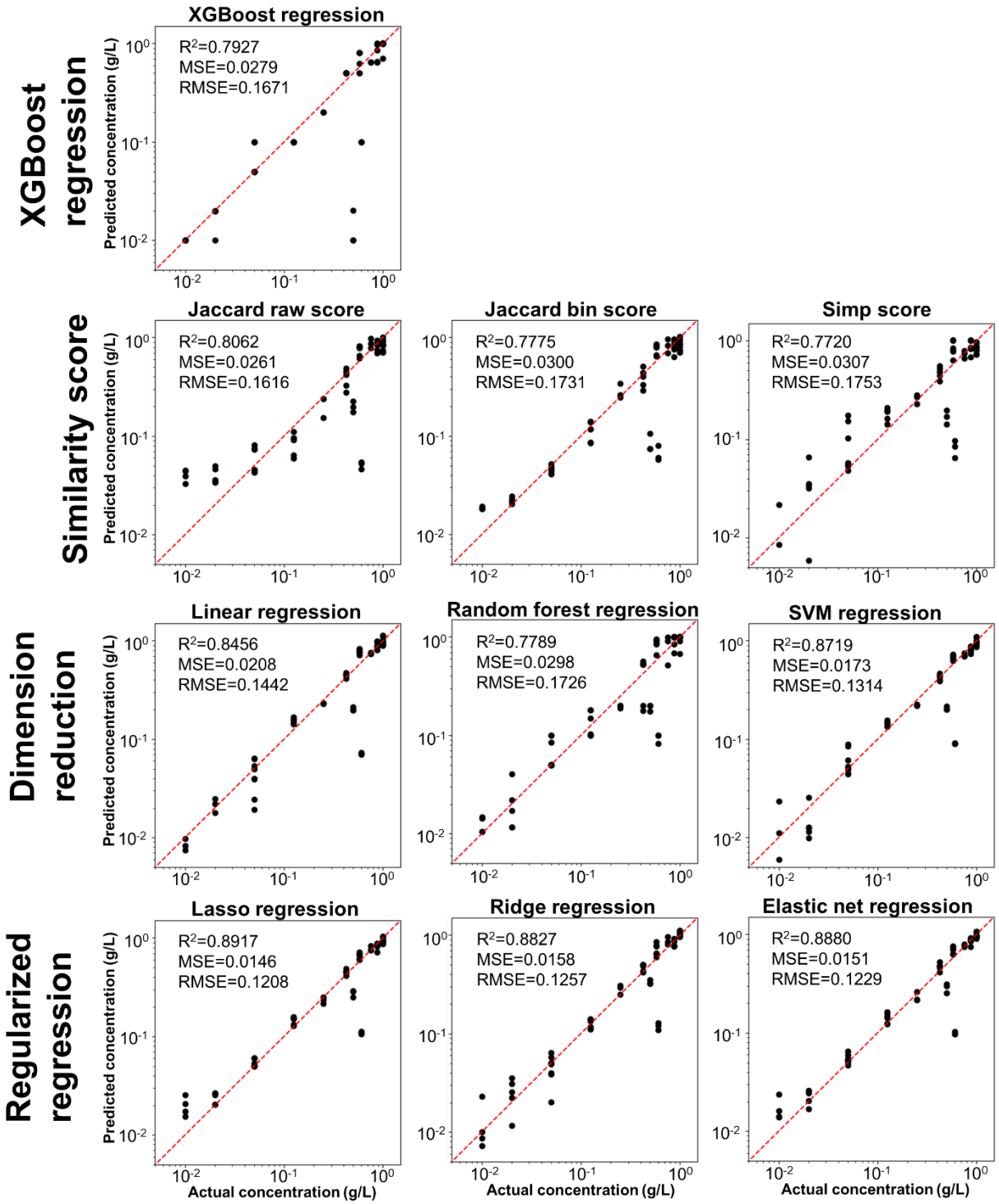


Figure 6.7. Summary of model performance for TL source concentration predictions on blind sample datasets. All the data were plotted in log-scale for visualization. Each data point represents individual instrumental injections. The red dashed line represents the ideal model performance (y_pred = y_actual).

For the spiked sample dataset (**Figure 6.6**), all models yielded an accurate (R^2>0.85) prediction for TL source concentrations. Dimension reduction regression and regularization regression groups both achieved the highest accuracies (lowest MSE or highest R^2 score). However, when the same models were applied to the blind sample dataset (**Figure 6.7**), model performance noticeably declined. This discrepancy highlights limitations to models' generalizability and the potential for model overfitting. Across all the models, XGBoost regression model was more susceptible to overfitting compared to other simpler models (Chu et al., 2022). Overfitting might be mitigated by tuning hyperparameters like *max_depth* or reducing the learning rate to avoid such problems with the support of larger datasets. The regression models based on scores also struggled with the blind dataset, potentially due to the models' simplicity (as single-variable models, only the similarity score was generated for model development and prediction). Combining multiple similarity scores might increase the model's ability to incorporate dataset complexity and variability. The dimension reduction strategy performed well while the SVM regression model outperformed the other two models (random forest and linear regression) with its flexibility. Among all the models tested, from both spike and blind sample datasets, the regularization regressions (i.e., lasso, ridge, and elastic net regression) generated the most accurate predictions. Compared to the other models, these regularized regression models have relatively simple assumptions and may generalize well to new/unknown data with a limited amount of training data. This generalizability derives from the use of the L1/L2 penalty that promotes sparsity or constraint non-zero coefficients, respectively, strikes the balance between model performance with training data and maintaining simplicity, and reduces the risk of overfitting.

Thus, the regularization regression model ability to perform feature selection and readily generalize, along with its tunability (via regularization parameter tuning) makes it a powerful

technique for a source tracking workflow, particularly when dealing with datasets with more features (i.e., FP features) than observations (e.g., data points, samples), as is common for HRMS datasets. Indeed, there exist fundamental mismatches between HRMS data collection capacity (e.g., very high numbers of features) and the practical limitations on the number of samples that can be collected and analyzed. In these situations, inherently prone to overfitting concerns, more training and testing data would be required to mitigate these risks and validate the model performance for application to field samples. Possible alternative models for development efforts include the optimized XGBoost, dimension reduction regression (PCA coupled with SVM regression), and the multiple score models suggested above.

6.3.4 Model performances with field samples

In addition to the lab simulated source dilution samples, the trained models were deployed on archived field samples (stored as methanolic extracts at -20℃) as illustrative examples of possible source identification and apportionment studies. Sample collection and preparation were described elsewhere (Peter et al., 2019). In short, hourly samples were collected at Miller Creek, Seattle, Washington with ISCO portable samplers (Teledyne ISCO, Lincoln, NE, USA) during storm events and processed to screen for vehicle derived chemicals (e.g., 6PPDQ). The samples are therefore representative of the potential presence of tire leachate in a field sample set. All the samples were re-analyzed by LC-HRMS as described in **Section 6.2.4**; the data was then processed for model predictions with models trained and described above. Because no actual source concentration (i.e., tire leachate quantification) data was available, the model performance was qualitatively and quantitatively evaluated by comparing the predicted source concentrations with vehicle-derived chemical concentrations or TL concentrations extrapolated from those chemical concentrations (namely 6PPDQ; **Figure 6.8**). The comparison showed a reasonable agreement

between model predictions and detected chemical concentrations; it clearly delineated low flow (lower chemicals release) and peak flow (higher chemicals release) sample types.

The qualitative model (dot product score coupled with logistic regression) was tested with the data, but did not yield any positive (i.e., presence of the source) predictions. This might be because the actual source concentration was lower than the lowest training concentration or the interference of a co-source of TL within the data, indicating that future development efforts should improve sample collection and coverage of trace source concentrations (<1%) where more training data was especially needed.

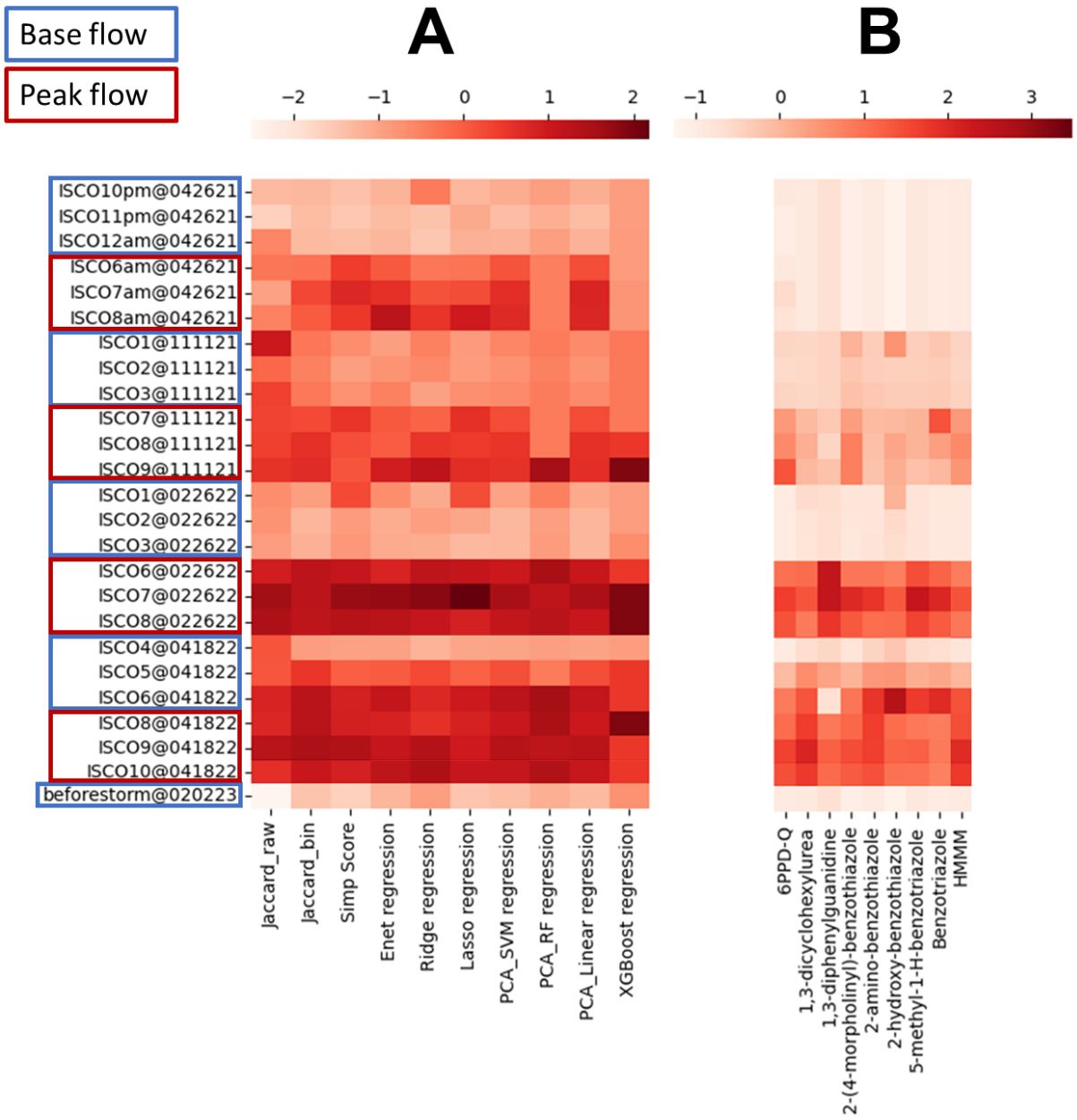


Figure 6.8. Heatmap for model predicted source concentrations and chemical concentrations detected in the samples. (A) heatmap for model predicted source concentrations and; (B) heatmap for chemical concentrations detected. All the values are column-wise standardized, the color bars represent standardized values for each cell.

The quantitative models were also tested with the dataset. To evaluate the quantification accuracy of different models, TL source concentrations were estimated by extrapolating 6PPDQ

concentrations (**Table 6.5**; 5.15~15.6 μg 6PPDQ/g TWP; Hu et al., 2022; Hu et al., 2023). Note that all the predicted and 6PPDQ-extrapolated source concentrations were back-calculated to the field concentration (i.e., the original concentration before sample extraction). Among all the predictions, Jaccard_bin and Jaccard_raw failed to predict the source concentration or only predicted low concentrations (Jaccard_bin and Jaccard raw based predictions predicted 0.08~0.09 mg/L and 0.2~0.6 mg/L for all samples, respectively), while XGBoost failed to make continuous predictions (all the predicted values were from the training data concentrations). Similar to the results derived from lab-simulated samples, the other models, especially regularized regressions and PCA coupled with SVM or RF regression, yielded the closest predictions compared to the extrapolated concentrations. Due to factors such as the potential inaccuracy of 6PPDQ extrapolation of TL concentration (e.g., actual TWP concentration is different than the 6PPDQ-predicted values) and the presence of confounding sources, we note that the current evaluation assists only in excluding models that are ineffective for source predictions and prioritizing potential model candidates for further development. Future efforts will require extensive sample collection and model optimization efforts.

Table 6.5. Predicted TL concentrations versus 6PPDQ extrapolated concentrations. All the predicted concentrations were back calculated to the field concentration (i.e., original concentration in the samples before SPE). The estimated TL concentrations were calculated from the 6PPDQ concentration based on the previous studies (Hu et al., 2022; Hu et al., 2023). Model predictions fall within the estimated TL concentration ranges were highlighted in red.

		Predicted concentration (mg/L as TWP)										Measurement and extrapolated concentration	
		Similarity scores			Regularized regression			PCA with regression				6PPDQ	Est. TL Conc
Sample Name	Flow condition	Jaccard_raw	Jaccard_bin	Simp	Enet	Ridge	Lasso	SVM	RF	Linear	XGBoost regression	(ng/L)	(mg/L)
ISCO10pm@042621	Base flow	52	15	48	51	191	53	76	100	72	10	6.5	**83~252**
ISCO11pm@042621	Base flow	45	15	26	28	-41	85	37	54	32	10	1.2	**15~47**
ISCO12am@042621	Base flow	68	15	54	45	-51	68	64	102	62	20	0.99	**13~38**
ISCO6am@042621	Peak flow	74	16	415	288	203	221	295	187	319	20	6.5	**83~252**
ISCO7am@042621	Peak flow	60	17	480	392	311	318	386	187	428	50	19	**244~738**
ISCO8am@042621	Peak flow	70	17	425	514	400	454	398	187	423	50	13	**167~505**
ISCO1@111121	Base flow	103	16	199	107	169	124	172	200	163	200	22.9	**294~889**
ISCO2@111121	Base flow	78	16	147	136	139	121	146	183	139	200	20.3	**260~788**
ISCO3@111121	Base flow	89	16	183	187	71	154	161	198	162	200	24.3	**312~944**
ISCO7@111121	Peak flow	88	17	439	288	261	387	311	200	321	200	61	**782~2369**
ISCO8@111121	Peak flow	89	17	371	279	393	359	370	200	391	500	70.2	**900~2726**
ISCO9@111121	Peak flow	93	17	352	452	552	396	376	580	399	1000	101.7	**1304~3950**
ISCO1@022622	Base flow	66	16	381	150	42	318	104	184	84	10	2.7	**35~105**
ISCO2@022622	Base flow	64	16	154	69	78	95	47	100	51	10	2.9	**37~113**
ISCO3@022622	Base flow	61	16	165	72	33	45	52	108	50	94	2.6	**33~101**
ISCO6@022622	Peak flow	101	18	554	432	547	483	452	576	482	500	86.5	**1109~3359**
ISCO7@022622	Peak flow	116	18	671	601	682	671	542	508	559	1000	114.2	**1464~4435**
ISCO8@022622	Peak flow	113	18	602	520	525	435	473	521	482	1000	103.3	**1324~4012**
ISCO4@041822	Base flow	83	16	141	100	14	95	100	102	101	10	4.1	**53~159**
ISCO5@041822	Base flow	83	17	326	285	333	269	306	200	318	500	40.1	**514~1557**
ISCO6@041822	Base flow	99	18	507	490	445	455	495	578	497	500	79.8	**1023~3099**
ISCO8@041822	Peak flow	97	18	508	435	412	428	456	564	469	1000	88.5	**1135~3437**
ISCO9@041822	Peak flow	109	18	613	486	578	452	508	509	520	500	111.4	**1428~4326**
ISCO10@041822	Peak flow	94	18	507	508	599	479	469	552	484	500	105	**1346~4078**
beforestorm@020223	Base flow	26	15	-3	44	79	12	33	56	38	73	3.9	**50~151**

6.4 Conclusion

In the current study, various models and algorithms were developed and tested for source tracking and apportionment outcomes. Sample collection efforts involved a representative lab simulated source dilution series in different chemical matrices and blind spiked samples to mimic actual field samples with unknown TL source concentrations. Both qualitative and quantitative methods were leveraged to facilitate the data analysis workflow. Source presence and background matrix identification were achieved using dot score logistic regression (Acc.=100% on blind sample data) and XGBoost classification model (Acc.=76% on blind sample data). On the other hand, different strategies were deployed for quantitative source apportionment and overcoming the limitations of data collection capabilities (sample collection and instrument data analysis). Such methods include feature selection, feature extraction, and regularization regressions. The results suggested that regularization models typically performed better on both spiked and blinded sample datasets with high accuracy and generalizability to the complex data set. The dimension reduction strategy coupled with SVM regression yielded accurate results; a direction that should receive further attention for future method development. These results proposed a prospective machine learning based workflow for accurate and automated source tracking.

Complementing traditional source tracking methods that focus on the concentration prediction or identification of pollution sources (Chae et al., 2021; Fauser et al., 1999; Rødland et al., 2022), the strategy proposed here provides a workflow for FP feature prioritization, source and background matrix identification, and source apportionment. These approaches were supported by machine learning algorithms, allowing users to better leverage data rich HRMS features for source tracking studies without substantial manual tuning, data processing, or identification efforts. Also,

using the FP features provided some advantages over using surrogate and marker chemicals for source tracking including: a) no commercial standards are required for FP features; b) as a cluster of features that were pre-selected by clustering analysis on different matrices, the FP features selected should be less affected by background matrix effects; and c) various options are available for users to select and optimize for accurate apportionment prediction results. We anticipate that the comprehensive and automated methodology described in the current study will be especially useful to the HRMS and environmental science communities to assist with source tracking studies.

Despite the advantages of the machine learning assisted source tracking methods, we reiterate that there still exist basic challenges of sufficient data collection (e.g., wider sample/matrix type coverage) that are critical to producing accurate and generalizable prediction models. Also, we recommend that optimization efforts focus on the FP feature selection and integrating background matrix information into models to avoid collinearity effects and promote improved performance. Also, due to typical limitations on funding and time, it is often impractical for a single lab to collect sufficient data that fully support model development and optimization for different source and background matrices. Therefore, there is a real need for a source tracking data repository where appropriate HRMS data can be shared and co-leveraged to support future development studies.

References

Aalizadeh, R., Nikolopoulou, V., Alygizakis, N., Slobodnik, J., Thomaidis, N.S., 2022. A novel workflow for semi-quantification of emerging contaminants in environmental samples analyzed by LC-HRMS. Anal Bioanal Chem 414, 7435–7450. https://doi.org/10.1007/s00216-022-04084-6

Barbosa, S., Saurina, J., Puignou, L., Núñez, O., 2020. Classification and Authentication of Paprika by UHPLC-HRMS Fingerprinting and Multivariate Calibration Methods (PCA and PLS-DA). Foods 9, 486. https://doi.org/10.3390/foods9040486

Brack, W., Hollender, J., de Alda, M.L., Müller, C., Schulze, T., Schymanski, E., Slobodnik, J., Krauss, M., 2019. High-resolution mass spectrometry to complement monitoring and track emerging chemicals and pollution trends in European water resources. Environmental Sciences Europe 31, 62. https://doi.org/10.1186/s12302-019-0230-0

Buerge, I.J., Poiger, T., Müller, M.D., Buser, H.-R., 2003. Caffeine, an Anthropogenic Marker for Wastewater Contamination of Surface Waters. Environ. Sci. Technol. 37, 691–700. https://doi.org/10.1021/es020125z

Campmajó, G., Saurina, J., Núñez, O., 2022. FIA–HRMS fingerprinting subjected to chemometrics as a valuable tool to address food classification and authentication: Application to red wine, paprika, and vegetable oil samples. Food Chemistry 373, 131491. https://doi.org/10.1016/j.foodchem.2021.131491

Chae, E., Jung, U., Choi, S.-S., 2021. Quantification of tire tread wear particles in microparticles produced on the road using oleamide as a novel marker. Environmental Pollution 288, 117811. https://doi.org/10.1016/j.envpol.2021.117811

Chu, B., Qureshi, S. 2022. Comparing Out-of-Sample Performance of Machine Learning Methods to Forecast U.S. GDP Growth. Comput Econ. 2022 Sep 16:1-43. doi: 10.1007/s10614-022-10312-z.

Cissé, G., 2019. Food-borne and water-borne diseases under climate change in low- and middle-income countries: Further efforts needed for reducing environmental health exposure risks. Acta Trop 194, 181–188. https://doi.org/10.1016/j.actatropica.2019.03.012

Dávila-Santiago, E., Shi, C., Mahadwar, G., Medeghini, B., Insinga, L., Hutchinson, R., Good, S., Jones, G.D., 2022. Machine Learning Applications for Chemical Fingerprinting and Environmental Source Tracking Using Non-target Chemical Data. Environ. Sci. Technol. 56, 4080–4090. https://doi.org/10.1021/acs.est.1c06655

Díaz, R., Ibáñez, M., Sancho, J.V., Hernández, F., 2011. Building an empirical mass spectra library for screening of organic pollutants by ultra-high-pressure liquid chromatography/hybrid quadrupole time-of-flight mass spectrometry. Rapid Communications in Mass Spectrometry 25, 355–369. https://doi.org/10.1002/rcm.4860

Drinking-water [WWW Document], n.d. URL https://www.who.int/news-room/fact-sheets/detail/drinking-water (accessed 4.20.23).

Du, B., Lofton, J.M., Peter, K.T., Gipe, A.D., James, C.A., McIntyre, J.K., Scholz, N.L., Baker, J.E., Kolodziej, E.P., 2017. Development of suspect and non-target screening methods for detection of organic contaminants in highway runoff and fish tissue with high-resolution time-of-flight mass spectrometry. Environ Sci Process Impacts 19, 1185–1196. https://doi.org/10.1039/c7em00243b

Du, B., Tian, Z., Peter, K.T., Kolodziej, E.P., Wong, C.S., 2020. Developing unique nontarget high-resolution mass spectrometry signatures to track contaminant sources in urban waters. Environmental science & technology letters 7, 923–930.

Ester, M., Kriegel, H.-P., Xu, X., n.d. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise 6.

Fauser, P., Tjell, J.C., Mosbaek, H., Pilegaard, K., 1999. Quantification of Tire-Tread Particles Using Extractable Organic Zinc as Tracer. Rubber Chemistry and Technology 72, 969–977. https://doi.org/10.5254/1.3538846

Fent, K., 2015. Progestins as endocrine disrupters in aquatic ecosystems: Concentrations, effects and risk assessment. Environ Int 84, 115–130. https://doi.org/10.1016/j.envint.2015.06.012

Hiki, K., Asahina, K., Kato, K., Yamagishi, T., Omagari, R., Iwasaki, Y., Watanabe, H., Yamamoto, H., 2021. Acute Toxicity of a Tire Rubber-Derived Chemical, 6PPD Quinone, to Freshwater Fish and Crustacean Species. Environ. Sci. Technol. Lett. https://doi.org/10.1021/acs.estlett.1c00453

Hu, X., Zhao, H.N., Tian, Z., Peter, K.T., Dodd, M.C., Kolodziej, E.P., 2022. Transformation Product Formation upon Heterogeneous Ozonation of the Tire Rubber Antioxidant 6PPD (N-(1,3-dimethylbutyl)-N′-phenyl-p-phenylenediamine). Environ. Sci. Technol. Lett. 9, 413–419. https://doi.org/10.1021/acs.estlett.2c00187

Hu, X., Zhao, H.N., Tian, Z., Peter, K.T., Dodd, M.C., Kolodziej, E.P., 2023. Chemical characteristics, leaching, and stability of the ubiquitous tire rubber-derived toxicant 6PPD-quinone. Environ. Sci. Process. Impact. 2023, Advance Article. https://doi.org/ 10.1039/D3EM00047H

Hug, C., Ulrich, N., Schulze, T., Brack, W., Krauss, M., 2014. Identification of novel micropollutants in wastewater by a combination of suspect and nontarget screening. Environmental Pollution 184, 25–32. https://doi.org/10.1016/j.envpol.2013.07.048

Jaccard index, 2023. . Wikipedia.

Kiefer, K., Du, L., Singer, H., Hollender, J., 2021. Identification of LC-HRMS nontarget signals in groundwater after source related prioritization. Water Research 196, 116994. https://doi.org/10.1016/j.watres.2021.116994

Klöckner, P., Seiwert, B., Wagner, S., Reemtsma, T., 2021. Organic Markers of Tire and Road Wear Particles in Sediments and Soils: Transformation Products of Major Antiozonants as Promising Candidates. Environmental Science & Technology. https://doi.org/10.1021/acs.est.1c02723

Lübeck, J.S., Alexandrino, G.L., Christensen, J.H., 2020. GC × GC–HRMS nontarget fingerprinting of organic micropollutants in urban freshwater sediments. Environmental Sciences Europe 32, 78. https://doi.org/10.1186/s12302-020-00353-2

Masiá, A., Campo, J., Blasco, C., Picó, Y., 2014. Ultra-high performance liquid chromatography-quadrupole time-of-flight mass spectrometry to identify contaminants in water: an insight on environmental forensics. J Chromatogr A 1345, 86–97. https://doi.org/10.1016/j.chroma.2014.04.017

Mazza, A., Piscitelli, P., Neglia, C., Rosa, G.D., Iannuzzi, L., 2015. Illegal Dumping of Toxic Waste and Its Effect on Human Health in Campania, Italy. Int J Environ Res Public Health 12, 6818–6831. https://doi.org/10.3390/ijerph120606818

Peter, K.T., Tian, Z., Wu, C., Lin, P., White, S., Du, B., McIntyre, J.K., Scholz, N.L., Kolodziej, E.P., 2018. Using High-Resolution Mass Spectrometry to Identify Organic Contaminants Linked to Urban Stormwater Mortality Syndrome in Coho Salmon. Environ. Sci. Technol. 52, 10317–10327. https://doi.org/10.1021/acs.est.8b03287

Peter, K.T., Wu, C., Tian, Z., Kolodziej, E.P., 2019. Application of Nontarget High Resolution Mass Spectrometry Data to Quantitative Source Apportionment. Environ. Sci. Technol. 53, 12257–12268. https://doi.org/10.1021/acs.est.9b04481

Peter, K.T., Herzog, S., Tian, Z., Wu, C., McCray, J.E., Lynch, K., Kolodziej, E.P., 2019. Evaluating emerging organic contaminant removal in an engineered hyporheic zone using high resolution mass spectrometry. Water Research 150 (2019) 140-152. https://doi.org/10.1021/ j.watres.2018.11.050

Qamar, K., Nchasi, G., Mirha, H.T., Siddiqui, J.A., Jahangir, K., Shaeen, S.K., Islam, Z., Essar, M.Y., 2022. Water sanitation problem in Pakistan: A review on disease prevalence, strategies for treatment and prevention. Annals of Medicine and Surgery 82, 104709. https://doi.org/10.1016/j.amsu.2022.104709

Rødland, E.S., Samanipour, S., Rauert, C., Okoffo, E.D., Reid, M.J., Heier, L.S., Lind, O.C., Thomas, K.V., Meland, S., 2022. A novel method for the quantification of tire and polymer-modified bitumen particles in environmental samples by pyrolysis gas chromatography mass spectroscopy. Journal of Hazardous Materials 423, 127092. https://doi.org/10.1016/j.jhazmat.2021.127092

Scholz, N.L., Myers, M.S., McCarthy, S.G., Labenia, J.S., McIntyre, J.K., Ylitalo, G.M., Rhodes, L.D., Laetz, C.A., Stehr, C.M., French, B.L., McMillan, B., Wilson, D., Reed, L., Lynch, K.D., Damm, S., Davis, J.W., Collier, T.K., 2011. Recurrent Die-Offs of Adult Coho Salmon Returning to Spawn in Puget Sound Lowland Urban Streams. PLOS ONE 6, e28013. https://doi.org/10.1371/journal.pone.0028013

Tian, Z., Peter, K.T., Gipe, A.D., Zhao, H., Hou, F., Wark, D.A., Khangaonkar, T., Kolodziej, E.P., James, C.A., 2020. Suspect and Nontarget Screening for Contaminants of Emerging Concern in an Urban Estuary. Environ. Sci. Technol. 54, 889–901. https://doi.org/10.1021/acs.est.9b06126

Tian, Z., Zhao, H., Peter, K.T., Gonzalez, M., Wetzel, J., Wu, C., Hu, X., Prat, J., Mudrock, E., Hettinger, R., Cortina, A.E., Biswas, R.G., Kock, F.V.C., Soong, R., Jenne, A., Du, B., Hou, F., He, H., Lundeen, R., Gilbreath, A., Sutton, R., Scholz, N.L., Davis, J.W., Dodd, M.C., Simpson, A., McIntyre, J.K., Kolodziej, E.P., 2021. A ubiquitous tire rubber–derived chemical induces acute mortality in coho salmon. Science 371, 185–189. https://doi.org/10.1126/science.abd6951

Triassi, M., Alfano, R., Illario, M., Nardone, A., Caporale, O., Montuori, P., 2015. Environmental Pollution from Illegal Waste Disposal and Health Effects: A Review on the "Triangle of Death." Int J Environ Res Public Health 12, 1216–1236. https://doi.org/10.3390/ijerph120201216

Tsugawa, H., Cajka, T., Kind, T., Ma, Y., Higgins, B., Ikeda, K., Kanazawa, M., VanderGheynst, J., Fiehn, O., Arita, M., 2015. MS-DIAL: Data Independent MS/MS Deconvolution for Comprehensive Metabolome Analysis. Nat. Methods 12, 523–526. https://doi.org/10.1038/nmeth.3393

Supplementary Materials

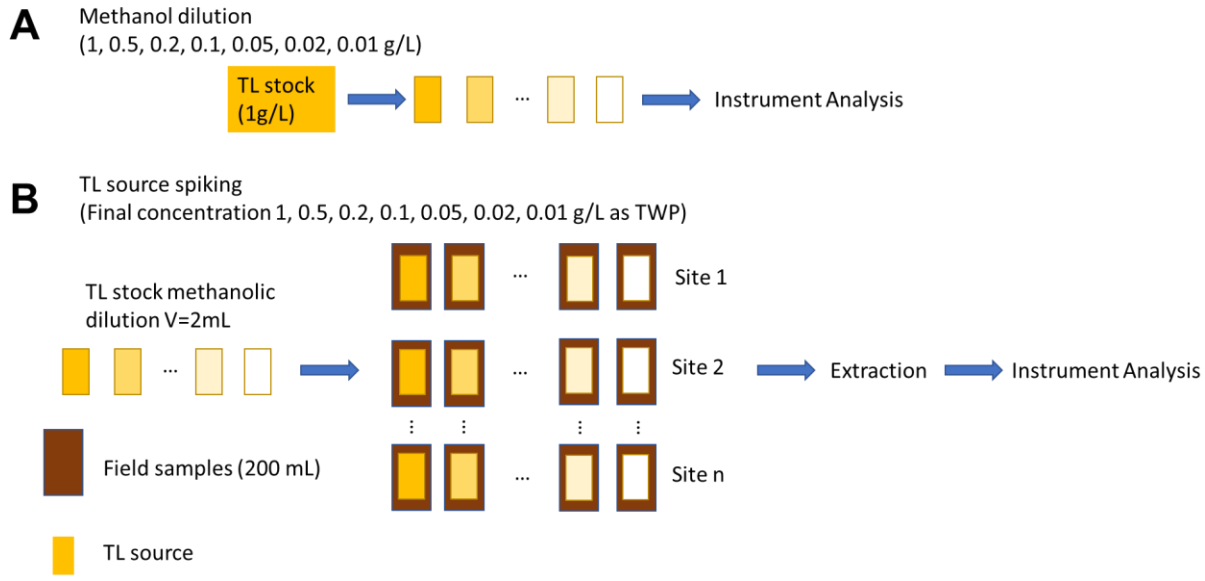


Figure S6.1. Diagram for TL dilution series sample preparation. (A) preparation of TL methanolic dilution series sample and; (B) preparation of TL source spiking series sample. All the samples were analyzed by LC-qTOF instrumentation, SPE was employed for TL source spiked samples.

Table S6.1. Spiking volumes of TL stock solution for field samples.

Dilution level	**L1**	**L2**	**L3**	**L4**	**L5**	**L6**	**L7**	**Blank**
TL concentration (g/L)	0.01	0.02	0.05	0.1	0.2	0.5	1	0
methanol volume spiked (µL)	1980	1960	1900	1800	1600	1000	0	2000
TL stock spiked (µL)	20	40	100	200	400	1000	2000	0

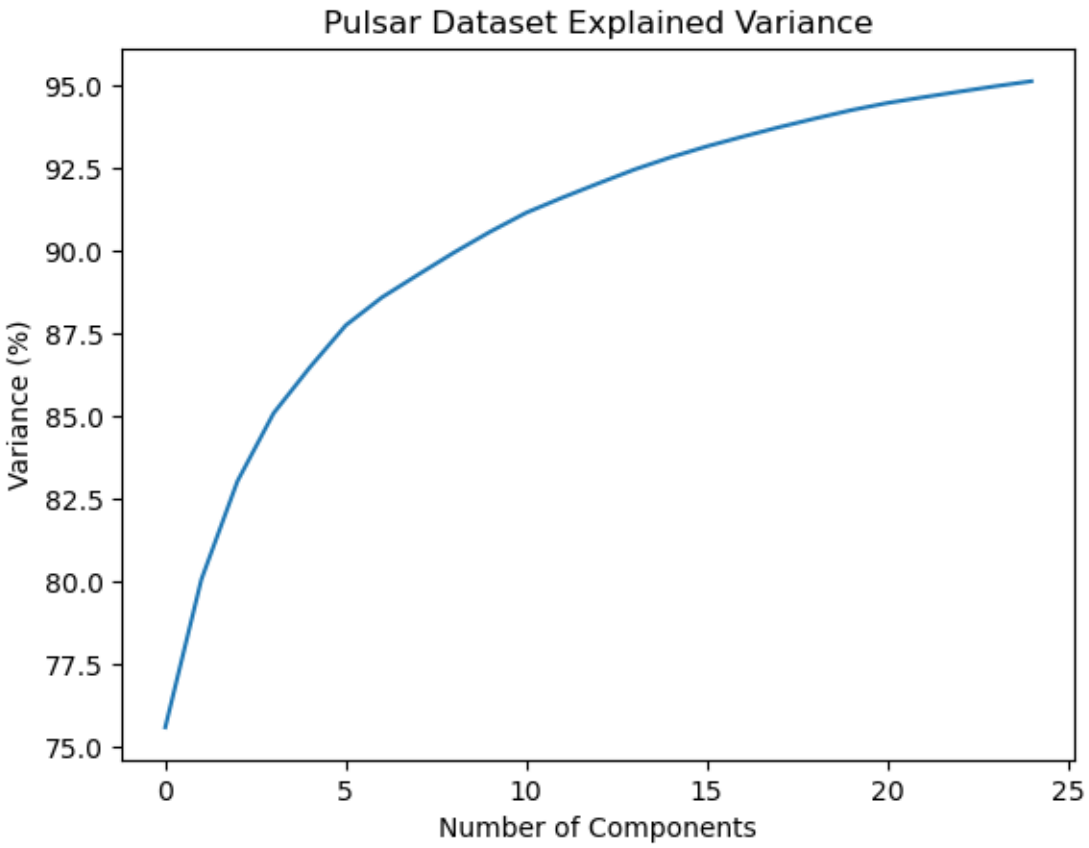


Table S6.1. Cumulative explained variance for principal components. For the 25 principal components reduced from the FP features, 95.12% of total variance was explained.

Chapter 7: General Conclusions and Recommendations

The overall objectives of this thesis consisted of two broad goals: (a) investigate the chemical characteristics, environmental transformation, and fate of the tire rubber antioxidant 6PPD and its TPs, especially 6PPDQ; and (b) develop improved workflows for HRMS data processing that facilitate NTA, especially with respect to source tracking and apportionment studies. Conclusions for the specific aims are outlined below.

Aim 1: Targeted quantification of 6PPDQ for physiochemical property measurements and environmental fate evaluations

A sensitive analytical method was deployed for detection and quantification of the 6PPD derived TP 6PPDQ in different samples to measure its physiochemical properties and environmental fate. We observed a moderately high logK_{ow} (4.30±0.02) for 6PPDQ, but far lower aqueous solubility (38±10 μg/L) than expected by software prediction modeling. We believe these effects are best explained by largely unrecognized solute-solute interactions discussed in **Section 2.3.1**. Given the widespread interest in PPD antioxidants, their quinones, and their alternatives, future studies should further investigate the role of solid-phase intermolecular interactions in explaining 6PPDQ characteristics and fate. The environmental fate of 6PPDQ was simulated with different approaches, the results suggested 6PPDQ is often highly sorptive to rubber and silicone (>92% 6PPDQ sorption during 30 min contact) materials while glass is a recommended for handling 6PPDQ (no obvious sorption of 6PPDQ under the tested conditions). Often, sorbed 6PPDQ could not be fully recovered from materials by typical solvent wash conditions. 6PPDQ leaching dynamics were determined with an up-flow column system; during the 6 h leaching tests,

a cumulative leached mass of 0.086±0.005 ng/cm^2 6PPDQ (5.15±0.3 μg 6PPDQ per g TWP; k_{diss}=0.0027±0.0005 min-1) was observed, indicating reasonably rapid discharge of 6PPDQ into roadway runoff during storms. Additionally, 6PPDQ aqueous stability was assessed under different pH conditions (pH 5, 7, and 9), the experiments suggested relatively high stability of 6PPDQ over 14 d (<10% decrease) and slight (26±10%) losses over 47 d in the tested buffer solutions. The observations above proved this chemical is a poorly water soluble but fairly stable contaminant in simple aqueous solutions. Further investigation regarding environmental fate and transport (e.g., soil sorption, multi-phase partitioning dynamics) of 6PPDQ is merited to facilitate risk assessment and remediation strategies.

Aim 2: Evaluate the transformation dynamics of 6PPD, 6PPDQ, and 6QDI, including formation of both 6PPDQ and other related TPs

We evaluated the transformation of 6PPD, 6PPDQ, and 6QDI and investigated their TPs formation dynamics under different environmentally relevant conditions (especially for 6PPD), including gas and aqueous phase ozone and aerobic exposures. The molar yield of 6PPD to 6PPDQ was initially determined ~9.7% from pure 6PPD and ~0.95% for 6PPD within TWP (Chapter 3), with nine potential environmentally relevant TPs prioritized via HRMS NTA approach and comparison with environmental occurrence data. Some TPs were then identified and quantified with commercial standards, including 1,3-DMBA, 4-NDPA, 4-ADPA, 1,3-DMBA, 4s DPA and 4-HDPA.

With refined targeted LC-MS/MS quantification methods developed by Zhao et al., the gas and aqueous phase ozonation of 6PPD was revisited under a broader set of conditions. Under the

tested conditions, 6PPD mass progressively reacted during short term gas (93±2% 6PPD transformed over 2.5 h) and aqueous ozone reactions (empirical reaction constant = $(1.18\pm0.16)\times10^{6}$). In addition to ozonation processes, stability tests of 6PPD under aerobic conditions indicated 6PPD instability under aqueous conditions (tested under pH7; 90.5±0.6% 6PPD transformed over 48 hours; half-life: 10.0 hours) and significantly more stability under gas phase exposures (41±3% 6PPD degraded over 42 days; half-life: 69.3 days). Under all tested conditions for 6PPD, only gaseous ozonation processes were observed to generate 6PPDQ, underscoring the importance of such gas phase abiotic interactions as the main contributor to environmental 6PPDQ formation. Notably, no 6PPDQ formation was observed in any aqueous solutions, even when high concentrations of ozone were present. To evaluate its stability, 6PPDQ ozonation in the gas phase was also investigated; these data suggested a slower reaction dynamic (60% 6PPDQ degraded over 4 hours) compared to 6PPD ozonation. Additionally, 6QDI ozonation was also tested to investigate potential 6PPD-6PPDQ transformation process and contradicted the presumption from other studies about 6PPD-6QDI-6PPDQ transformation pathway. These studies confirmed the presence of major TPs (1,3-DMBA, 4-NDPA, 4-ADPA, 1,3-DMBA, 4s DPA and 4-HDPA) for 6PPD family chemical transformation. Future studies should focus on precise characterization of 6PPD abiotic transformation process, including elucidating reaction mechanisms and pathways of 6PPD transformation or determining accurate 6PPD-O_3/O_2 reaction stoichiometries and rate constants. The information would be critical to environmental fate assessment and treatment method development for 6PPD.

Aim 3: Workflow development for HRMS NTA studies, with a focus on source identification and apportionment approaches

To facilitate open source HRMS NTA data processing, we developed a novel *MSS* package with Python programming language. The package consists of 5 modules with 24 features providing a full stack HRMS data processing capacity for data import, feature extraction and alignment, data reduction, advanced data mining (statistical analyses, feature clustering, and a novel source tracking function), data visualization, feature annotation, and reporting. Machine learning based algorithms (e.g., Random Forest modeling, PCA, DBSCAN) were employed to facilitate accurate and efficient data processing. Notably, *MSS* could accurately identify 99.5% of the peak signals and was >5 times faster than other similar platforms tested. The DBSCAN clustering analysis tool was also used in Chapter 3 to prioritize potential 6PPD TPs and should be broadly applicable to a wider range of study outcomes that screen environmental samples for contaminant prioritization. Future package updates should focus on integrating other NTA associated functionalities that are necessary to enable a full-stack HRMS data processing workflow, including structural annotation, in-silico fragmentation prediction, etc.

Beyond general functionalities that assist HRMS NTA, methodologies were developed and optimized for source tracking and apportionment studies based on bulk unidentified chemical features (FP features) within HRMS datasets. Lab-simulated source dilution samples were collected and processed using LC-HRMS (injections n=549) to generate training and testing data. Qualitative and quantitative models were developed for source identification and apportionment outcomes. Source presence and background matrix identification were achieved with dot score logistic regression (Acc.=100% on blind sample data) and XGBoost classification model (Acc.=76% on blind sample data). Several strategies were used to mitigate the challenge of model overfitting for quantitative analysis, including feature selection (XGBoost model), feature extraction (similarity scores and PCA) and regularization regression (lasso, ridge, and elastic net regressions).

Regularization regressions outperformed other algorithms (R^2 score = 0.94~0.98 and 0.88~0.89 for spike and blind samples), while PCA coupled with SVM modeling also produced good accuracy (R^2 score = 0.96 and 0.87 for spike and blind samples) and merits further investigations. To overcome the challenge of limited data collection capacity, future efforts should primarily focus on data collection and data simulation method development, as well as model optimization and workflow integration. There is also a pressing need to build a public database for HRMS data and establish a standardized protocol for sample collection and data processing to facilitate source tracking studies.