Prediction and Hazard Estimation of Polycyclic Aromatic Hydrocarbon Transformation Products

by

Trevor Sleight

Bachelor of Science, Geneva College, 2008

Masters of Science, University of Pittsburgh, 2010

Submitted to the Graduate Faculty of the

Swanson School of Engineering in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2021

UNIVERSITY OF PITTSBURGH

SWANSON SCHOOL OF ENGINEERING

This dissertation was presented

by

Trevor Sleight

It was defended on

June 21, 2021

and approved by

Ioannis Bourmpakis, Phd, Associate Professor, Department of Chemical and Petroleum Engineering

Vikas Khanna, Phd, Associate Professor, Department of Civil and Environmental Engineering

Leanne M. Gilbertson, Phd, Assistant Professor, Department of Civil and Environmental Engineering

Carla A. Ng, Phd, Assistant Professor, Department of Civil and Environmental Engineering

Prediction and Hazard Estimation of Polycyclic Aromatic Hydrocarbon Transformation Products

Trevor W. Sleight, P.E., PhD

University of Pittsburgh, 2021

Polycyclic Aromatic Hydrocarbons (PAHs) are a group of compounds containing at least two aromatic rings. Generated from natural or industrial processes, their degradation half-lives can range from weeks to months, as they undergo numerous environmental reactions resulting in diverse transformation products (TPs). While some PAHs possess known hazardous properties, relatively little is known about the hazards of their TPs. An increase in mutagenicity (the ability to cause genetic errors) has been observed as PAHs biodegrade. Since numerous TPs can be generated from each original PAH, evaluating which structures contribute to a potential increase in mutagenicity becomes a complex problem for remediators and regulators. The objective of this work was to build tools utilizing new and existing approaches to predict the most likely PAH TPs, identify which contributed to mutagenicity, and test the tools via an empirical experiment.

To achieve this objective, a network-based tool was developed to refine datasets of over 20,000 predicted PAH TPs to less than 200, for the parent PAHs acenaphthene, anthracene, fluorene, and phenanthrene, creating a manageable number of the highest likelihood compounds. Within this subset, the tool predicted up to 48% of PAH TPs found by previous empirical studies, aiding in the first step, predicting likely TPs. To address the second step of PAH degradation risk assessment, a method to predict the hazard – here mutagenicity - of the most likely TPs was needed, as available tools were not designed for biodegradation-induced mutagenicity. A QSAR

for PAH TP mutagenicity was developed which outperformed the best available QSARs when evaluating for biodegradation-induced mutagenicity and suggested that certain structural features corresponded to mutagenic mechanisms. Finally, the predictive tools were tested in an empirical study, aiming to identify the approximate time in a PAH's degradation that mutagens emerge. Biodegradation cultures with phenanthrene and fluorene suggested that the networks tool and the QSAR together could help target the occurrence of mutagenicity in a PAH's degradation timeline. Overall, this work provided two computational tools, the networks model for predicting the likely TPs, and the QSAR for estimating the mutagenicity of actively degrading PAHs and demonstrated their utility in biodegradation experiments.

Table of Contents

List of Tables

List of Figures

Acknowledgements

3 years went very quickly. I am grateful to my advisors, Dr Ng and Dr Gilbertson, my committee, comprised of my advisors and Dr Bourmpakis and Dr Khanna. I also appreciate the friends and family and our church that were part of this journey. Serving as president of EGSO was a unique opportunity and the efforts of my team made me proud of the quality of students we have at Pitt. May we always strive to make the world a better place.

1.0 Introduction

1.1 PAH Sources

Polycyclic aromatic hydrocarbons (PAHs) are a diverse group of compounds containing at least two aromatic rings. Approximately 500 Gg of Polycyclic Aromatic Hydrocarbons (PAHs) are emitted into the atmosphere every year.[1,2] Environmentally relevant sources include large-scale combustion events, including natural sources such as forest fires or volcanic activity, and anthropogenic sources such as power generation, some manufacturing operations[3] and transportation.[4,5] PAHs can be emitted into the atmosphere from incomplete combustion,[6] or leaked directly onto the surface from oil and gas production or irresponsible industrial practices.[7,8] In most areas with significant PAH contamination, anthropogenic sources significantly outweigh all natural sources.[9–12] The pollution burden varies by multiple orders of magnitude from country to country, so although the overall anthropogenic emissions of PAHs worldwide are slowly declining, the contamination level in many local regions is either static or increasing.[13,14]

Atmospherically emitted PAHs deposit onto the surface environment, contaminating soil and water. There, many are persistent contaminants, with half lives of up to several months.[15,16] Elevated levels of PAHs in various stages of transformation can be found in land and water near major PAH sources, even if the source has not been active for many years.[17–19] Biodegradation is the primary means of PAH mineralization and ultimate removal. However, an individual PAH can generate thousands of unique degradation intermediates and toxicological data is not available for all of these structures. Numerous studies (Table 2) have demonstrated an increase in mutagenicity, genotoxicity, or developmental impacts after environmental transformation of PAHs. These

transformations result in a changing hazard profile for PAH-contaminated environments, with the potential for multiple mechanisms of toxic action depending on the transformation products (TPs) that form as the PAH degrades. Estimating the potential harm to the environment in this dynamic system is a challenge for both remediators and risk assessors. De Souza et al, 2019 recommended specifically considering mutagenicity as a metric for the success of bioremediation.[20] Unfortunately, measuring mutagenicity directly requires a bioassay, which are typically time consuming and expensive. For example, the most well established bioassay for mutagenicity, the Ames Test,[21] requires almost a full day to perform and 2-3 additional days for the bioassay to run. It is also expensive and difficult to conduct, requiring a Biosafety Level 2 capable facility and all appropriate supporting equipment to appropriately prepare samples, store reference strains and maintain sterile procedures.[22] Due to the varied and dynamic nature of PAH contamination, implementing bioassays as a standard measure for successful bioremediation at the scale necessary to address the hazards involved in PAH remediation would be difficult. A computational or analytical method that would allow a reasonable estimate of the mutagenicity of partially degraded PAHs and enable a more targeted empirical approach would be extremely beneficial. Such an approach is explored in this research.

1.2 PAH Sources and Transformation Products

As with many forms of pollution, the specific PAHs present in a region tend to reflect the profile of the local pollution sources.[23] Certain PAHs are more strongly associated with combustive processes from either biomass or petrogenic sources. For example, vehicular emissions and similar petrogenic sources tend to cause a higher proportion of lower molecular weight PAHs,

including 3-4 ring PAHs, such as phenanthrene, fluorene and pyrene, with lower concentrations of pyrene being correlated to greater catalytic converter use.[24] Residential or agricultural burning of biomass fuels such as wood, dung, or crop waste primarily contributes 3-4 ring PAHs.[25] A greater proportion of 5 and 6 ring PAHs are found in emissions originating from coal combustion.[26]

Table 1 Mean atmospheric concentrations of unsubstituted, oxy, and nitro-PAHs

Location and Date of Study	Unsubstituted PAHs ng/m^3 (unless otherwise indicated)	Oxy-PAHs ng/m^3	Nitro-PAHs ng/m^3	Publication
Beijing, China 2012-2013	143	55.9	1.73	Lin et al, 2015[27]
Grenoble, France 2013	24.5	10.3	0.22	Tomaz et al, 2016[28]
Beijing, China, 2015 (Heating Season)	264	N/A	2.51	Zhang et al, 2020[29]
Rio de Janeiro, Brazil, 2017-2018,	Particle: 3.76 Gas: 10.03	Particle:1.62 Gas: 0.86	Particle: 1.00 Gas: 0.83	Santos et al, 2020[30]

The major subcategories of environmental PAHs include the unsubstituted PAH structures, oxy-PAHs, and nitro-PAHs. Although nitro-PAHs can be formed in the combustion process itself, the majority of those detected in environmental samples are believed to be the result of atmospheric reactions.[31–33] Oxy-PAHs may arise either from pyrogenic reactions alongside unsubstituted PAHs, or they may result from photochemical processes in the air or microbial processes in the

soil.[34,35] Several atmospheric studies have evaluated the relative concentration of nitro- and oxy-PAHs vs the unsubstituted PAHs (Table 1). The distribution of transformation products also varies with the source as well as with the reactions in the environment that begin to occur immediately after emission, and the specific PAHs, oxy-PAHs, and nitro-PAHs which were quantified vary among studies. However, the overall trends are consistent: unsubstituted PAHs are roughly 2-3 times as abundant as oxy-PAHs, and 10 to 100 times as abundant as nitro-PAHs. Once associated with soil or water, oxy-PAHs become the dominant TP, as these are formed from microbial degradation.[36,37]

1.3 Fate and Transport

The mean half-life of a PAH in the atmosphere is typically estimated to be only a few days.[38,39] PAHs can sorb onto atmospheric particles, or be collected by rain and subsequently deposit out of the atmosphere,[40] with the dominant mechanism (wet or dry) varying based on the local weather conditions.[41] This results in higher levels of PAH contamination close to major sources of PAH emission. Once on the surface PAHs exhibit half-lives of weeks to months as they slowly degrade through many intermediates.[15,16] PAHs can also be transported by water movement on the earth's surface, either by dissolving or by sorbing onto suspended solids and following the particle's transport phenomena. PAHs can also be transported by man-made water-flows such as wastewater or sewers. Hot spots of concentration can occur in local regions due to specific sources as well as environmental fate and transport. Local contamination levels have also been tied to seasonal variations in environmental factors.[11,42] Consequently, locations such as estuaries or

rivers can have higher concentrations relative to the areas around them and some sites have seasonal variations in PAH concentrations.[42–44]

1.4 Environmental Impacts

The U.S. EPA has identified 16 PAHs as priority pollutants.[45] These span a range of sizes and possible ring arrangements. There are thousands of possible PAHs, so although the priority PAHs are a useful starting point, they are by no means exhaustive. Although the PAHs on the priority pollutants list have been well studied[14], hazard assessment for their biodegradation TPs remains an unsolved challenge for both regulating agencies and remediators trying to rehabilitate areas of PAH contamination.

Despite the fact that some microorganisms can degrade PAHs, contamination has been shown to suppress the overall respiration of microbial communities,[46] as well as lead to shifts in community diversity.[47,48] Benthic invertebrates are impacted by PAHs[49] and toxicity towards small earthworms[50] and fish[51,52] has also been documented (Fig 1). Bivalves are particularly vulnerable as they lack the capability to effectively degrade and eliminate PAHs.[53,54] Tumors were found in oysters in Black Rock Harbor, Connecticut, US where they were exposed to PAHs and other co-occurring chemicals.[55] Additionally, many PAHs can bioaccumulate through the food web, causing higher level predators such as fish, mammals, and humans to be exposed through the food chain.[56] Mutagenicity of environmental samples has been found to be associated with PAH contamination in highway runoff and cokery pollution.[57,58]

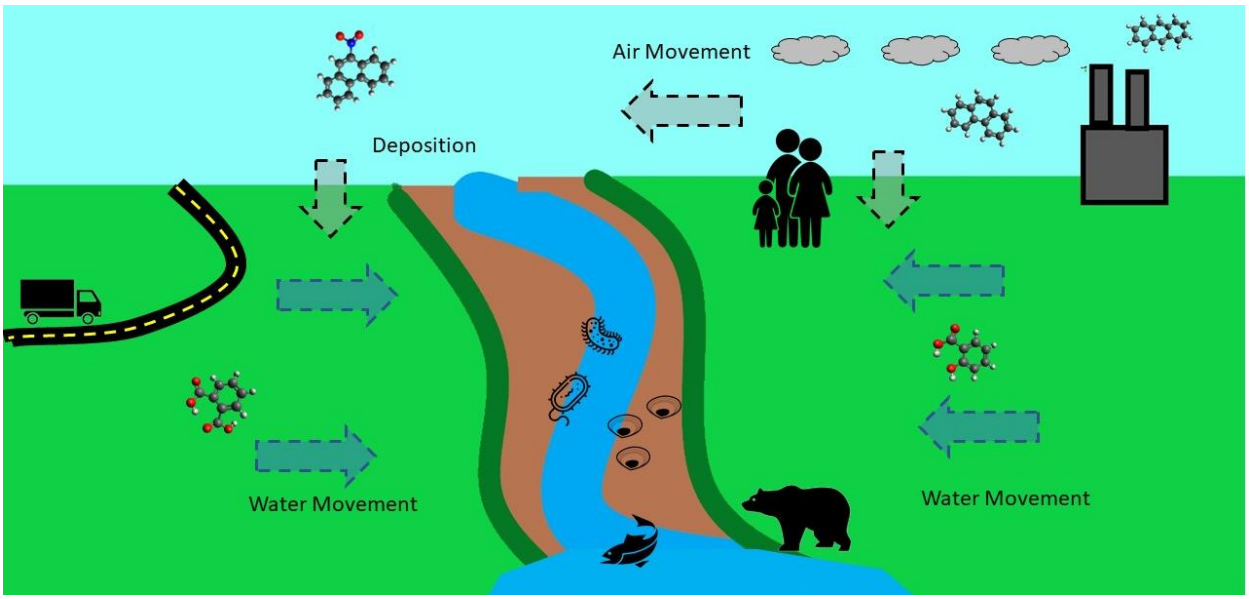


Figure 1 PAHs Fate and Transport pathways. Atmospheric and surfance environmental factors transport PAHs through air and water movement. The PAH TPs cause toxic impacts to organisms living in the different stages

1.5 PAH Biodegradation

Literature indicates that the primary mechanism of PAH transformation leading to ultimate removal from the environment is biodegradation.[59–62] Much of the microbiome that is capable of degrading PAHs lives in the aerobic region of soil and surface water. PAHs found in the anerobic region of deep lakes or aquifers degrade far more slowly.[63,64] Furthermore, the microbiome of the aerobic region is critical to the health of the overall ecosystem, playing a crucial role in nutrient cycling.[65]

As PAHs aerobically degrade, polar oxygen groups are added onto their structures. In some PAHs this can also occur abiotically due to photooxidation.[66] During this process, PAHs can

generate dozens to thousands of possible intermediate transformation products.[67–69] The intermediates generated may depend on the original PAH, the available local microbes and the degree of degradation or remediation that has already occurred. The complexity of the possible different degradation pathways renders hazard assessment a significant challenge for both regulators and risk assessment professionals.[70–72]

1.6 Toxicity Mechanisms of PAH Biodegradation Metabolites

PAH's baseline toxic mechanism is narcosis; resulting from the compound's interfering with the cell membrane.[73,74] Narcosis correlates well with a compound's octanol-water partition coefficient (K_{ow}). Under aerobic conditions, where the majority of PAH degradation occurs,[63,64] PAHs become smaller as rings open and are cleaved off, and more polar as oxygen groups are incorporated into their structure. This lowers their K_{ow} and also lowers their narcotic toxicity. K_{ow} can be estimated with reasonable reliability using a variety of methods, and quantitative structure activity relationships (QSARs) based on K_{ow} have been found to adequately predict narcotic toxicity.[75–77] However, different methods are needed to estimate toxicity by other mechanisms.

Studies of carcinogenicity and mutagenicity have shown that PAHs require metabolic "activation" by enzymes before they become carcinogenic.[78–82] This activation adds oxygen groups to the perimeter of the PAH molecules,[83,84] and can occur in an organism through metabolic pathways as the organism attempts to digest or eliminate the PAH. Alternatively, environmental microorganisms can cause a similar effect as they degrade PAHs.[37] This results in the potential for the presence of direct-acting mutagens, which are mutagenic without need of enzymatic activation, after PAHs have been biodegraded either due to natural processes or deliberate bioremediation.[36,37]

Several types of studies (Table 2) have demonstrated that it is the oxygenated or nitrogenated PAH derivatives that are responsible for the mutagenic impact of PAH contamination in the environment.

Table 2 Selected previous studies demonstrating increased mutagenicity after environmental transformation.

Media	**PAH Derivative**	**Study**
Air, gas phase	nitro-fluorene, nitro-naphthalene	Arey, 1992,[85]
Sediment	Polar vs non polar fraction	Fernandez, 1992[86]
Air, particles	nitro-PAHs, oxy-PAHs	Topinka, 1998[87]
Air, particles	nitro-PAHs	Topinka, 2000[88]
Water	oxy-phenanthrene	Schrlau, 2017[89]
Air, particles	nitro-PAHs, oxy-PAHs	Umbuzeiro, 2008[90]
Soil	oxy-PAHs	Mattsson, 2009[91]
Air, particles	Nitro-PAHs, oxy-PAHs	Wang, 2011[92]
Coal	Polar vs non-polar	Meyer, 2014[93]
Water	Oxy-PAHs	Chibwe et al, 2015[94]
Soil	Oxy-PAHs*	Wincent, 2015[95]
Sediment	Polar vs non polar	Di Giorgia, 2016[96]
Soil	Not specified	de Sousa, 2019[20]
Soil	Polar, oxy-PAHs	Park, 2008[97]

*Developmental Toxicity in Zebrafish

All of the studies in Table 2 were conducted with environmental samples. In addition to these studies, there are also several studies where known PAH intermediate TPs are evaluated as individual molecules, further demonstrating the mutagenic potential of PAH TPs.[98–100]

1.7 Computational Tools

With complex problems such as PAH degradation, experiments cannot be conducted for every scenario due to time and resource constrains, or to theoretical conditions that cannot be achieved in a reproducible manner. In these cases, modeling approaches can provide fruitful insight into how and where to spend precious resources in order to collect the most relevant empirical results. This work incorporates both computational and empirical tools to address the challenge of evaluating the mutagenicity that may result from aerobic PAH biodegradation. Available tools were used where possible to expand upon existing work and provide unique options for improving the field of ecological risk assessment. The novel computational tools were then used to guide and design an empirical experiment to attempt to detect the occurrence of mutagenicity in a live biodegradation culture.

1.7.1 Networks Analysis

Network theory is a powerful tool for studying a complex system of interacting objects. The first known network, or graph, was drawn by the German mathematician Leonhard Euler, to represent a transportation problem in the city of Konigsberg.[101] This representation of physical space highlighted the relationships between different points rather than the details of exact

measurements, as represented in Euclidean geometry. The focus on relationships laid the groundwork for Einstein's theory of relativity.[102] Networks can be directed or undirected and can contain additional attributes about the objects within them. Network theory has shown substantial value in numerous fields including social sciences,[103] trade[104] and food webs.[105,106] Optimizing algorithms to expand its applicability and usefulness is an area of active interest.[107,108] The capability of the networks approach to elucidate information about the elements of the network based on the relationships between its members has significant promise in the study of PAH biodegradation intermediates. The connections between the different TPs can be represented as a network and analyzed to learn more about both the individual TPs and the network as a whole.

1.7.2 Structural Activity Relationships

Structural activity relationships are a method of estimating chemical information based on a molecule's structure. These relationships have been recognized since the early understanding of molecular structure. However improvements in computational tools over the last several decades have dramatically improved their utility and accessibility.[109–112] QSARs remain an active area of research for drug design, chemical safety and regulatory evaluation. These relationships can predict chemical properties, such as solubility or boiling point, or biological properties such as toxicity or mutagenicity. Used within the proper applicability domain, QSARs can be valuable screening tools for focusing more time-consuming empirical work.

1.7.3 Empirical Analysis

PAH biodegradation experiments can be accomplished using known PAH degrading strains.[113,114] The concentration of the original PAH and in some cases the transformation projects can be monitored with analytical equipment, reflecting the progress of the PAH degradation. Bioassays are available to evaluate the toxic properties of the TPs resulting from biodegradation. The gold standard bioassay for mutagenicity, and the one on which most QSARs are trained is the Ames Test.[21,115] Ames Test kits are commercially available, and this bioassay can be conducted in a few days in a BSL-2 capable facility. Degrading a single PAH in a batch culture in a laboratory allows us to test the tools in a controlled setting and employ the Ames Test to measure the occurrence of mutagenicity in live PAH biodegradation cultures, over the course of the PAH's degradation timelines. The results can be compared against the predicted likelihood of mutagenicity from the computational tools.

1.8 Dissertation Research Objectives and Scope

The *overall objective* of this research was to develop a method of analysis that could be used to predict the most likely PAH metabolites from aerobic environmental degradation, classify them as mutagens or non-mutagens, and test the method's predictive abilities with an empirical experiment.

The *central hypothesis* was that network theory combined with a pathway predictor, would provide a realistic estimate of the most likely metabolites, that a combination of energetic and

topological molecular descriptors will be able to classify the most likely mutagens, and that both of these tools will yield predictions which can be empirically verified.

The following specific aims were pursued to develop a networks method to identify the structures of the most likely metabolites, identify the most likely mutagens with a QSAR that was simple enough to be connected with known mechanisms, and use the Ames Test to detect the occurrence of mutagenicity based on the predictions from networks method and the QSAR.

Specific Aim 1: ***Develop a networks-based model for environmental PAHs to predict the most likely metabolites to form in a PAH biodegradation system***. In this aim we collected all of the realistic possible pathways from the EAWAG-Pathway Prediction System (PPS) in order build a large network of the possible intermediate TPs. A simple, but unique algorithm was developed in order to reflect the weighted contribution of multiple pathways, yielding a much smaller dataset of the most likely structures to occur in an environmental context.

Specific Aim 2: ***Develop a Quantitative Structure-Activity Relationship (QSAR) to identify mutagenic PAH metabolites and determine the properties that are associated with mutagenicity.*** This objective explored the structures most likely to be suggestive of a mutagen, allowing us to identify where in the degradation chain mutagenic compounds are most likely to occur. A combination of supervised and unsupervised machine learning methods was used to develop a system that could reflect a wide range of different sizes of molecules with different mutagenic mechanisms.

Specific Aim 3: ***Experimentally determine the occurrence of mutagenicity in a PAH degradation batch culture to confirm outputs from the model and/or identify new mutagenic intermediate metabolites.*** A known hydrocarbon degrading strain was used to biodegrade PAHs in a laboratory flask culture. The degradation was monitored with a High-Performance Liquid

Chromatograph (HPLC), and aliquots were run on the Ames Test at regular intervals. The media prior to exposure to the bacterial degradation was not mutagenic, but the occurrence of mutagenicity was observed after partial degradation.

1.8.1 Dissertation Organization

Chapter 1 provides a summary of the overall environmental context of the problem of hazard assessment for PAH transformation products. Chapter 2 summarizes the work done to generate our first publication from this project.[68] This component of the project focused on utilizing a well-established degradation pathway predictor to generate a networks of possible PAH TPs and the biodegradation reactions that connect them, and used custom algorithms based on network theory to refine the network down to a more feasible scale. Chapter 3 is an explanation of the custom QSAR that we developed to evaluate the mutagenicity of potential PAH TPs. The currently available QSAR tools did not provide specific enough resolution to discriminate between mutagens and non-mutagens for the PAH TPs which occur in surface soil and water: oxy-PAHs. Chapter 4 contains an overview of the batch culture experiments conducted to validate the tools developed in the earlier chapters. Code for all of the analysis methods developed is available on the Ng Lab github, https://github.com/ngLabGroup or on Trevor Sleight's github account. https://github.com/twsleight/Environmental_PAH_Mutagenicity.

1.8.2 Intellectual merit and broader impacts

This project was designed to utilize the tools of network analysis and predictive toxicology to aid in the problem of hazard assessment for PAH biodegradation. PAHs are one of several

categories of compounds that have long, complicated degradation chains. The techniques explored in this work provide insight into the hazards of the degradation TPs of PAHs. However, they could also be applied to other molecules with similar complex degradation chains and potentially hazardous intermediates. Both the U.S. EPA and The European Union's Registration, Evaluation, Authorization and Restriction of Chemicals (REACH) require evaluation of environmental transformation products in certain situations,[116,117] and QSAR use is encouraged and accepted by both agencies for many requirements.

2.0 Networks Analysis

2.1 Introduction

Polycyclic Aromatic Hydrocarbons (PAHs) are environmental contaminants containing at least two aromatic rings in their chemical structure. PAHs are released into the environment from either pyrogenic or petrogenic sources. Pyrogenic PAHs are created by incomplete combustion of a carbon based fuel including wood, coal, or dung.[118] Petrogenic PAHs are caused by refining petroleum products, and to a lesser extent, releases from the production and use of plastic consumer products.[119] The contribution from each anthropogenic source varies widely across different countries depending on their industry, agriculture, and power production.[1] Natural sources of PAHs include volcanic activity and forest fires. PAH exposure is a significant human health concern, contributing to cancer and cardiovascular diseases.[120] While PAHs can be inhaled, the typical half-lives for mid-sized PAHs to settle out of the atmosphere are estimated to be only hours to days, depending on atmospheric conditions,[38,121] while PAHs can persist for

several months in soil or water before degrading, with larger PAHs demonstrating greater persistence.[15,16]

PAHs are introduced to soil or water through deposition from the atmosphere, direct spillage from petroleum production, and/or impermeable surface runoff.[122] They can dissolve into water or sorb onto soil colloids and be transported by water movement, causing areas of concentrated contamination where they are deposited. From soil or water, PAHs can undergo volatilization, photochemical transformations, and microbial biodegradation.[121] Under sunlight, photochemical degradation can rapidly oxygenate the perimeter carbons of PAHs, especially the larger ones,[123] and in some cases open rings.[120] These pathways are often complementary to biodegradation pathways and generate similar products.[66] However, biodegradation is the most effective mechanism for mineralization of PAHs, resulting in their complete removal from the environment.[59–61] Microorganisms utilize a variety of oxygenase enzymes to hydroxylate the outer rings of PAHs, enabling ring cleavage enzymes to open the rings and break down the aromatic structure of the compound.[59] Some studies show low or minimal PAH transformation without the presence of microorganisms when comparing biotic and abiotic soil conditions.[60]

From a hazard assessment perspective, the aerobic region of soil and surface water is of particular interest because much of the microbiome capable of degrading PAHs exists in this region, and PAHs accumulate in soil and water rather than in the atmosphere due to longer soil and water half-lives and low volatility. If PAHs settle into the anaerobic zone in the sediment of a deep lake, river, or aquifer, their degradation rates drop significantly.[63,124] Many organisms including benthic invertebrates, fish, amphibians, and small mammals live or feed in the aerobic region of the environment, and are exposed to the toxic effects of PAH contamination. PAHs can bioaccumulate in food webs, and are found in human food sources.[125] Intermediate metabolites

created during the biodegradation process are a subject of concern for PAHs,[36,126] and also many other environmental contaminants including pharmaceuticals, solvents, and antibiotics.[127] Estimating the overall environmental impact of intermediate metabolites, which could include persistence, bioaccumulation potential, or toxicity, poses a challenge for both regulators and remediation professionals due to the broad spread of possibilities, low concentrations, and complex degradation pathways. Environmental transformation products are beginning to be covered by regulations. The EU's Registration, Evaluation, Authorization and restriction of CHemicals (REACH) legislation requires evaluation of the transformation products of chemicals produced or imported in quantities above 100 tonnes/yr.[116] Although PAHs are not common commercial products, the interest in including transformation products in regulations may eventually apply to contaminants also. The US Environmental Protection Agency (EPA)'s methods of PAH risk assessment currently evaluates the parent compounds only, although there is interest in improving the accuracy of these methods and reflecting the uncertainty of mixtures.[128]

Multiple mechanisms of toxicity are possible for PAHs and their metabolites. Some PAHs can be metabolically activated to become carcinogens.[81] However, in absence of other knowledge, narcosis is considered to be the baseline mechanism of toxicity to most invertebrates,[129–131] and can be caused by both directly emitted PAH compounds and their degradation products. Narcotic toxicity is well correlated with a compound's octanol – water partition coefficient (log K_{ow}), and is used in many quantitative structure activity relationship (QSAR) estimation tools, such as those published by the EPA.[75] However, recent research suggests that narcotic toxicity does not represent the entire picture of PAH toxicity, and that PAHs may exhibit synergistic effects through more complex toxicity mechanisms.[132] Numerous

studies have noted increased developmental toxicity or genotoxicity in the presence of PAH metabolites, including PAHs of low molecular weight (3 rings or less).[89,94,132–136]

Unfortunately, in many cases it is not possible with existing methods to identify exactly which chemical structures are causing toxic effects on an ecosystem. Sources of PAHs may not be constant, and seasonal weather patterns as well as wind and rain can impact PAH deposition rates.[137] The variable nature of the emission of PAHs to the environment, as well as the complexity of the environmental microbiome can result in some intermediate metabolites occurring very briefly or only for a limited period of time during the degradation process.[137,138] Numerous analytical techniques can be used to empirically study PAH transformation products, usually involving a combination of a chromatograph to separate compounds and a detector to measure them.[69] However, all currently available methods have difficulty resolving very similar compounds, and many intermediates may occur outside of the sampling period and in low concentrations.[89,139,140] Additionally, even without the challenge of co-eluted compounds, many empirical studies of PAHs are challenged by unidentifiable compounds due to either a lack of available standards or limits in mass spectrum libraries.[136,141–143] As such, there is a need for improved methods of prioritizing potentially formed metabolites by their frequency or likelihood of occurrence as well as the potential to cause an environmentally relevant toxic impact.[70]

Such an approach was developed herein, using an innovative integration of degradation product prediction and network theory. We used the EAWAG Pathway Prediction System (EAWAG-PPS http://eawag-bbd.ethz.ch/predict/), hosted by the Swiss Federal Institute of Aquatic Science and Technology (EAWAG), to predict potential transformation products of PAHs.[144] EAWAG-PPS is a biodegradation tool which predicts microbial reactions using a molecular substructure search and a transformation rule database. Thousands of possible

biodegradation intermediates can be generated from each original PAH, resulting in a rich, but overwhelming data set.

Network theory leverages the relationships between interconnected items to elucidate specific information about the individual objects and the relationships between them.[145] Network theory has seen significant utilization in multiple domains including social sciences, biology, transportation, and food webs.[105] Individual objects are referred to as nodes and the links between them are referred to as edges. Network edges can contain information such as weight or direction, and individual characteristics can be attached to the nodes. Mathematical relationships describing the connections between the different nodes and edges can be used to represent the entire network, or extract specific information about particular sub-sections of the network. In the current application, the nodes represent the plethora of possible degradation metabolites, and the edges represent the biodegradation reactions from one compound to another. Each node is represented by a Simplified Molecular Input Line Entry Specification (SMILES) code, which indicates a specific chemical compound and each edge is a biodegradation reaction as defined by EAWAG-PPS. Networks analysis has previously to study metabolic networks, although in many cases data are organized and presented as a network, but not investigated using network analysis methods. When network analysis is used, it has demonstrated significant value in revealing new possible metabolic pathways,[146,147] understanding robustness in complex reaction networks,[148,149] and in locating the most biologically relevant pathways among known reactions.[150] Our methodology is a contribution to the techniques used to refine networks for biological relevance.

We present a feasible model that is able to focus the user on the most relevant possible degradation compounds. This method is able to elucidate the hidden portions of the degradation chain, which are often difficult to study empirically due to low instantaneous concentration,

numerous complex structures, and the lack of available standards. Our work will enable a better understanding of the pathways that are most likely to form toxic transformation products in the environment.

2.2 Data and Methods

Four of the PAHs on the EPA's Priority Pollutants List;[45] acenaphthene, anthracene, fluorene, and phenanthrene were selected for this analysis. These compounds were chosen as candidates for testing the network analysis approach because of the variations in the ring structure between them and a mix of heavily studied compounds such as phenanthrene and less studied compounds such as acenaphthene.

2.2.1 Chemical Characteristics

The following chemical characteristics were considered for correlations with network data: log Kow, number of oxygens, number of carboxylic acid groups, number of carbonyl groups, number of aromatic and aliphatic hydroxyl groups, number of nonring carbons, and number of rings. These descriptors were selected as they reflect the different ways that the PAH rings are oxygenated, opened, and ultimately degraded. Because of its capability to handle high volumes of data, RdKit[151] was used to calculate all molecular structure descriptors based on the SMILES codes obtained from EAWAG-PPS, with the exception of log K_{ow}. The Kowwin module of the Estimation Program Interface Suite (EPI Suite) was used to predict log K_{ow},[152] as the Kowwin method performed better than the RdKit Wildman–Crippen method on several literature data

sets.[153–156] The Python libraries, Pandas,[157] NumPy,[158] and MatplotLib[159] were used to manage the data with custom scripts, available on github: https://github.com/ngLabGroup. A summary of the scripts and how to use them is provided in the Chapter 2 Supporting Information. (see attached file Chap2_networks_analysis_SI.docx)

2.2.2 Network Construction

The network of degradation metabolites was constructed using batch processing from the EAWAG-PPS[144] (batch access kindly provided by Dr. Kathrin Fenner, EAWAG). Only the aerobic likelihood categories of "neutral", "likely", and "very likely" were included. Aerobically "unlikely" reactions were assumed to be negligible. An Edge Aerobic Likelihood Score was assigned to each category and used as the numerical weighting in the network calculations. The Edge Aerobic Likelihood Score was selected based on a sensitivity analysis designed to optimize the network's agreement with empirical literature data. The "Neutral" category was assigned an Edge Aerobic Likelihood Score of 1, "Likely" a score of 33, and "Very Likely" a score of 65. The full sensitivity analysis is presented in the SI. All edges were assumed to be unidirectional and irreversible. EAWAG-PPS does not predict methylation or nitrification of PAHs, so pathways containing these steps could not be included in the model, although these could be added to future versions. Compounds that did not contain at least one aromatic ring were only included in the network as sinks, defined as leading directly from a ring-containing compound to a non-aromatic compound. Compounds beyond sinks would not contain rings and were assumed to have low toxicity and persistence relative to the ring-containing compounds, and were not considered in the network.

2.2.3 Network Analysis

Three methods of network analysis were employed to describe the complex transformation pathways of PAH biodegradation. First, analyzing the network as a whole, network characteristics were correlated with the molecular descriptors of individual nodes. Second, the compounds that have the highest betweenness centrality contain identifiable structural patterns. These compounds were grouped into similar substructures based on Morgan Fingerprints[160] using Tanimoto Similarity[161] and Butina Clustering[162] and the most common structures were identified. Finally, the compounds identified by the novel Node Throughput algorithm were compared against the compounds identified by the empirical literature survey.

2.2.4 Betweenness Centrality

Betweenness centrality is a commonly used node-level metric in network analysis. A node's betweenness centrality is the number of geodesic paths between any other two other nodes in the network that pass through the specified node, normalized by all of the geodesic paths between those other two nodes.[105] The Edge Aerobic Likelihood Score is used as the edge weighting for this calculation. This measure is computed for each node in the network and is used to reflect the influence that a node has over all the other nodes in the network. In the chemical degradation network, betweenness centrality is used to highlight nodes (compounds) that are particularly influential in the pathways that are followed during network degradation. Betweenness centrality was obtained using the python package Networkx.[163] In our chemical degradation networks, a high value of betweenness centrality indicates that a compound falls on a large number of heavily weighted pathways in the degradation chain.

2.2.5 Node Throughput Algorithm

In addition to using an established network algorithm, we developed a new algorithm, hereafter termed "Node Throughput", designed to highlight specific nodes and edges that occur with a higher likelihood than other nodes. Each edge's "Throughput" value is assigned by the following:

$$\text{Edge Throughput Value} = \text{Edge Aerobic Likelihood Score} \times \frac{\text{Source Node Throughput Value}}{\sum \text{All Outgoing Edge Aerobic Likelihood Scores}} \quad (1)$$

and each node's "Throughput" value is assigned by:

$$\text{Node Throughput Value} = \sum \text{Incoming Edge Throughput Values} \quad (2)$$

Both equations are applied recursively until the entire network is accounted for. Once the edge weightings have been assigned by equation 1, Node Throughput could also be computed with a node strength function, however, the recursive method is necessary to compute the Edge Throughput values, so it is simplest to compute them together. For the purposes of this paper, "Node Throughput" refers to the node strength assigned by the above method, so as not to confuse it with a node strength value that could be computed by another method. Figure 2 provides an example of Node Throughput, in which all edges are weighted equally for simplicity in presenting the concept. However, in all calculations they were weighted as described above.

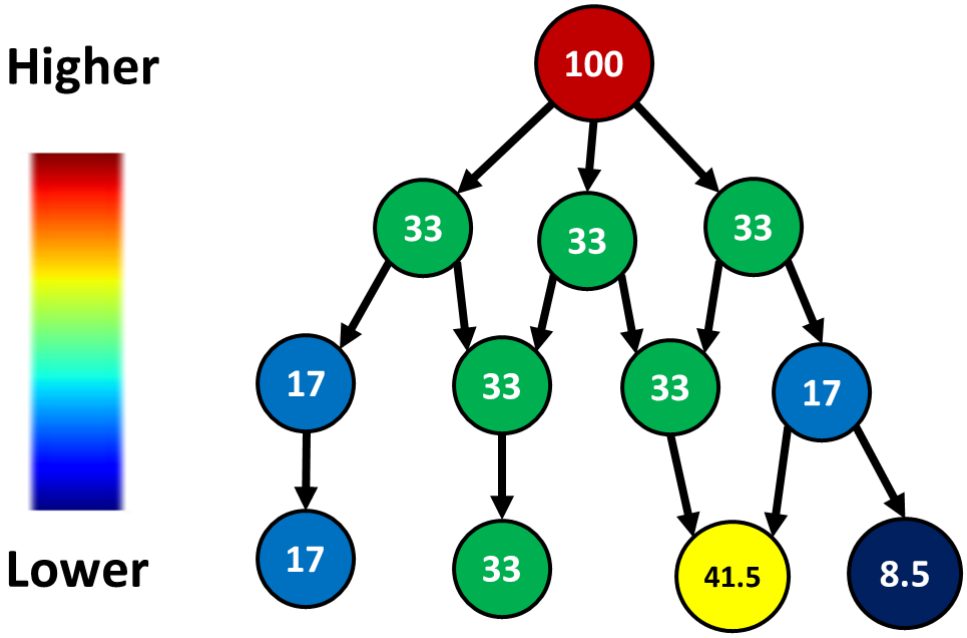


Figure 2 Schematic of the Node Throughput metric. The Throughput divides at the outgoing nodes and sums when multiple edges point to the same ingoing node. After several generations, certain portions of the network accrue more Throughput than other portions. Each layer sums to the total at the start.

2.2.6 Relevant Pathway Filtering

In addition to identifying nodes of interest using the networks metrics described above, we developed a systematic procedure for trimming a network to include only the portions of it that contain nodes with a specified property. The nodes are selected based on a molecular descriptor of interest, and then the pathways that flow to the nodes of interest are traced back to a source node. This reduces the size of the network without sacrificing relevant information about the desired parameters. In this analysis, nodes with a log K_{ow} value > 2.0 were selected as nodes of interest, based on a study by Voutchkova et al, 2011 which indicated that log K_{ow} > 2.0 may indicate

potential toxicity.[164] Log K_{ow} was used as an example for the analysis below; however, any desired property that could be computed from a SMILES code could be used (Fig 3).

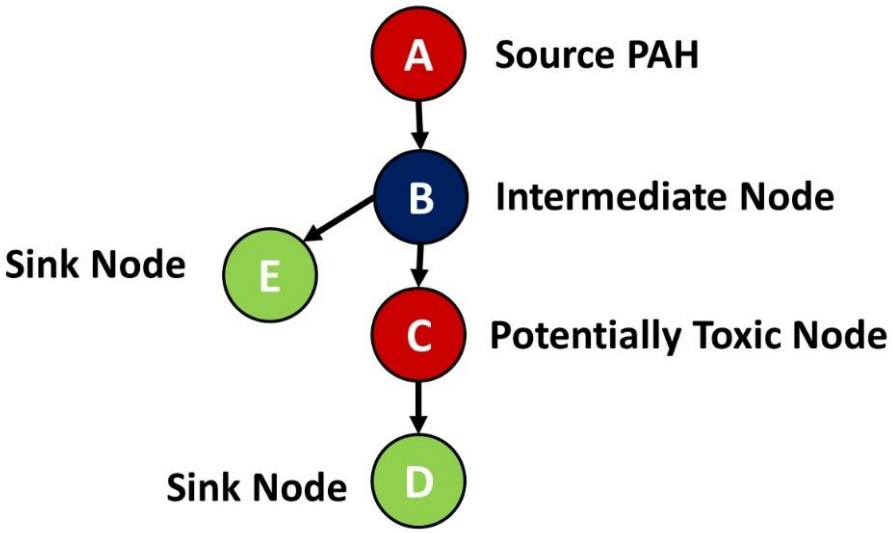


Figure 3: Relevant Pathway Filtering. Node A is one of the parent PAHs; acenaphthene, anthracene, fluorene, or phenanthrene. Its log Kow is greater than 2.0, so it is coded as a potentially toxic node. Node C, a metabolite, also has a log Kow greater than 2.0 and is considered a po tentially toxic node. Node B has a log Kow value less than 2.0, however, it leads to Node C, so it is considered an Intermediate Node and is included in the network. Nodes E and D have log Kow values less than 2.0, and have no daughter products that are greater than 2.0, so these are considered sinks and no compounds below them are included in the Relevant Pathway filtered network.

2.2.7 Empirical Literature Survey

A comprehensive literature survey was used to identify all available empirical studies on individual PAH degradation in a natural freshwater or soil aerobic environment. The following search terms were used to identify studies in the databases Scopus, Web of Science, and

Compendex: [anthracene, acenaphthene, fluorene, phenanthrene] AND [degradation, biodegradation, biotransformation] AND [metabolite, intermediate]. Studies that included deliberate augmentation such as surfactants or heat that would not be reflective of an environmental degradation process were omitted, as were studies that only used a mixture of PAHs without separating metabolites by their original source. Saltwater studies were also omitted as they would not be comparable with the EAWAG Pathway Prediction system. After this screening process, a total of 176 studies, some including multiple PAHs, were included for comparison with our modeled degradation networks. A detailed summary is provided in the SI. (see attached file Chap2_networks_analysis_SI.docx)

2.3 Results and Discussion

See attached files Chap2_networks_analysis_SI.docx and Chap2_networks_analysis_pathmatches.xlsx for Supporting Information

The batch process from EAWAG-PPS provides a data set of several thousand possible metabolite structures for each individual PAH. This data set is impractically large for the detailed analysis of each individual structure. However, by applying the techniques of networks analysis, concise information about nodes that have the highest likelihood of formation, or that occur in specific parts of the degradation network can be extracted. For clarity, nodes in the metabolite network are hereafter referred to as "metabolites" and edges are hereafter referred to as "biodegradation reactions". A summary of the numbers of metabolites (nodes) and biodegradation reactions (edges) in the predicted networks and effects of each method of filtering selected, as well as the number of empirical papers cited for each PAH is provided in Table 3.

Table 3 Summary of the number of nodes and edges in the predicted network and in literature. Relevant Pathways filtering was conducted, targeting nodes with log Kow > 2.0. High Node Throughput was defined as a Node Throughput value > 0.01

	Acenaphthene	Anthracene	Fluorene	Phenanthrene
Initial Predicted Nodes	6,104	26,694	30,759	22,996
Nodes after Relevant Pathway Filtering	3,591	8,365	16,451	14,651
Initial Predicted Edges	20,154	100,572	104,725	84,374
Edges after Relevant Pathway Filtering	9,190	23,311	52,385	40,544
High Node Throughput Compounds (> 0.01)	165	39	98	39
Empirical Literature Metabolites	14	46	46	88
Number of Empirical Literature Papers Cited	11	45	35	108

2.3.1 Network Level Correlations:

Correlations can be drawn between chemical characteristics and the node-level metrics betweenness centrality and Node Throughput. Relevant pathway filtering reduces the number of metabolites in the network by eliminating pathways containing metabolites that do not exhibit selected chemical characteristics, yielding a smaller network. The reduced network is more feasible to visualize or utilize for further analysis. The correlations between predicted chemical properties and the network metrics are very consistent across different PAH degradation networks. Some correlations with betweenness centrality are stronger if the unfiltered data are used; however,

relevant pathway filtering can be applied without substantially weakening the correlations for most of the chemical characteristics with Node Throughput. (Fig 4)

Correlations between the network algorithms and molecular descriptors can be used to prioritize compounds based on specific criteria. For example, log K_{ow} is strongly associated with narcotic toxicity, and certain patterns of aromatic hydroxyl groups are associated with carcinogenic toxicity. Log K_{ow} is weakly correlated with Node Throughput with an average Spearman correlation coefficient (r_s) of 0.21, and correlated more strongly with betweenness centrality with an average r_s of 0.40. Aromatic hydroxyl groups are correlated with betweenness centrality with an average r_s of 0.56. These aromatic hydroxyl groups are part of the structure of compounds that adduct onto DNA causing genotoxic and developmental impacts.[81,165] Carbonyl groups are associated with certain types of toxicity[166] and are negatively correlated with betweenness centrality with an average r_s of -0.42, suggesting that these compounds form closer to the edges of the network, or deeper in the degradation chain. These compounds may cause toxic impacts even after substantial degradation of PAH compounds have occurred. A discussion of the potential applications and strengths and weaknesses of the Node Throughput algorithm is provided in the sections below.

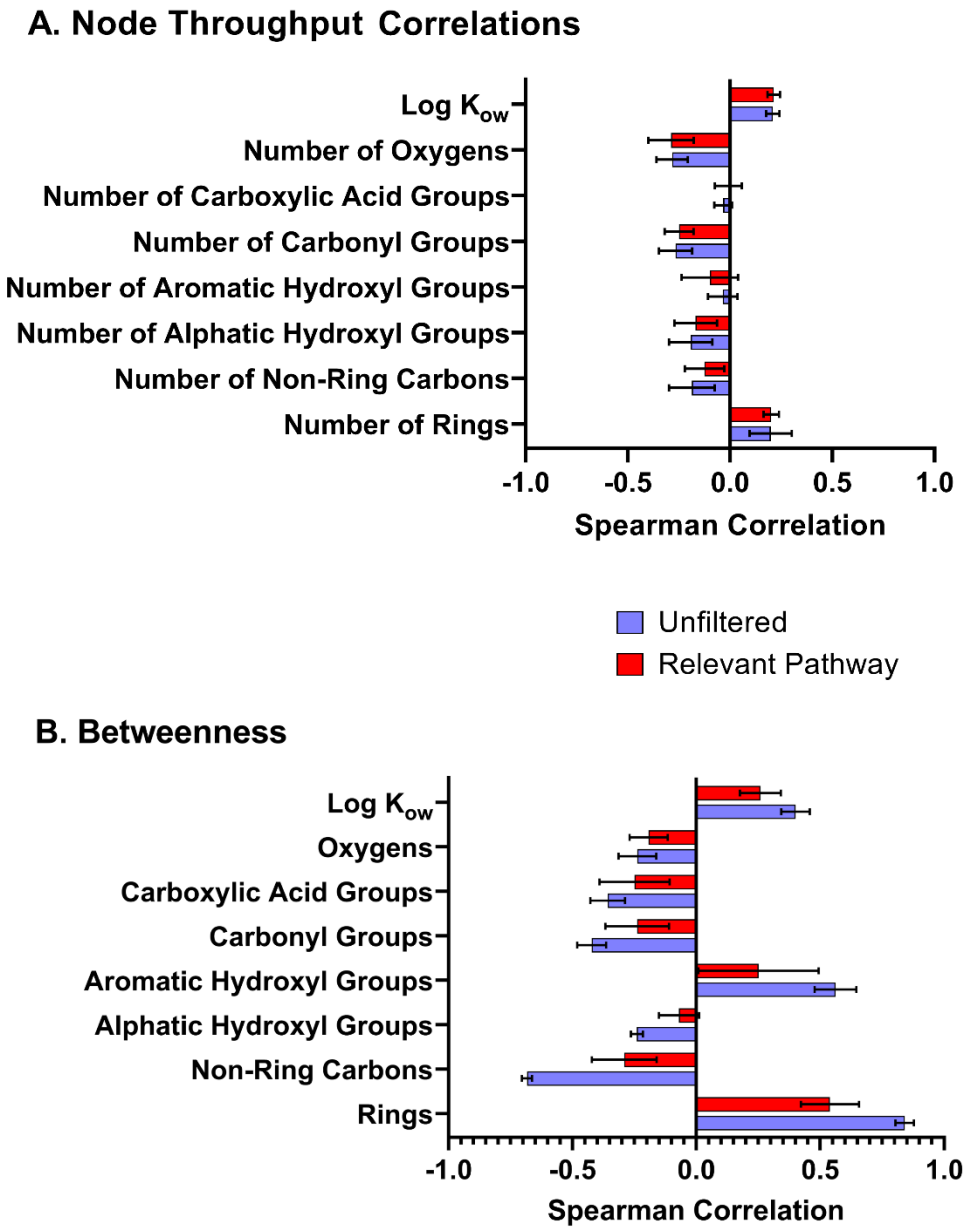


Figure 4 A. Mean Spearman correlation coefficients (r_s) between Node Throughput and molecular descriptors. B. Mean Spearman correlation coefficients (r_s) between betweenness centrality and molecular descriptors. 4 PAH degradation networks were considered and averaged: (acenapthene, anthracene, fluorene, and phenanthrene). Relevant Pathway Filtering was applied selecting nodes with log Kow > 2.0.

2.3.2 Betweenness Centrality Substructure Identification

Betweenness centrality reflects the compounds that occur on the greatest number of degradation pathways in the network. In this portion of the network, with high betweenness centrality, many predicted compounds exhibit similar substructures. In order to analyze the substructures in these compounds, the top 1% of compounds in each PAH degradation network with the highest betweenness centrality were selected. For clarity, these compounds will hereafter be referred to as high betweenness compounds. These compounds were clustered using Morgan Fingerprinting[160,167] as described in the SI (see attached file Chap2_networks_analysis_SI.docx), and the Maximum Common Substructure (MCS) for each cluster was then computed. The clusters represent groups of compounds with the same substructure. The tightness of the cluster can be adjusted as appropriate for the application; tighter clusters will have larger MCS's, but it will take more clusters to describe the high betweenness compounds. The MCS's for phenanthrene's five largest clusters are summarized in Figure 5.

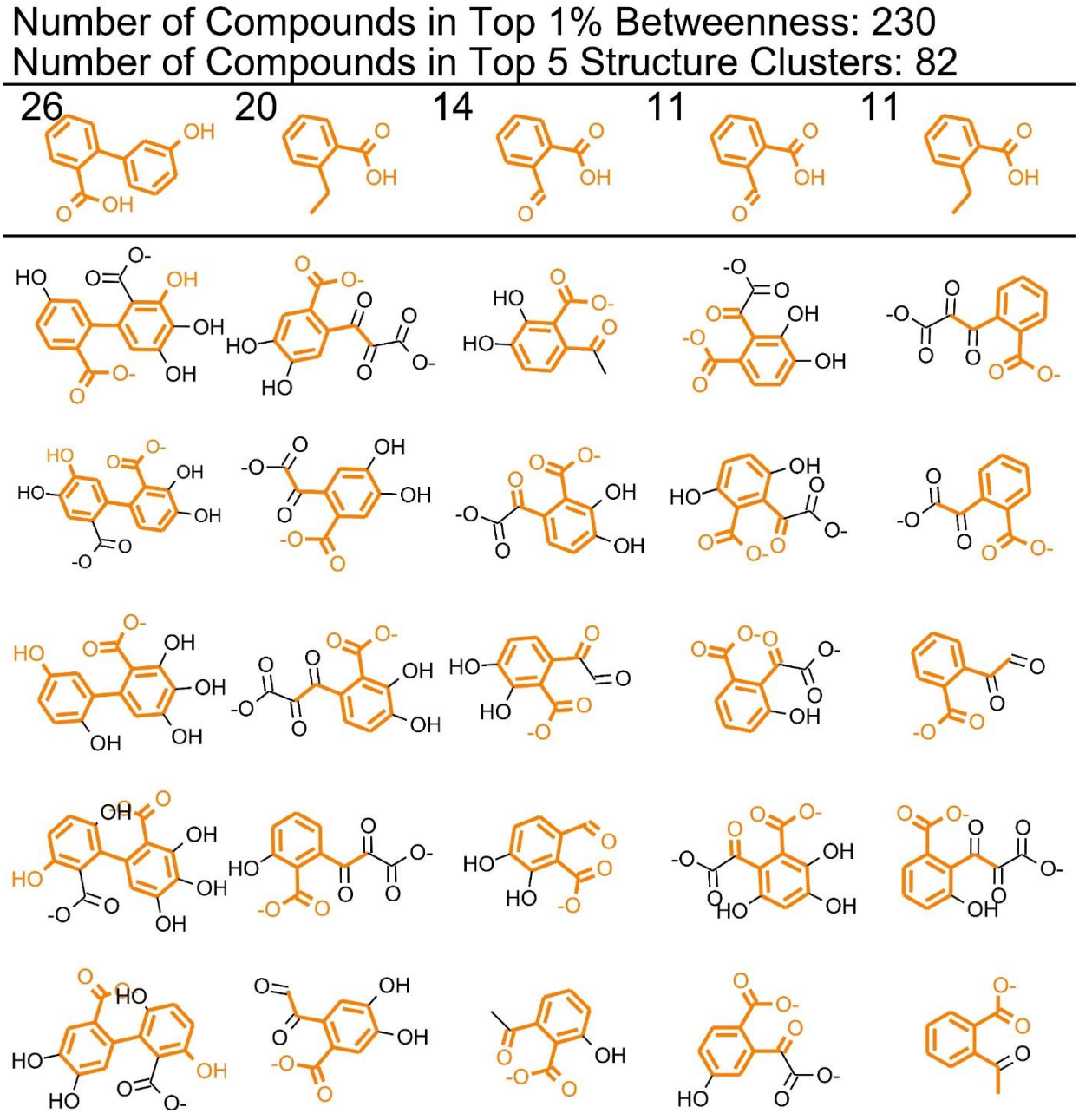


Figure 5 Maximum Common Substructures (MCS) in the top 1% of betweenness for phenanthrene. The first row shows the MCS in top 5 largest clusters. The numbers indicate how many compounds are contained in each cluster. The compounds in the vertical columns with the substructure highlighted in orange are presented as examples of which compounds are included in the cluster. In the case of phenanthrene, the same MCS occurs in several different clusters.

Locating common substructures is one possible way to predict the types of compounds that occur deep in the degradation network where low concentrations, intermittent occurrence, and brief persistence may make it more challenging to study the network with experimental methods. Additionally, high betweenness compounds may lie along key pathways that govern the rest of the network. If the primary degradation pathways can be highlighted by studying the high betweenness compounds, accelerating or restricting these pathways by providing specific nutrients, bacterial strains or enzymes could influence the overall decay chain. For example, Tian et al, 2003 found that salicylate affected the degradation kinetics of phenanthrene by *Pseudamonas Mendocina*,[168] and Sponza et al, 2012 suggested a biodegradation model including inhibition kinetics.[169] Optimizing the PAH degradation chain by deliberately managing the intermediate compounds could lead to less creation of toxic metabolites, or enable them to be removed more quickly. For example, a bioremediation professional should select an augmentation method that either avoids toxic intermediates, or continues the degradation pathway beyond them, and regulations should consider degradation metabolites.

2.3.3 Identification of Likely Empirical Compounds

The Node Throughput algorithm can be used to predict compounds that are likely to be detectable with instruments in an empirical study, as well as those that are on pathways to or from a detectable compound. We posit that nodes having a higher Node Throughput score have a greater likelihood of formation. A threshold value can be used as a cutoff to create a subset of "High Node Throughput" compounds, and construct a new network just from these compounds. The threshold could be set wherever desired; however, for this analysis, a threshold of 0.01 Node Throughput was found to perform well. It should be noted that the algorithm does not incorporate kinetic data

and therefore some High Node Throughput compounds may form that do not persist for very long. Figure 6 demonstrates this analysis for phenanthrene.

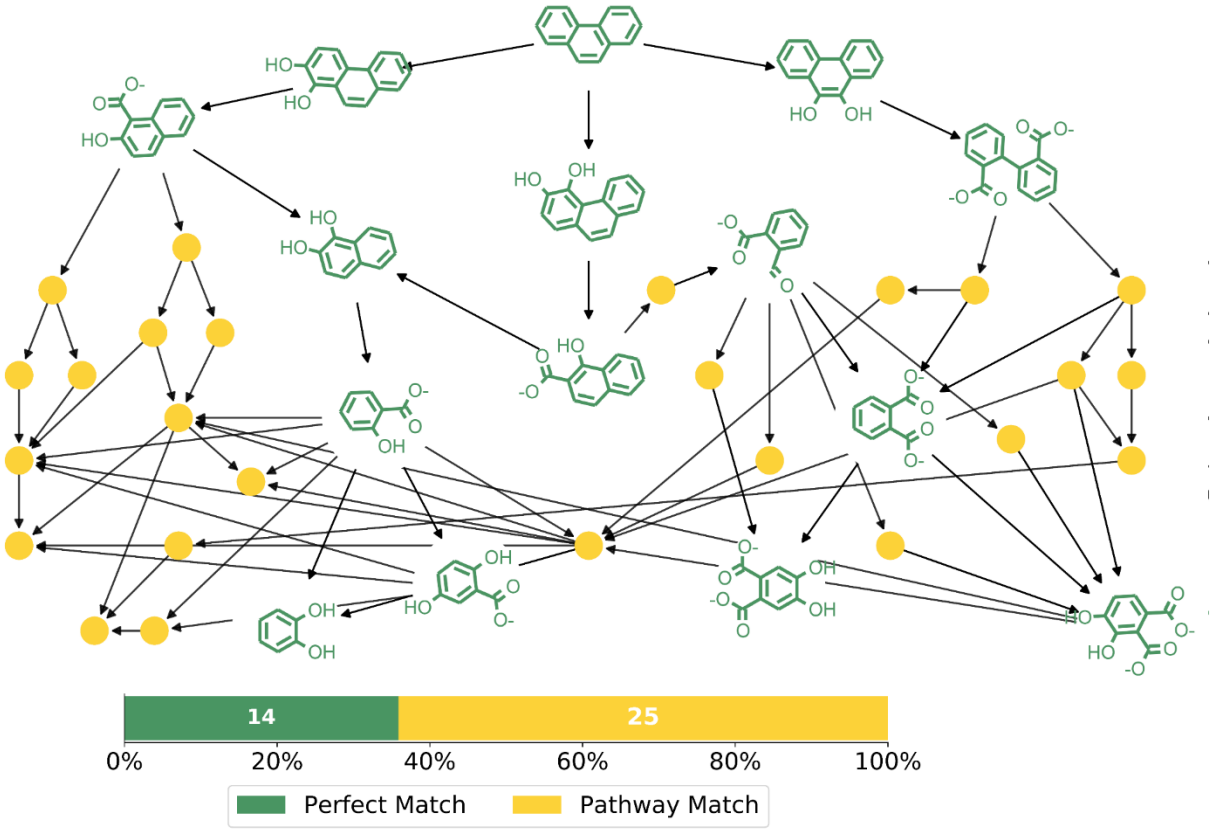


Figure 6 Phenanthrene Degradation Network High Throughput (> 0.01) predicted compounds for phenanthrene degradation compared against the available empirical literature data. "Perfect Match" compounds that match an empirical study exactly are shown as structures, and high Node Throughput predicted compounds that are on pathways to or from "Perfect Match" compounds are defined as "Pathway Match" compounds and are shown as yellow dots. The bar on the bottom of the figure shows the distributions of different types of matches. Comparisons for the other 3 PAHs and identification of the Pathway Matches are included in SI sections S1-S4. (see attached file Chap2_networks_analysis_SI.docx and Chap2_networks_analysis_pathmatches.xlsx)

More compounds are predicted by the high Node Throughput model than are found in the literature. This is anticipated because many possible intermediates on the high throughput pathways are included in the prediction. In the example shown in Figure 6, all predicted metabolites selected by the high Node Throughput filter were either exactly matched or included in high throughput pathways to or from matched compounds. Although some compounds leading to and from the experimentally identified compounds may not be easily detected with experimental approaches, they may still occur and cause toxic effects in organisms exposed to them. These additional compounds represent possible targets for analysis in future experiments to further illuminate PAH degradation metabolites. This process could be applied to less studied PAHs, providing a data set of the most likely structures and the links between them.

To assess the completeness of the High Node Throughput filtering method, a thorough literature review was conducted as described in the Methods to identify all available metabolites found by empirical studies. The comparison between the metabolites from empirical literature and the compounds identified by High Node Throughput filtering is summarized in Figure 7.

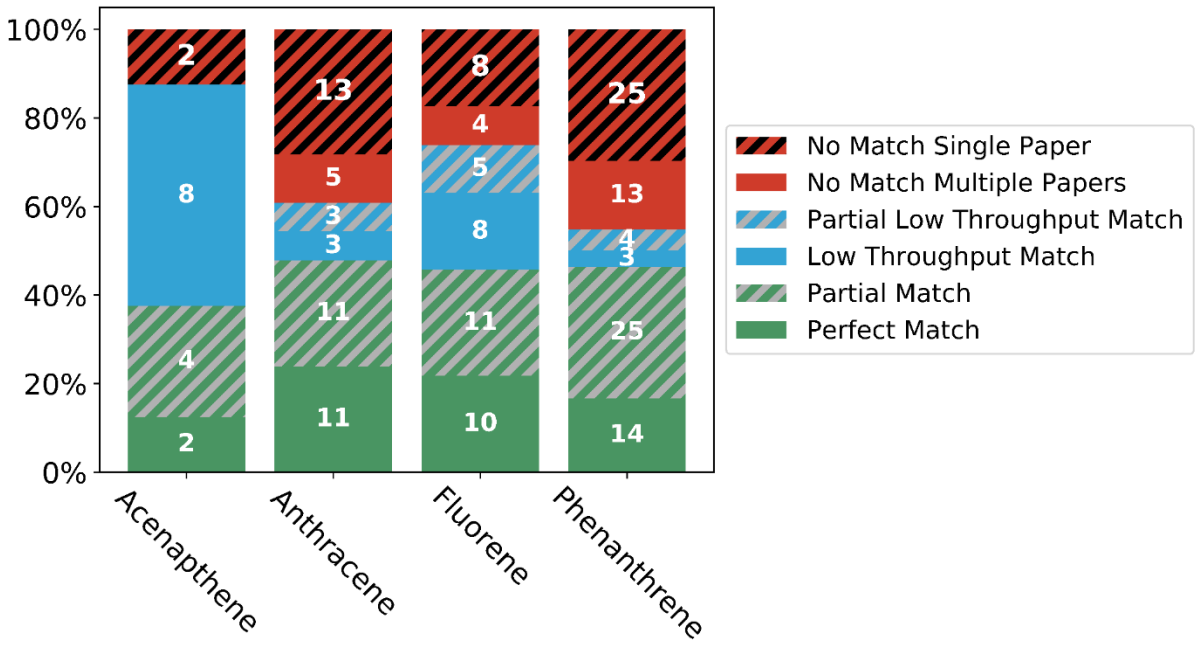


Figure 7 Analysis of whether the empirical compounds identified in literature are predicted by the High Node Throughput network algorithm. Perfect Match: The predicted structure and the literature structure match exactly. Partial Match: One of the following conditions is true: 1. The compounds have the same carbon backbone, including aromaticity of bonds. The oxygenation groups may differ. 2. The predicted network contains an exact match for both the immediate precursor and daughter product of a compound from literature. Low Throughput Match/Partial Match: The same criteria as Perfect/Partial Match is applied, but the predicted Throughput is less than 0.01. No Match: The "No Match" category is assigned if none of the other categories can be assigned. No Match compounds are separated into compounds found by only one paper and those found by multiple papers.

If Partial Matches are included, 38% to 48% of the compounds discovered by empirical studies in the literature are also predicted by the High Node Throughput algorithm. If the compounds that are only represented by a single paper are removed, this improves to 42% to 66%. In many cases there are reasonable explanations for why compounds are not matched due

to the EAWAG-PPS rules. For example, EAWAG-PPS does not predict methylations, and out-and-back pathways are not predicted. Common reasons for a "No Match" classification are included in the SI.

Uncertainty in the network arises from the broad spread of possible compounds predicted. It is expected that not all of the predicted compounds have a realistic environmental impact. The EAWAG aerobic likelihood categories used in this study were limited to three levels[60] and their relative weight to each other can only be estimated. To account for this uncertainty, the SI contains detailed sensitivity analysis showing that even with a significant amount of variability in the network weights, the Highest Node Throughput compounds are very consistent and the literature matches are found with very high fidelity.

2.3.4 Environmental Implications

Hazard assessment for intermediate degradation compounds is a challenging problem. Network analysis can assist in highlighting desired portions of the complicated and often undetectable degradation network. As the understanding of the toxic mechanisms of these compounds improves, the correlations between the chemical property data can be used to predict the portions of the degradation network that may contribute to specific toxic impacts. Compounds that show high betweenness can be analyzed for substructures, indicating the types of compounds likely to form during degradation. The novel High Node Throughput algorithm can predict many of the same compounds identified over the last 30 years of experimental research into PAH metabolites, as well as the compounds that flow to and from them. Furthermore, the comparison of network metrics with chemical property data suggests that the relationships both with betweenness centrality and with Node Throughput are consistent regardless of which PAH is

modeled, so this method may provide a good tool for estimating the degradation pathways of less studied PAHs or for studies to target intermediate metabolites of previously studied PAHs.[170]

3.0 Direct-Acting Mutagenicity Quantitative Structural Activity Relationship (QSAR)

3.1 Introduction

Reproduced with permission from Chemical Research in Toxicology, Sleight, T.; Sexton, C.; Mpourmpakis, I.; Gilbertson, L.; Ng, C. A Classification Model to Identify Direct-Acting Mutagenic Polycyclic Aromatic Hydrocarbon Transformation Products. Chem. Res. Toxicol. Rev. 2021. Copyright 2021, American Chemical Society (in review)

Polycyclic Aromatic Hydrocarbons (PAHs) contain at least two aromatic rings and are the byproduct of both natural and industrial pyrogenic processes, including forest fires, extraction and burning of fossil fuels, plastic manufacturing and municipal waste incineration.[171,172] Consequently, PAHs can be found in all environmental compartments - the atmosphere, soil and water.[173] Atmospheric deposition of PAHs occurs within a few days of emission,[38,121] and some PAHs can persist for up to several months on surface soils or in water as they slowly degrade.[15,16] While in the aerobic region of soil and/or water, PAHs can form a multitude of different transformation products (TPs), many of which are hazardous.[174,175]

The U.S. Environmental Protection Agency (EPA) currently classifies 16 PAHs as priority pollutants.[176] While these compounds have been well studied, there is limited research on the hazards of TPs that result from PAH biodegradation, which occurs naturally in the environment and can also be induced with deliberate bioremediation at contaminated sites. While biodegradation typically eliminates the original unsubstituted PAHs, it does not necessarily eliminate the health risk due to their toxic TPs.[177–179] Numerous studies have noted that PAH TPs

exhibit greater developmental toxicity, mutagenicity or genotoxicity than the original PAHs, including PAHs of lower molecular weight, containing three rings or less.[20,36,89,94,98,126,134,177,180–182]

3.2 PAH Toxic Impacts Background

Narcosis is the baseline mechanism of PAH toxicity and typically correlates well with a compound's octanol-water partition coefficient (K_{ow}).[129–131] Many toxicity assessment tools, including the Ecosar module of the U.S. EPA's Estimation Programs Interface (EPI) Suite,[183] rely heavily on K_{ow} and provide adequate prediction of narcotic toxicity.[75–77] However, as compounds are degraded, numerous transitions occur including oxygenation, hydroxylation and ring cleavage, which can impact the TP's potential toxicity.[184] PAH degradation intermediates are smaller than their parent compounds and possess polar oxygen groups, lowering their K_{ow}. However the addition of polar oxygen groups facilitates other toxicity mechanisms, particularly mutagenicity.[185,186] Since the mutagenicity of biodegraded TPs is not directly related to K_{ow}, QSARs based on K_{ow} do not effectively predict mutagenicity. Therefore, a different approach is needed for assessing this component of PAH toxicity.

The recognition of a relationship between a molecule's structure and its toxic properties is almost as old as molecular theory. In his thesis in 1893,[109] Cros noted a relationship between the lipophilicity of primary aliphatic alcohols and their toxicity, a principle which remains useful to this day. As molecular representations and numerical analysis methods have improved, quantitative structural activity relationships (QSARs) have gained increased acceptance as screening and prediction tools for physical properties and biological activity.[187,188] In order to

predict the specific toxic properties of a molecule, empirical training data is needed. Most publicly available or commercially licensed QSARs for mutagenicity are trained on Ames test data,[189–191] the gold standard bioassay for mutagenicity.[115,189,192] The test involves culturing bacterial strains with specific reverse mutations in their histidine (an essential amino acid) coding gene on histidine deficient media. When exposed to a mutagen, the reverse mutation reverts to restore the histidine gene, allowing the bacteria to grow, thus indicating the presence of a mutagen. The test typically includes a set of replicates spiked with rat liver extract (S9 fraction), to reflect an organism's metabolism. This is designed to evaluate the formation of mutagens upon biotransformation within organisms higher on the food chain that might be exposed to PAHs through ingestion or inhalation. However, direct-acting mutagens such as those created through biodegradation may be difficult to detect if S9 activation is used. Publicly available ready-to-use Ames test QSAR programs are typically trained on data both with and without S9 activation and consequently are not suitable for evaluating the direct-acting mutagenicity that is induced through biodegradation (that is, occurring without any S9 activation).[193] Additionally, several more specific theoretical QSAR relationships have been developed, many focusing on select nitrogenated PAHs which occur in the atmosphere (Fig 8).[194–198] Although this body of work aids in evaluating the mutagenicity of nitrogenated PAHs, which can occur in the atmosphere, there is a lack of similar evaluation of the environmental transformation products resulting from the terrestrial biodegradation of PAHs, which are primarily oxygenated derivatives.[36,37,98]

Due to the narrow applicability of the available theoretical QSAR relationships designed specifically for PAHs, and the lack of specificity of public QSAR programs, there is a gap in assessment methods for mutagenic PAHs TPs activated by environmental processes.[18–21] The 16 PAHs recognized as priority pollutants by the U.S. EPA are commonly used as a starting point for

PAH studies, but do not contain any nitro-PAHs or oxy-PAHs, which are common among intermediate TPs. Thus, a new tool is needed that is both able to cover a wide range of different sizes of potential compounds and is trained only on data that does not use S9 activation. In this study, we develop such a QSAR that is able to reflect the mutagenic potential of molecules representative of potential PAH TPs from aerobic biodegradation, with higher performance than publicly available QSAR tools and with broader applicability than previously published theoretical relationships. The methods used allow the resulting characteristics of PAH TPs to be connected with known pathways of PAH mutagenicity for a mechanistic interpretation.

3.3 Experimental Procedures

3.3.1 Training Data

Data for Ames test *Salmonella typhimurium* strains TA98 (frameshift mutation) and TA100 (base-pair substitution mutation) were selected, as they are commonly used in studies focused on environmental exposure, and are specified in ISO 11350 for the genotoxicity of water and wastewater.[203] These strains have shown good responsiveness to the oxy-PAH mutagens that are generated from PAHs by environmental degradation processes (Table 4). Most of the training data were acquired from the Chemical Carcinogenesis Research Information System (CCRIS).[204] The data were limited to hydrocarbons that only contain carbon, oxygen, and hydrogen, weighing less than 500 amu and containing between 1 and 5 aromatic rings, in order to be representative of environmentally relevant PAH TPs. Twelve additional molecules were obtained from the Organization for Economic Co-operation and Development (OECD) eChemPortal.[205] Other data

sources[206–208] were reviewed but it was found that these sources did not discriminate between positive Ames test results that used S9 activation and those for direct-acting mutagens (which are mutagenic without metabolic activation).

3.3.2 Molecular Descriptors

A search for existing studies on QSAR development for environmental PAH TPs uncovered 7 studies that contained QSAR relationships for specific sub-categories of PAHs (Fig 8, Table 4).[196,199,200,202,209–211] These studies were used to inform the types of descriptors that were likely to be helpful in predicting PAH mutagenicity. The Highest Occupied Molecular Orbital and Lowest Unoccupied Molecular Orbital energies (ϵ-$_{HOMO}$ and ϵ-$_{LUMO}$, respectively), the HOMO-LUMO gap, the ionization potential, and electron affinity were calculated using electronic structure calculations in Gaussian 09.[212] Density Functional Theory calculations were applied using the B3LYP[213] hybrid functional and the 6-311g(d,p) basis set.[214] This method and basis set were selected due to good performance in calculating electronic properties.[215–217] All molecules studied were fully optimized to their ground state geometry in neutral charge state. From the relaxed structures, vertical ionization potentials and electron affinities were calculated with single-point energy calculations (in positive and negative charge states, respectively). Pharmaceutical Data Exploration Laboratory (PaDEL)-Descriptor[218] was used to calculate topological descriptors for the Ames test data sets. The coordinates from the DFT optimization were supplied to PaDEL-Descriptor for use in the calculation of 3-dimensional descriptors. PaDEL has the capability to calculate up to 1875 unique descriptors. Of these, approximately 1500 descriptors were applicable to our dataset with the others reflecting atoms or structures that not relevant to PAHs. Many of PaDEL's descriptors contain weightings that reflect different types of atoms (Table S4.1). All of

these are set for each atom, except for the weighting for intrinsic state, abbreviated "s", and charge, abbreviated "c". An atom's intrinsic state is the ratio of the atom's π bonds and lone pair electrons over the number of σ bonds for the atom considered. Atoms with more single bonds will have a lower intrinsic state, relative to the same atom with double bonds, and most oxygen atoms will have a higher intrinsic state than most carbon atoms (Table S4.2). In order to explore the overlap between our dataset and existing studies, PaDEL structural descriptors[218] were computed for each of the datasets used in these studies and principal component analysis (PCA) was used to visualize the feature space of each dataset as well as the dataset used for this study.

3.3.3 Analysis

The descriptors were standardized by standard deviation, and recursive feature elimination was applied to select the most relevant features. Logistic Regression (LR) was used to develop a model for distinguishing between mutagens and non-mutagens across the entire dataset, considering mutagens detected by either the TA98 or the TA100 strain. Previous studies suggest that there are multiple possible mechanisms by which small hydrocarbons may exhibit mutagenicity;[99,219–224] thus, we anticipate that there may be multiple mutagenic mechanisms within our dataset. PCA clustering was completed without considering whether the data was positive or negative for mutagenicity. We hypothesized that structural subclasses with similar mechanisms could be uncovered using unsupervised clustering methods to categorize the data before attempting to extract the descriptors that corresponded to mutagenicity within each cluster.

Established classification metrics were applied for evaluating the performance of the classifier.[225] Mutagenic compounds are considered positive data for the classifier metrics. Accuracy is the overall fraction of correct classifications of either class. Precision is the proportion

of mutagenic classifications that were correct. Recall/Sensitivity (terms used interchangeably) is the proportion of the total number of true mutagens that were correctly classified. Specificity is the proportion of the total number of non-mutagens that were correctly classified. The publicly available QSARs are expected to have low precision and specificity for this dataset as they are trained on data that includes S9 activation. F1 is the harmonic mean of precision and recall. F1 reflects errors in both forms of classification, so it is the most comprehensive single overall metric for evaluating a classifier's performance.

In each intra-cluster regression analysis, 10 iterations of k-fold validation using 1/3 of the data for testing were conducted, sweeping through different numbers of features to determine the ideal number to select. The number of descriptors to use was determined by incrementally adding descriptors and calculating the weighted F1 scores. Once performance stopped improving, no further descriptors were added. Although recursive feature elimination is an effective method of selecting the most impactful descriptors, it does not eliminate correlated descriptors that provide the same information. A variance inflation factor (VIF) of < 5 was used to iteratively remove highly correlated, redundant descriptors[226,227]. If the VIF was within 0.5, an assigned preference was used as detailed in Table S2.16. Full details including correlated descriptors and example structures are provided in the SI and the code used to extract the data from the CCRIS and OECD repositories, efficiently calculate the PaDEL descriptors, and perform the analysis is available at https://github.com/twsleight/Environmental_PAH_Mutagenicity.

3.4 Results

See attached file Chap3_classification_model_SI.docx for Supporting Information and additional data

Based on PCA analysis of the PaDEL descriptors for each dataset, none of the existing individual studies covered the entire feature space that we wished to consider (Fig 8).

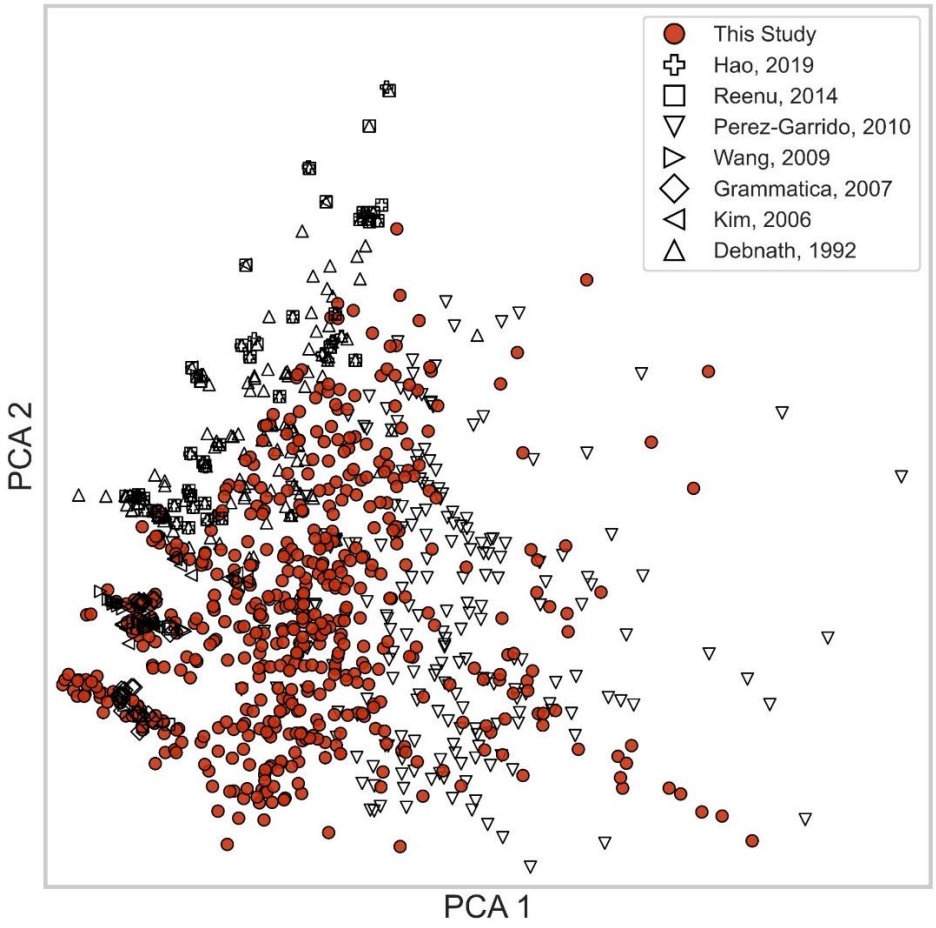


Figure 8 Principal Component Analysis (PCA) plot of the features of the available literature QSAR relationships for PAHs. The PCAs were calculated with the PaDEL topological descriptors to visualize the feature space of the different descriptors.

Table 4 Existing QSARs that apply to specific PAH derivaties

Descriptors	**Training Data Type and Size**	**Target Assay**	**Study**
ϵ-$_{LUMO}$, K_{ow}	Nitro-PAHs, 162	TA98/TA100 (w/o S9)	Debnath et al 1992[196]
Gs (molecular symmetry) Rm5+ (max autocorrelation of Atoms)	Nitro-PAHs, 48	TA98/TA100, (w/o S9)	Gramatica et al, 2007[202]
ϵ-$_{HOMO}$ Total Information Content Index, 2D structure-based atom-pair descriptor representing topological information, Hypnotic-80* Infective-50**	Nitro-PAHs, 48	TA100, (w/o S9)	Hao et al, 2019[209]
Electron-Correlation Energy (E_{CORR}) Correlation Contribution of Electrophilicity Index (ω_{CORR})	Nitro-PAHs, 51	TA100, (w/o S9)	Reenu[210]
q_{+H} (indicates hydrogen bonding)	OPAHs 2-6 rings, 26	DNA Binding	Wang et al, 2009[199]
Substructure alert	α, β Unsaturated Carbonyl Compounds, 235	Multiple Ames Tests (w & w/o S9)	Perez-Garrido et al, 2010[200]

Table 4 (Continued)

ϵ-$_{LUMO}$, and several charge descriptors	Benz[a]anthracene Derivatives, 29	Myco-bacterium strains	Kim et al, 2006[211]
Charge and Structural Descriptors	Oxy-PAHs w/1-5 aromatic rings. ~550	TA98, TA100, (w/o S9)	This study

*Ghose-Viswanadhan-Wendoloski hypnotic-like at 80% (drug-like index)[228]

**Ghose-Viswanadhan-Wendoloski anti-infective at 50% (drug-like index)[228]

The descriptors used by the studies in Figure 8 (Table 4) were used to inform the descriptors explored in this study. Several studies indicate that both the ϵ-$_{HOMO}$ and the ϵ-$_{LUMO}$ energies can be relevant descriptors. Additionally, there are several descriptors that reflect specific molecular shapes or charge distributions. We are focusing on oxygenated PAHs rather than nitrogenated PAHs, in order to reflect the most likely TPs generated in surface water and soil. However, the mechanisms of mutagenic toxicity are likely to be similar. Furthermore, as Table 4 illustrates, PAHs of different sizes and with different functional groups may require different descriptors to evaluate their mutagenic potential.

The ability of existing publicly available QSAR programs to classify direct-acting PAH mutagens was evaluated with our dataset (Table 5). The other common Ames test strains of TA97, TA102, TA1535, TA1537, and WP2 were also considered for additional positive data to evaluate the publicly available QSARs, but no additional positives were found.[229]

Table 5 Classifier metrics of publicly available Ames QSARs vs the QSAR developed in this work, evaluated with our dataset, which contains unsubstituted and oxygenated hydrocarbons.

	Vega[191]	T.E.S.T[190]	This Study All Data	This Study Small Molecules	This Study Large Molecules
Accuracy	0.80	.73	0.82	0.91	0.80
Precision	0.56	.48	0.60	0.75	0.63
Recall/Sensitivity	0.95	.85	0.84	0.86	0.84
Specificity	0.75	.69	0.81	0.93	0.78
F1	0.70	0.61	0.70	0.80	0.72

Training our entire dataset of 557 molecules (141 mutagens) with LR yields a Receiver Operating Curve (ROC) with an Area Under the Curve (AUC) of 0.90. The F1 score equals or exceeds those of both the Vega-HUB and EPA Toxicity Estimation Software Tool (T.E.S.T) models (Table 5). Forty-Eight descriptors are used in this model. The ROC and classifier metrics are developed using 10 rounds of 3-fold cross validation, using 2/3 of the data for training and 1/3 for testing, randomizing the data between each round. The final regression coefficients are obtained by regressing across the entire dataset and the selection threshold is set at 0.5 in order to be consistent with the other QSARs. The simplicity of LR made it a good candidate to connect the descriptors to specific mechanisms.[230] Regression coefficients from the scaled LR input data provide a simple and easily understood reflection of a descriptor's predictive strength for or against mutagenicity. Several other methods including k-nearest neighbors, decision tree, support vector machines and random forest were considered. However, these methods are more difficult to interpret due to the need to tune additional hyperparameters.

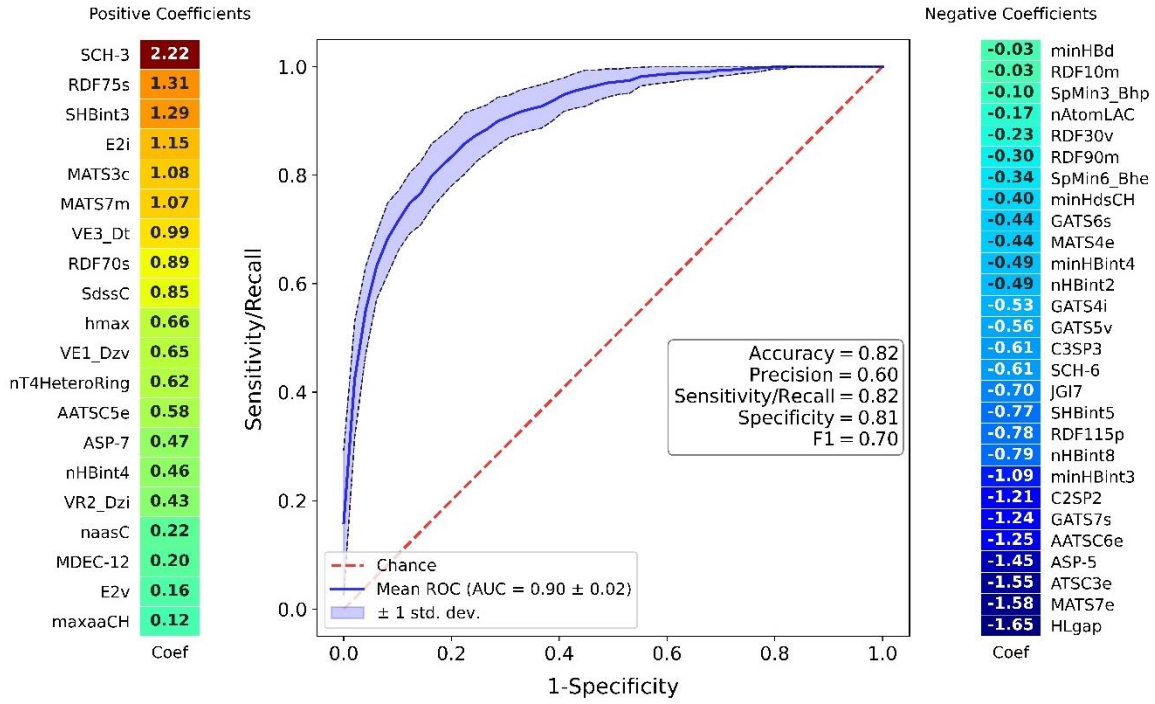


Figure 9 Receiver operating (ROC) with logistic regression for the entire dataset. The two descriptors with the two largest positive and negative coefficients are the simple chain, order 3 (SCH-3) which tends to highlight epoxide groups and HOMO-LUMO gap, which has lower values associated with mutagenicity. See S-2 for details of other descriptors.

Table 6 Confuction matrix for 10 rounds of cross validation for the entire dataset

	True Mutagens	True Non-Mutagens
Predicted Mutagens	(True Positives) 1167	(False Positives) 762
Predicted Non-Mutagens	(False Negatives) 243	(True Negatives) 3398

Although the performance of the classifier using all of the data across the entire dataset has an equal or higher F1 score than the publicly available QSARs (Table 5), clustering the available data by the principal components of the descriptors allows us to achieve a higher level of

performance with fewer descriptors. It is reasonable to anticipate that a broad study of hydrocarbon mutagenicity such as this one would include data reflecting multiple mutagenic mechanisms.

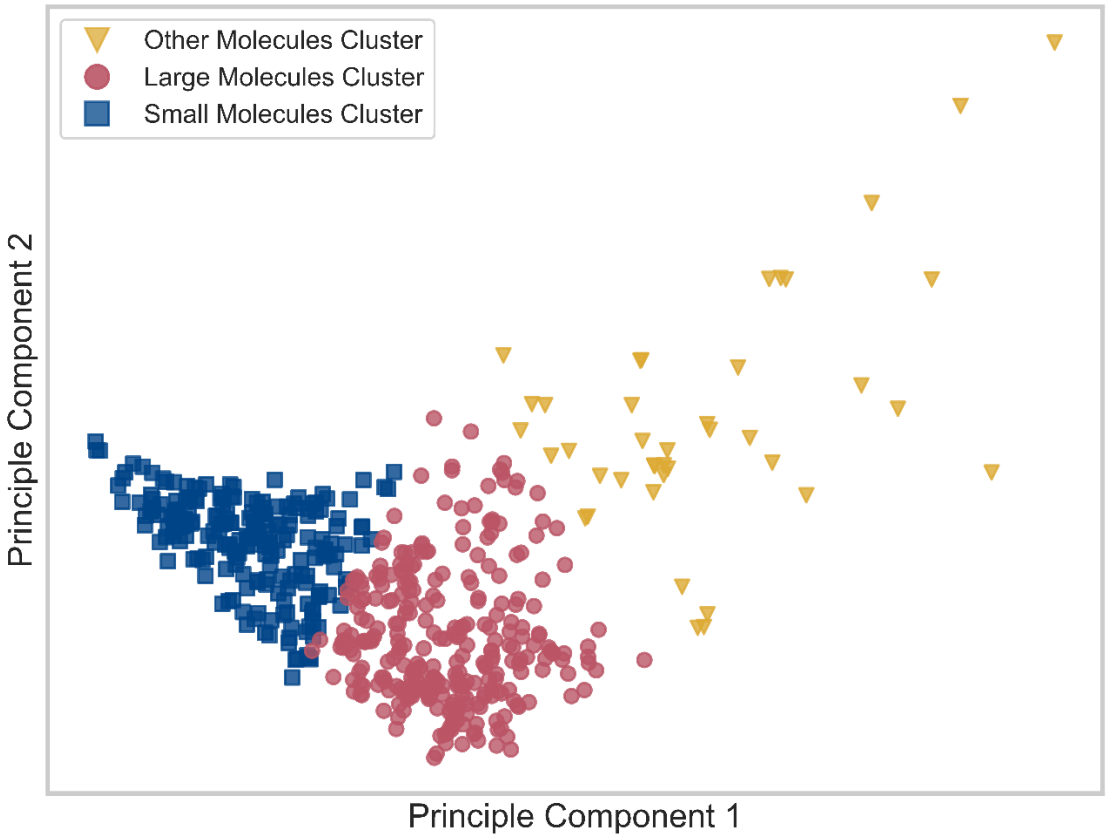


Figure 10 Principal Component (PC) plot of the dataset showing the results of k-means (k=3) clustering.

Three clusters, based on the first five principal components was found to be the ideal number to maximize the silhouette plot (Fig S1.3) and minimize the Davies-Bouldin Plot (Fig S1.4), indicating that this is the ideal number of clusters to describe this dataset. Within each cluster, specific descriptors emerge which allow a better characterization of the mutagenic compounds than if the dataset is considered as a whole. The two major clusters are still larger than most of the studies in Table 4 and classification scores exceeding those by existing QSARs can be

obtained with this method. There are 301 molecules in the cluster of "large molecules". This cluster has an average molecular weight of 254 amu, with an average of 32 atoms per molecule and most molecules have 2 or more aromatic rings. The "small molecules" cluster contains 215 molecules. Most of these molecules have 1 ring, with a few 2 or 3 rings structures. The average molecular weight is 153 amu and there is an average of 21 atoms in each molecule. See Figures S6.1-S6.4 for representative structures of both clusters. The "other molecules" cluster was not analyzed due to its complicated molecular structures and the small size of the cluster. Aliphatic 6-membered rings were much more common in this cluster than in the other two clusters, and these compounds were not structurally similar to common environmental TPs of PAHs (Fig S6.5).

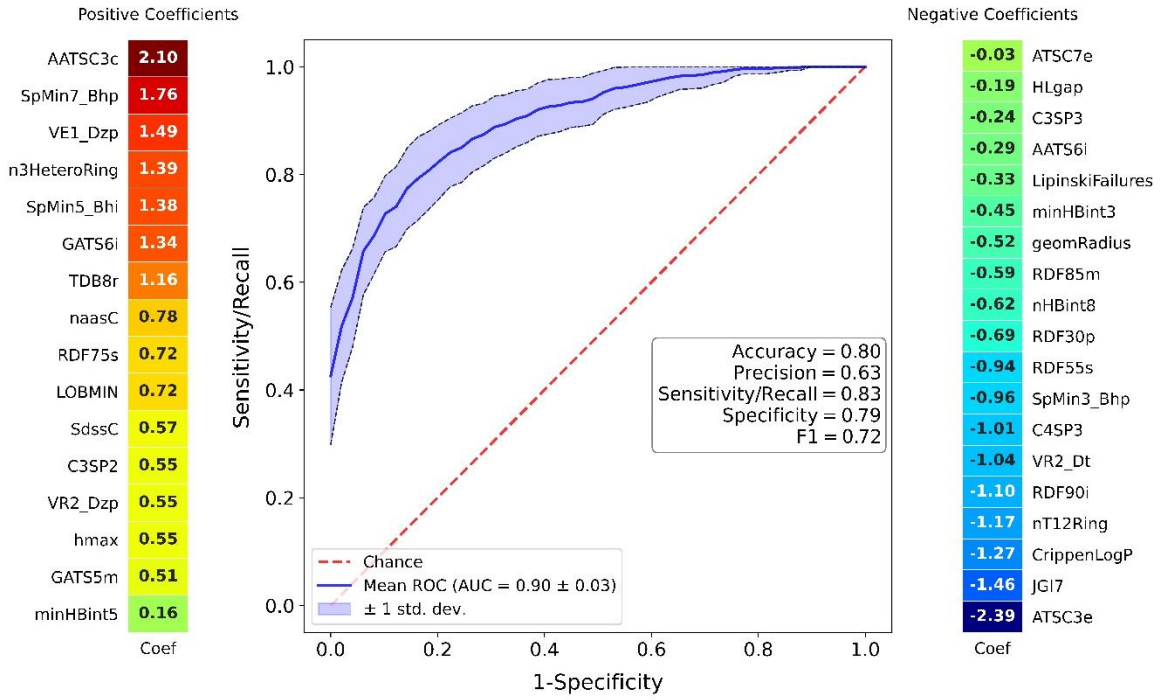


Figure 11 Large molecules cluster (blue in Figure 10, 301 molecules, 92 mutagens) Logistic regression receiver operating characteristic with classifier metrics and coefficients for the larger molecules in the dataset.

Table 7 Confuction matrix for 10 rounds of cross validation for the large molecules cluster

	True Mutagens	True Non-Mutagens
Predicted Mutagens	(True Positives) 768	(False Positives) 447
Predicted Non-Mutagens	(False Negatives) 152	(True Negatives) 1643

In the large molecules cluster, a ROC curve with an AUC of 0.90 can be achieved with thirty-five descriptors. All descriptors have a VIF of < 5 and the maximum absolute correlation between any two descriptors is 0.76. This results in a classifier that achieves an F1 score of 0.72 (Fig 11, Tab 7), indicating a performance improvement by regressing within this cluster as opposed across the entire dataset, as well as reducing the number of descriptors used.

In the small molecules cluster, a ROC curve with an AUC of 0.96 is achieved with twenty-five descriptors, resulting in a classifier with an F1 score of 0.80. (Fig 12, Tab 8) The highest Spearman correlation coefficient between any two descriptors is 0.94, between two of the molecular edge descriptors.

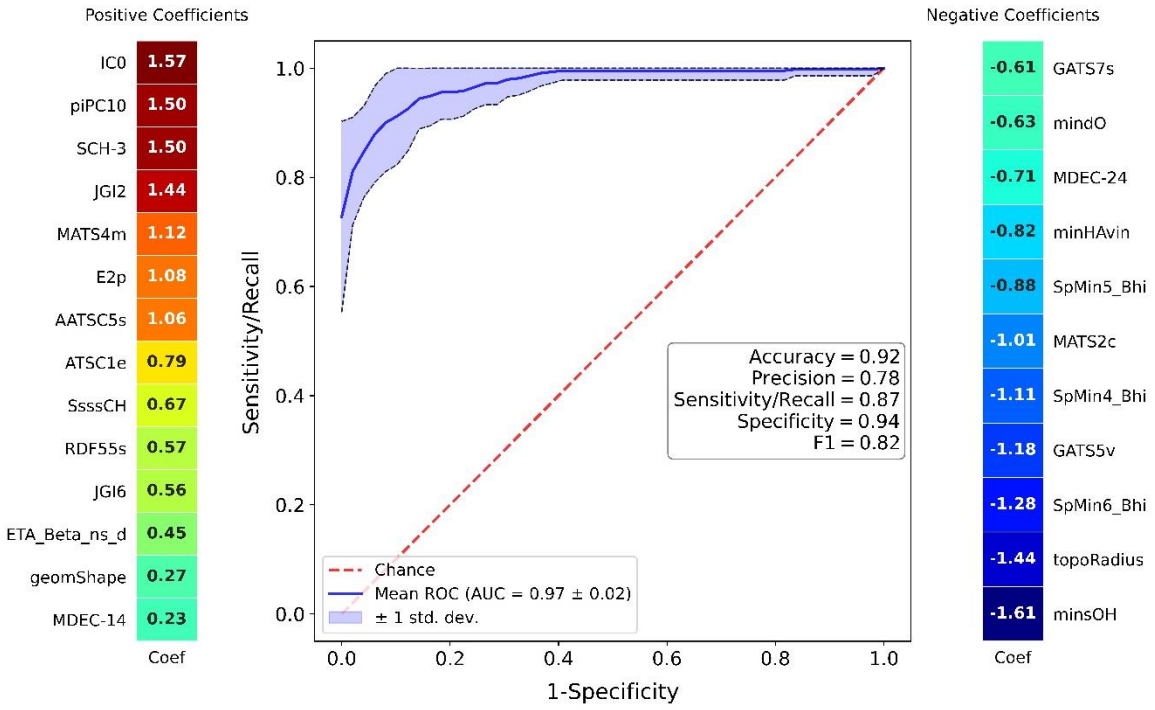


Figure 12 Small molecules cluster (rose in Figure 10, 215 molecules, 43 mutagens) Logistic regression receiver operating characteristic with classifier metrics and coefficients for the small molecules in the dataset.

Table 8 Confuction matrix for 10 rounds of cross validation for the small molecules cluster

	True Mutagens	True Non-Mutagens
Predicted Mutagens	(True Positives) 376	(False Positives) 108
Predicted Non-Mutagens	(False Negatives) 54	(True Negatives) 1612

Further details of the descriptors with examples as well as information about which descriptors were removed by the VIF function due to correlation with other descriptors are provided in Figures S2.1-S2.15. PaDEL-Descriptor's documentation includes a summary of all possible descriptors that can be computed by PaDEL including descriptors not selected for this analysis. Further information can be obtained from Todeschini and Consonni, 2009.[228]

3.5 Discussion

3.5.1 Positive coefficients in the large molecule cluster

Since the dataset is only comprised of compounds containing only carbon, oxygen, and hydrogen atoms, most descriptors correlate with specific atomic substructures. Some substructures can be connected to specific known mutagenic pathways and certain substructures emerge as likely non-mutagenic structures.

The average centered Broto-Moreau autocorrelation - lag 3, weighted by charges (AATSC3c) has the largest positive coefficient in the large molecules cluster. Lag refers to how close the atoms are together in the molecular structure. The descriptor reflects oxygen groups 3 atoms apart. It is balanced against the strongest negative descriptor, ATSC3e, the centered Broto-Moreau autocorrelation, lag 3, weighted by Sanderson electronegativities, which is more associated with non-mutagens. The primary differences between the two descriptors are that AATSC3c is an average over the entire molecule, and is weighted by charges, while ATSC3e is weighted by Sanderson electronegativities. AATSC3c tends to select molecules with ortho-oxygen groups, which can be mutagenic[222,231–233] and ATSC3e tends to select any pattern of where there are multiple sets of oxygen groups 3 spaces apart, which may be mutagenic or not.

There are two descriptors based on eigenvalues of the Burden modified matrix. SpMin7_Bhp is the smallest 7^{th} absolute eigenvalue of Burden modified matrix, weighted by relative polarizabilities. SpMin5_Bhi is the smallest 5^{th} absolute eigenvalue, weighted by relative first ionization potential. Both of these descriptors tend to favor larger molecules, but the relative polarizability weighting of SpMin7_Bhp favors molecules with oxygen groups on part of the perimeter of the molecule. VE1_Dzp is the sum of the coefficients of the last eigenvector from

Barysz matrix, weighted by polarizabilities. This sum will be higher for asymmetric molecules with oxygens on one side of the molecule, which results in a higher overall sum due to the coefficients being different. Another strong positive coefficient, GATS6i, refers to the Geary autocorrelation lag 6, weighted by the ionization potential (abbreviated "i", Table S4.1) of each individual atom. This Geary autocorrelation number highlights molecules with a different atom 6 positions away in the molecular structure, as carbon and oxygen have different "i" weighting values. This suggests that mutagens must have oxygen groups on one portion of the molecule, but not all over it, and this descriptor selects such dissimilar molecules.

Several descriptors indicate structures that are recognized indicators of mutagenicity. TDB8r is the 3D topological distance-based autocorrelation, lag 8, weighted by covalent radius. Since the covalent radii of carbon and oxygen atoms are similar, this descriptor does not weight the difference between oxygens and carbons as heavily as the other weighting schemes and tends to select longer molecules, particularly with aromatic rings that are several bonds apart and have a bend in their structure, referred to as the "bay" region, which is a recognized mutagenic structure.[78,234,235] The n3HeteroRing descriptor usually indicates the presence of an epoxide group. This is a well-recognized indicator of mutagenicity[236] and is often used as structural alert for mutagenicity.[200] The maximum electrotopical state of a hydrogen (hmax) is similar to a descriptor that was reported by Wang et al, 2009.[199] Wang suggested that the hydrogen atom from a hydroxyl group might engage in hydrogen bonds with DNA bases, assisting the mutagenic interaction.

3.5.2 Negative coefficients in the large molecule cluster

The negative descriptors are equally important in sorting the mutagenic structures from the non-mutagenic. As mentioned above, the largest negative coefficient is the centered Broto-Moreau autocorrelation, lag 3, weighted by Sanderson electronegativities (ATSC3e). This descriptor reflects the same atom 3 bonds apart in the molecular structure and favors unsubstituted rings as well as oxygens embedded a ring or aliphatic chain 3 bonds apart. JGI7, the mean topological charge index of order 7, has a strong negative coefficient. This descriptor indicates molecules that have a high charge transfer 7 atoms apart, and tends to highlight molecules with oxygen groups distributed all over the entire molecules surface, further confirming that the more mutagenic structures have a portion of the molecule with a large amount of charge transfer and a portion with less charge transfer. CrippenLogP is the octanol-water partition coefficient (Log P, also called Log K_{ow}) calculated via the Ghose-Crippen method. Log P has been previously reported as a positive variable for mutagenicity with nitro-PAHs.[196] However, in the large molecule cluster, it is assigned a negative coefficient. This difference may be due to controlling for size through the cluster assignments, or because this QSAR selects only direct-acting mutagens.

Certain carbon molecular structures are less likely to be mutagenic. The number of 12-membered rings, nT12Ring, is associated with non-mutagenicity. This descriptor tends to select molecules with at least 3 rings that are all joined together on the edges, resulting in more circular structures, which tend to be non-mutagenic. VR2_Dt is the normalized randic-like eigenvector-based index from detour matrix. The detour matrix reflects the longest path in a molecule's graph between each set of atoms. This descriptor is highest in molecules where most paths between atoms are also the longest path, which results in the descriptor selecting molecules with long aliphatic chains and unsubstituted rings. C4SP3 refers to a carbon bound to 4 other carbons. There are no

mutagens in the large molecule cluster that contain this particular structure. While a single methyl group in specific positions on the perimeter of a planar aromatic structure can enhance mutagenicity,[237–240] it is likely that the three-dimensional structure of the C4SP3 grouping inhibits a molecule's ability to cause mutagenicity; similar to how long aliphatic or substituted chains tend not to be mutagenic. There are also patterns of oxygen groups which can be non-mutagenic. nHBint8 is the count of E-State descriptors of strength for potential hydrogen bonds of path length 8. This descriptor tends to reflect hydroxyl groups on opposite sides of the molecule, which tend to be non-mutagenic in this cluster, whereas mutagens tend to have hydroxyl groups on only one side of the molecule.

There are three radial distribution functions (RDF) including both positive and negative coefficients. Although there is some correlation between them, the weightings used with these functions and the specific radii that apply to each account for the differences. RDF75s (radius of 7.5 Å, intrinsic state weighting), is the only RDF descriptor with a positive coefficient, and reflects the likelihood of finding a high intrinsic state structure within 7.5 Å of each atom in the molecule. RDF90i, (radius of 9.0 Å, ionization potential weighting) tends favor molecules with long aliphatic chains. RDF30p (radius 3.0 Å, relative polarizability weighting) highlights molecules with a large number of singly bonded carbons. RDF55s is the radial distribution function for intrinsic state at a distance of 5.5 Å. This is a 3D descriptor which reflects the likelihood of finding high intrinsic state molecules, particularly doubly bonded oxygens, within 5.5 Å of each atom in the molecule, which is approximately 3-4 carbon-carbon bonds. Molecules with high values for this descriptor have numerous oxygens evenly distributed over the molecules surface.

3.5.3 Positive coefficients in the small molecule cluster

In the small molecule cluster, several descriptors indicate dense structures with a high proportion of oxygen as likely mutagens in this cluster. The information content index (neighborhood symmetry of 0-order), IC0, has higher values when there are approximately equal numbers of carbons and oxygens in the molecule. The 10th bond order index ID (piPC10) reflects molecules that have at least 10 bonds in a series and has a stronger weighting for double or aromatic bonds, so this descriptor is highest when a large proportion of the molecule is aromatic. MATS4m is the Moran autocorrelation, lag 4, weighted by atomic mass. This descriptor tends to highlight molecules with oxygens groups close together on a molecule's perimeter, which like the epoxide group, tend to be a mutagenic sub-structure.[99,231] AATSC5s, the average centered Broto-Moreau autocorrelation, lag 5, weighted by intrinsic-state, indicates oxygen groups 5 atoms apart, such as para-oxygen groups. ATSC1e is the centered Broto-Moreau autocorrelation, lag 1, weighted by Sanderson electronegativities, and highlights molecules where there are oxygen-oxygen bonds.

As with the large molecules cluster, in the small molecules cluster, the presence of an epoxide group (indicated by SCH-3: simple chain, order 3) is a strong descriptor for mutagenicity. The mean topological charge index of order 2, JGI2, is highest when there are single atom perimeter groups, such as hydroxyl groups or methyl groups. Unsubstituted rings or rings joined by aliphatic chains have lower scores. E2p is the 2nd principal component accessibility directional Weighted Holistic Invariant Molecular (WHIM) index, weighted by relative polarizabilities. WHIM indexes are 3D molecular indices that reflect the covariance matrix based on the weighting.[241] Molecules with symmetric branches have a higher E2p value.

Unlike in the large molecule cluster, RDF55s has a positive coefficient in this cluster of small molecules. These descriptors select molecules with high intrinsic state values, such as hydroxyl groups or double bonded oxygens within short radii of most atoms in the molecule.

3.5.4 Negative coefficients in the small molecule cluster

Again, the negative descriptors are also necessary to distinguish mutagenic vs non-mutagenic structures. There are three negative descriptors highlighting Burden modified matrix minimum absolute eigenvalues, SpMin4_Bhi, SpMin5_Bhi, and SpMin6_Bhi. These descriptors all tend to select molecules with symmetric patterns, some of which can fold into 3D structures that are not good candidates for mutagenicity. The topological radius (topoRadius) is the minimum number of atoms across a molecule. In the small molecules cluster, molecules with larger topological radii typically have long, aliphatic chains, potentially substituted with oxygen groups, or rings linked by single bonds (such as the biphenyl structure) which are non-mutagenic structures.

As with the large cluster, certain structural patterns are less likely to be mutagenic. The Geary autocorrelation, lag 5, weighted by van der Waals volumes (GATS5v) highlights atoms 5 bonds away that have a different van der Waals volume. These may be a different atom, such as carbon vs oxygen, or they may be bonded differently. This tends to reflect aliphatic rings or chains, which are non-mutagenic structures. MATS2c is the Moran autocorrelation lag 2, weighted by charges. This descriptor tends to select carboxylic acid groups. There are no mutagens in the small cluster that contain carboxylic acid groups. The mindO and minsOH descriptors indicate the minimum energy state of a double bonded oxygen and an OH group, respectively. If the structure is not present, the value of 0 is assigned (which is then scaled with the rest of the data as part of

the analysis), so these two descriptors reflect the presence of OH and doubly bonded oxygens. Both structures are found in mutagenic and non-mutagenic structures and in the absence of other descriptors these structures are not always mutagenic.

3.5.5 Global summary of regression coefficients in both clusters

The OECD guidelines for QSAR development state that a QSAR should provide a mechanistic interpretation if possible.[230,242] Across both clusters, there are both key similarities and differences that may give us insight into the likely mechanisms of mutagenicity. Although there are a few mutagenic molecules without oxygen groups, mutagenic structures typically have oxygen groups on a part of the molecule, but not all of it. The epoxide group is a strong, although not a sufficient predictor of mutagenicity in both large and small molecules. Long aliphatic or oxygen-substituted chains are anti-mutagenic structures unless they contain an epoxide group close to the rings on the molecule.

In the large molecule cluster, the positive descriptors favor molecules that are asymmetric and have both a polar and a non-polar portion of the molecule. As reported previously, asymmetric structures such as the bay or fjord region contribute to the likelihood of a TP being mugatenic,[78,234,235,243,244] and more symmetric structures such as unbranched biphenyl groups tend to be non-mutagenic.[245,246] In the small molecules cluster, oxygens close together on the perimeter of a ring, either as hydroxyl groups or doubly bonded oxygens, contribute to the mutagenicity of the PAH structure and mutagens may be more symmetric.

3.5.6 Proposed mutagenic mechanisms

In the absence of studying each molecule experimentally, a potential mechanism of action for any individual TP can only be hypothesized. However, based on the strongest descriptors in each cluster, and what is known about the mechanisms of PAH mutagenicity, the following two mechanisms are proposed as the dominant mechanisms within the two major clusters. These are not exclusive categories and there is overlap, but the differences in the selected descriptors suggest some differences in the mutagenic mechanism.

3.5.6.1 Covalent Adduct Pathway

One of the most well studied mechanisms of PAH-induced mutations is formation of covalent adducts with DNA bases. These adducts can be either stable adducts, which persist within the DNA strand and interfere with proper replication of the strand, or unstable adducts, which remove purine bases from the DNA strand. The bay region diol epoxide group is well established as a structural indicator of covalent adduct mutagenicity.[83,243,247] The ortho-quinone structure has also been shown to be directly mutagenic through this mechanism in the Ames test.[99,219,231,248] Benzo[a]pyrene is the most studied PAH in this category, but it is not the only PAH that exhibits this mechanism of mugatenicity.[99,223,224]

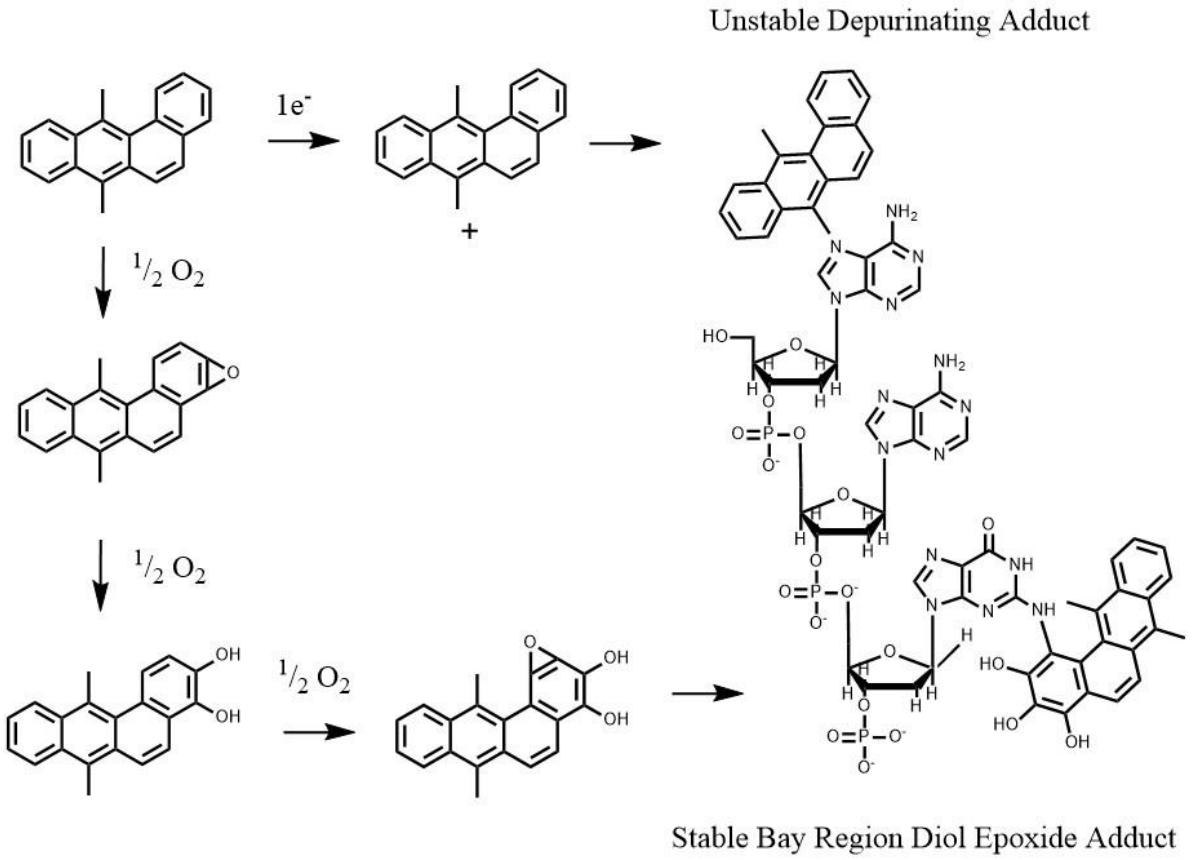


Figure 13 Covalent adduct and radical cation pathway. An example diol epoxide pathway (bottom), proposed by Huberman et al[249], and the radical cation pathway (top), proposed by Borosky and Laali[237], to covalent adduct formation.

Stable covalent adducts may be formed with either of the purine bases, primarily at the exocyclic nitrogen group. If the adducts are not repaired, the wrong pyrimidine base can be introduced on the opposite strand, leading to a base pair substitution.[250] Many PAHs can also lose an electron and become radical cations, which may create unstable DNA adducts with the purine bases (Fig 13). This results in the removal of a purine from the phosphate backbone of DNA.[250–252] Some studies suggest that an ionization potential of less than 7.35 eV can be used as the cutoff for predicting if a compound will be active via the radical cation pathway or not. Compounds with

greater ionization potential will likely only be activated by the diol-epoxide pathway, whereas compounds with a lower ionization potential may be activated by both pathways.[250] Since there are multiple mechanisms of mutagenicity, including some that do not require ionization, ionization potential alone is not a good predictor of mutagenicity.

The cluster of large molecules contains several descriptors that suggest that the covalent adduct pathway is relevant. The descriptors that indicate the bay or fjord region and the epoxide group contribute to this. The small molecules cluster also contains epoxide groups, so there may be some formation of adducts from this cluster, even though the molecules in the small molecules cluster are not large enough to contain bay and/or fjord regions.

3.5.6.2 Reactive Oxygen Species (ROS) Pathway

Several PAH intermediates can generate ROS through enzyme and metal catalyzed reactions,[222] and a subset of PAH TPs have the potential to follow both ROS generating pathways and covalent adduct pathways. Many quinone and catechol-like structures can cause ROS damage, including several smaller structures.[219–221] PAH catechols can generate ROS as they are metabolized to ortho-quinones. The resulting PAH ortho-quinones can enter futile ROS cycles (Fig 14) when cellular reducing agents such as NAD(P)H and NADH are present,[220,253] greatly amplifying their toxic impacts. This process generates hydroxyl radicals, hydrogen peroxide, and superoxide anion radicals ($\bullet OH^-$, H_2O_2, $\bullet O_2^-$).[232,254] Enzymes[219] and trace metals such as Cu(II)[254] or Fe(III)[255,256] can act as catalysts in all of the oxidation steps of the futile ROS cycle. NAD(P)H reduces the quinone back to the catechol structure, perpetuating the cycle. Other polar structures resulting from oxygenation of hydrocarbons may also generate ROS through similar pathways.[36,221] Although the quinone or catechol structure can autoxidize, the presence of a

catalyst has a significant impact on how quickly the ROS cycle iterates and consequently how many ROS are generated.

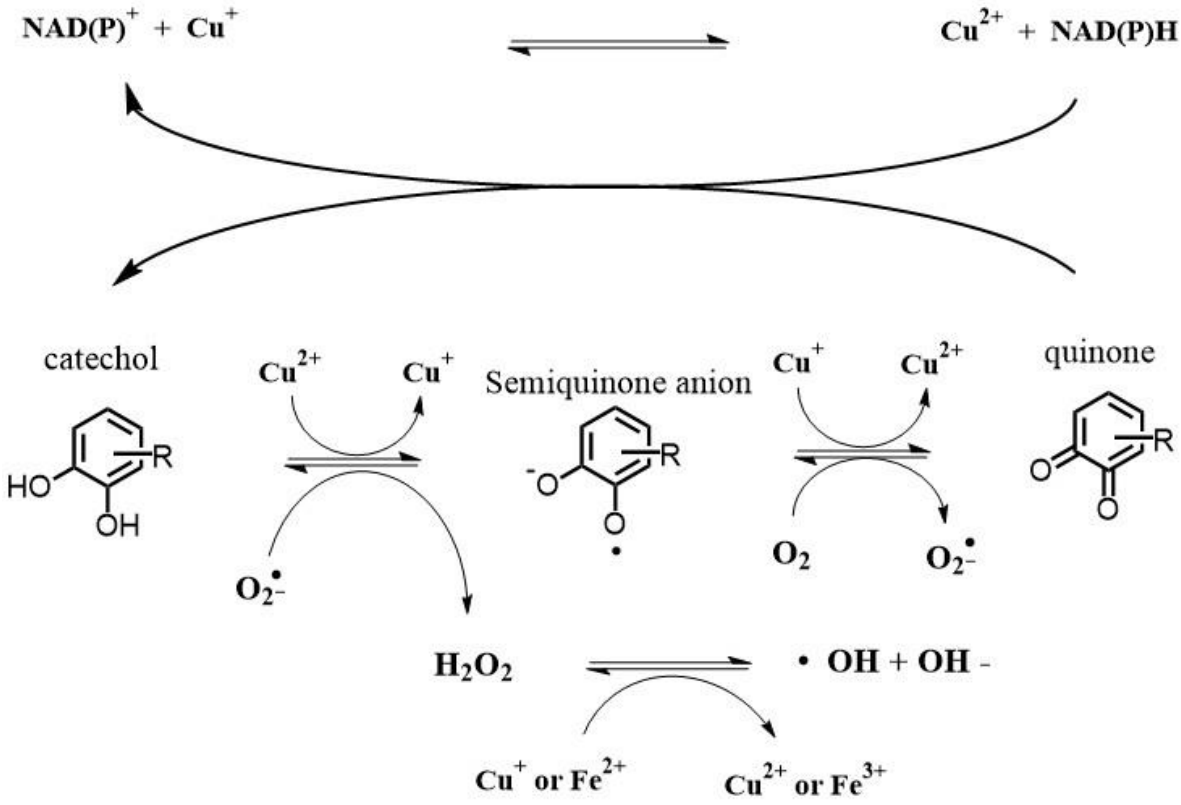


Figure 14 Possible ROS generating pathways for PAH catechol and quinone-like structures. Different steps of the process can generate hydrogen peroxide (H_2O_2), hydroxyl radical ($\bullet OH^-$), and the super-oxide anion radical ($\bullet O_2^-$). The Fenton reaction converts hydrogen peroxide into a hydroxyl radical. Reproduced from Flowers, L.; Ohnishi, S. T.; Penning, T. M. DNA Strand Scission by Polycyclic Aromatic Hydrocarbon *o* -Quinones: Role of Reactive Oxygen Species, Cu(II)/Cu(I) Redox Cycling, and *o* -Semiquinone Anion Radicals †, ‡. *Biochemistry* 1997, *36* (28), 8640–8648. https://doi.org/10.1021/bi970367p. 1997 American Chemical Society.[254] with additions adapted from Park et al, 2006[232]

While all of the DNA bases can be oxidized, the deoxyguanosine oxidation is the most common. Although this oxidation occurs naturally and can be rapidly repaired at normal levels, high levels of oxidized DNA can result in base pair mutations[257] and ultimately carcinogenesis. 2'-deoxyguanosine reacts with hydroxyl radicals to form 8-hydroxy-2'-deoxyguanosine (8-OHdG),[258] which is a potent enough biomarker for oxidative DNA damage that it has been proposed as an early indicator of colorectal cancer in humans.[259]

There are descriptors in both clusters that are representative of the ROS pathway, and molecules of various sizes can engage in these ROS generating pathways.[260] However, it appears to be the dominant mechanism of mutagenicity in the small molecules cluster. The descriptors MATS4m, RDF55s and AATSC5s are all indicative of catechol- or quinone-like structures, supporting the ROS pathway as a likely explanation of mutagenicity.

3.5.7 Benefits of clustering

Using laboratory bioassays to study the mutagenic mechanisms from the vast array of all possible molecules that can result from PAH degradation would be prohibitively expensive and time consuming. Thus a computational approach to estimating the most likely method of mutagenicity to guide experiments has substantial promise. The differences in the descriptors selected in the two major clusters of this study suggest that there may be different dominant mechanisms of mutagenicity in the two different clusters. These distinctions are not exclusive, and it is likely that both known mechanisms of mutagenicity are active in each cluster. However, certain differences are quite pronounced. There are no descriptors with a lag of 1 selected in the large molecules cluster, but in the small molecules cluster, ATSC1e and IC0 both indicate similar atoms close to each other in 3-D space or next to each other in the molecular structure, respectively.

The impact of HOMO-LUMO Gap (HLgap) changes between the two clusters. In the large molecules cluster, HOMO-LUMO gap has a moderately negative coefficient, which generally indicates that mutagens are likely to be more reactive molecules with smaller gaps and in the small molecules cluster it is not selected as a relevant descriptor. We hypothesize that the absence of selection of HOMO-LUMO gap in the small molecules cluster may indicate the molecule persisting long enough to enter the futile ROS cycle. Once in ROS cycle, catalysts may account for the ability of molecules with the larger HOMO-LUMO gaps to be mutagenic. The radial distribution function with a radius of 5.5 Å, weighted by intrinsic state (RDF55s), is a negative descriptor in the large molecules cluster, but a positive descriptor in the small molecules cluster. This is attributable to the need for mutagens in the large molecules cluster to have a greater portion of the molecule without oxygen groups (which have a high intrinsic state value, Table S13), whereas the small molecules which tend to cause mutagenicity via ROS can have a higher proportion of oxygen groups over their surface area.

Between the two clusters, there are differences in the RDF55s coefficients as well as in the variables (including HOMO-LUMO gap) selected in each cluster (Figs 11-12). Additionally, there is improvement in the performance of the F1 classifier metric as well as a reduction in the number of descriptors needed by analyzing within each cluster rather than by analyzing the entire dataset all at once. These distinctions suggest that the small molecules and large molecules have different structural properties that contribute to mutagenicity, and likely different mechanisms, and that the other molecules should be excluded from the training data on their structural differences from both the large and small molecule clusters.

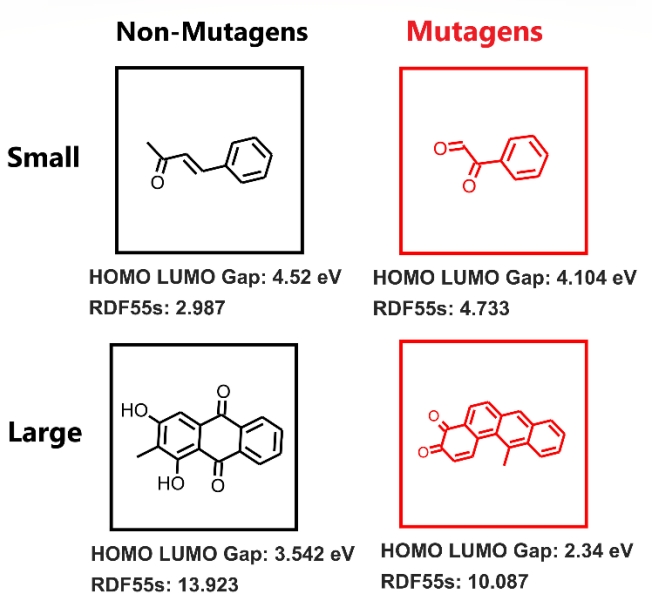


Figure 15 Examples of structures illustrating the differences in selected negative and positive descriptors in each cluster. In the small molecules cluster, high values of RDF55s (Radial Density Function at a radius of 5.5 Å, weighted by intrinsic state) are associated with a higher likelihood of mutagenicity, and the HOMO-LUMO gap values are similar. In the large molecules cluster, Low values of HOMO-LUMO gap indicate a molecule that is more reactive, and are associated with a higher likelihood of mutagenicity. High values of RDF55s are less likely to be mutagenic in the large molecule cluster, but more likely to be mutagenic in the small molecule cluster. Also see Figures S3.1-S3.6

3.6 Conclusions

The improvement in performance of our QSAR (F1 scores of 0.7 to 0.8) compared to those trained on data that used metabolic activation (F1 score of 0.61 to 0.7) shows how crucial it is to document whether or not metabolic activation was used in Ames Tests. Future datasets[189] developed for the Ames test should specify this difference. Sub-dividing the data using unsupervised clustering before applying machine learning techniques for classification is a novel

approach that shows significant promise in the area of topological-descriptor-based classifiers. The reduction in the number of descriptors needed in order to achieve high classifier performance by evaluating within each cluster rather than across the entire dataset suggests that there may be different mechanisms (covalent adducts and ROS) which dominate each cluster. In large datasets where there may be multiple mechanisms of mutagenicity, unsupervised clustering may provide a simple unbiased method of sorting compounds to create more constrained, but more accurate QSARs.

Empirical studies of PAH contamination demonstrate that their TPs are a significant contributor to the mutagenic hazards at PAH-contaminated locations. The QSAR developed in this work improves upon previous PAH-QSARs by more accurately detecting direct-acting mutagens. Known mutagenic structures such as the bay region epoxide group and quinone-like compounds are useful descriptors. However, the distribution of oxygen groups over the surface of the molecule and the molecule's overall shape are also necessary to determine mutagenicity or non-mutagenicity. The number of 3D descriptors selected shows the importance of the 3D structure to mutagenic action. Future mutagenicity classifiers should use optimized structures and include 3D descriptors in their training dataset. Although a larger dataset was used than in previous PAH studies, a simple, regression-based classifier was still achievable, demonstrating that more advanced algorithms which make it harder to obtain a mechanistic interpretation may not be ideal or even necessary for predicting PAH mutagenicity.

Due to the vast number of possible PAH TPs[175] that can result from PAH environmental contamination, computational approaches to estimating the potential hazards are the most feasible and accessible solution. Future regulation and remediation approaches may benefit from a better understanding of pathways that lead to mutagenic compounds, ultimately leading to

biodegradation approaches that avoid or remove these mutagenic TPs. Hazard assessments for PAH contamination should incorporate degradation products and evaluate the most likely mutagenic structures in order to avoid trading one hazard for another in the process of remediating PAH soil contamination.

4.0 Empirical Evaluation of Mutagenic TP Formation

4.1 Introduction

4.1.1 Motivation

The networks model developed in Chapter 2 provided a method to estimate the most likely compounds that would occur in the environment. In order to estimate which PAHs would degrade into mutagenic transformation products, the QSAR developed in Chapter 3 was applied to the high throughput compounds that were predicted by the networks model. These compounds were assumed to have the greatest probability of occurring in an actual degradation experiment. The direct-acting mutagenicity QSAR is a tool for predicting whether a given structure is likely to be mutagenic or not. These two tools can be used in combination to estimate whether a given PAH is likely to degrade into mutagenic TPs and if so, where in the degradation chain they are likely to occur. In order to verify the predictive tools, we designed an experiment to evaluate whether, when, and to what extent mutagenic transformation products would form from a given PAH in a laboratory biodegradation experiment.

4.1.2 Distinctions from previous biodegradation mutagenicity studies

The occurrence of mutagenicity from biodegradation intermediates has been previously evaluated with various experimental designs, and environmental transformation of PAHs can increase their mutagenicity (Chap 1, Table 2). In many previous studies, known or suspected PAH

intermediates were synthesized and evaluated for mutagenicity. While valuable in the structural specificity offered through compound synthesis, this approach does not capture the potential contribution of TP mixtures[89,261] or the formation of unknown metabolites. In other studies, the mutagenicity of a PAH degradation culture (such as phenanthrene) was evaluated before and after a period of degradation, but not at intermediate time points (Table 4). This approach runs the risk of missing mutagenic TPs which persist long enough to cause harm, but ultimately degrade before the final time point. Such intermediate TPs are particularly important when translating results of a batch culture study, which proceeds from a deliberate starting point to a deliberate end point, to an environmental context. Unless the contaminant source has been eliminated or controlled, there is typically ongoing input of the parent contaminant compound (here, PAHs) even as legacy pollution transforms, causing new TPs continue to be generated from the new contamination.

Another key difference between this study and previous work on PAH biodegradation mutagenicity is the simplicity of the sample preparation. Many previous studies employ techniques such as liquid-liquid extractions,[134] fractionation,[93,261] or directly synthesizing known metabolites.[244] While these studies are valuable for their specificity and their enhanced detection ability, their results may not be as representative of the actual degradation conditions as a sample preparation with fewer purification and concentration steps. Additionally, some studies only take samples at the beginning and end of a degradation culture,[262] but not at intermediate time points during the degradation culture. This study provides a complementary option to more processed samples with a very simple sample preparation procedure and samples taken at intermediate time points.

4.1.3 Selection of PAHs for Batch Degradation

The dominant PAHs in any given contaminated environment vary widely. However, phenanthrene is one of the most abundant PAHs found in environmental studies and is also a smaller PAH, making it a preferable starting point for designing the predictive tools in this work due to a manageable set of possible TPs and good availability of empirical literature.[24,57,263,264] The empirical literature (Chap 2) was reviewed to identify intermediate aerobic biodegradation products of phenanthrene, although only a few studies included the associated hazards. Additionally, phenanthrene has several known microbial degraders (e.g., *Pseudomonas Putida*, used here) which are available from laboratory suppliers, facilitating laboratory degradation cultures without needing to isolate an environmental strain from a contaminated site. Furthermore phenanthrene has a "bay" region, a recess in the aromatic rings, making it a good candidate for exploring mutagenicity induced by biodegradation.[78] The availability of existing microbial degraders and the bay region[80,235,235,243] suggest that phenanthrene may introduce observable mutagenicity in a biodegradation experiment. Several previous studies suggest that phenanthrene may generate mutagenic TPs (Table 9).

Fluorene was selected as the second PAH for use in batch degradation studies. Pilot experiments confirmed that *Pseudomonas Putida* was also able to degrade fluorene. Fluorene is typically found in lower concentrations than phenanthrene in environmental contexts and thus, is less studied. The empirical literature review conducted for the networks model found approximately half as many empirical TPs reported for fluorene as for phenanthrene. The QSAR developed in Chapter 3 was used to estimate the likelihood of mutagens occurring based on the predicted degradation projects from the network models from these two PAHs.

Table 9 Previous Phenanthrene Mutagenicity Studies

Experimental Design	Strains*	Results	Study
Batch culture, bacteria immobilized on porous surfaces, glucose as a co-metabolite, ethyl acetate extraction used	TA98	Positive when co-metabolized with glucose. S9 fraction reduced mutagenic effect.	Traczewska et al, 2000[134]
Individual compounds that are potential metabolites	TA98, TA100, TA1535, TA1538	Some positives	Bucker et al, 1979[244]
Biodegradation with alfalfa	TA98, TA100	Negative	Flocco et al, 2002[265]
Individual compounds	TA98, TA100	Some positives	Wood et al, 1979[84]
Fungal enzyme degradation	TA98, TA100	Negative	Wulandari et al, 2021[262]

*Notes on Ames test strains: TA98: Frameshift Mutation, TA100; Basepair substitution with SOS promoting plasmid. TA1535: Basepair Substitution Mutation

In the case of fluorene, minimal information is available for the mutagenicity of degradation TPs. Most of the available studies for mutagenicity in relation to fluorene have been conducted with either unmodified fluorene, or fluorene bound to groups not likely to occur through environmental transformations. Two fluorene amides that were being developed as insecticides in the 1940s, 2-Aminofluorene and 2-acetylaminofluorene, were discovered to be carcinogenic.[266] Mutagenicity studies using fluorene tend to focus on these and their derivatives.[246,267,268]

4.2 Materials and Methods

4.2.1 Overall Experimental Design

The high-throughput networks predicted by the networks model (Chap 2) were evaluated for potential mutagens with the QSAR (chap 3). The placement of mutagens on the pathways most likely to occur based on empirical literature (see 4.3.1) was used qualitatively to design the degradation experiments.

Based on the results from a pilot phenanthrene degradation culture, (Appendix C), the location in the degradation chain of the mutagens predicted by the networks model (Chap 2) and the QSAR (Chap 3), a 2-day time interval for the Ames FT samples was selected. From previous pilot work, (Appendix C) most degradation runs in this experimental setup were anticipated to be complete (either fully degraded, or the degradation significantly leveling off) by approximately 1 week, so time points of 2, 4, 6, and 8 days were selected. (Fig 16).

Phenanthrene and fluorene were degraded by *Pseudomonas Putida* in batch degradation cultures. Aliquots were pulled from the PAH degradation cultures, measured with the HPLC to determine the extent of degradation, and tested with the Ames FT. Two weeklong batch culture experiments were conducted, using *Pseudomonas Putida* as the PAH degrading strain, and the Ames tester strains TA98 and TA100 to measure mutagenicity.

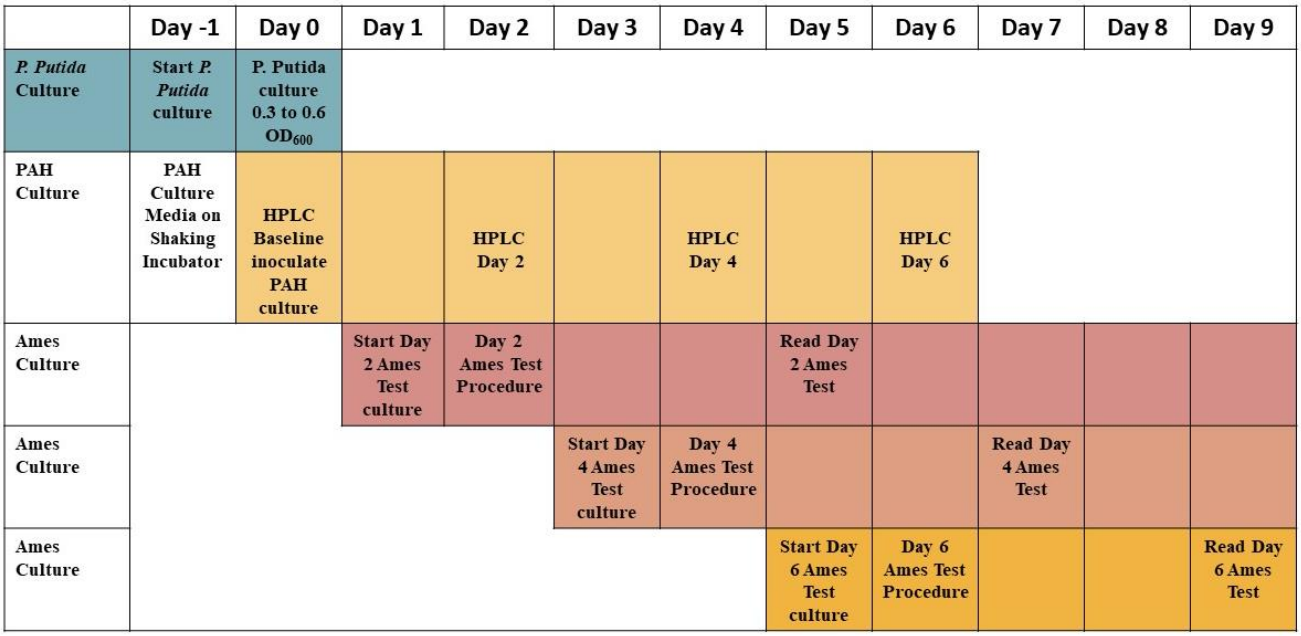


Figure 16 Gantt chart for the full experimental design, showing 6 days. Each Ames FT overnight culture was started the day prior to the Ames Test procedure and the P. Putida degradation culture ran the entire time. If the degradation culture ran for 8 days the same process was extended for 2 more days.

Three replicates were used for HPLC analysis and were filtered to remove cells and any suspended solids prior to analysis. An abiotic control (Figs 20-22) was included to determine if any induced mutagenicity could be attributable to photo-degradation. The abiotic controls were prepared in the same manner as the degradation culture, placed on the shaking incubator and exposed to sunlight, but not inoculated with bacteria.

4.2.2 PAH-Media Preparation

Bushnell-Haas (BH) Media was prepared according to the manufacturer's instructions and autoclaved at 121 ºC for 20 minutes. Tween® 80 was used to help stabilize the PAH in solution

with the bacteria. Since Tween cannot be autoclaved, it was added to previously autoclaved BH media at a concentration of 2% (v/v) using a hotplate at 120 ºC and frequent manual swirling. The BH media/Tween solution was then filter sterilized through 0.1 µm or 0.2 µm filters based on availability, with a sterile syringe and diluted to a concentration of 0.5% Tween (v/v), which was adequate to stabilize our PAHs. The PAH (phenanthrene or fluorene) was dissolved in acetone at 5,000 mg/L and an appropriate amount added to the BH media/Tween to yield 100 mg/L of PAH. This BH media/Tween/PAH solution was left open in a chemical hood, covered with sterile cotton overnight to allow the acetone to evaporate, leaving the PAH behind. This stock solution (hereafter referred to as "PAH culture media") was then stored away from light to prevent any photo-transformations.

4.2.2.1 *Pseudomonas Putida* Freezer Stock Preparation

Pseudomonas Putida was selected for the degradation experiments because it is a known PAH degrader.[113,114,269–273] *Pseudomonas Putida* (ATCC 17484) was purchased from the American Type Culture Collection, Stock cultures were prepared by growing a culture overnight in BD Difco™ growth media (a beef peptone media). 0.5 mL of this culture was combined with 0.5 mL of 50% glycerol (v/v) for a final volume of 1 mL and 25% glycerol (v/v) in 1.5 mL microcentrifuge tubes. The prepared microcentrifuge tubes were labeled and stored in a -80 ºC freezer.

4.2.2.2 PAH Degradation Culture Inoculation

Pseudomonas Putida cultures were started from freezer stock as described in Appendix B and grown overnight in a shaking incubator 28 ºC and 120 rpm to early exponential phase (approximately 18 hrs, OD_{600} 0.3 to 0.6) in BD Difco media. If the culture proceeded all the way to the stationary phase, it was passaged once and the passaged culture was used for inoculating the PAH culture media while in the early exponential phase. The PAH culture media flasks to be used for inoculation were filled with 100-120 mL of PAH culture media from the pre-prepared stock and placed in the shaking incubator, wrapped in foil, at the same time as the *Pseudomonas Putida* cultures. This provided several hours of shaking prior to inoculation or light exposure in order to fully solubilize the PAH.

Cells were extracted from the *Pseudomonas Putida* cultures in the early exponential phase and 40 mL added to a sterile 50 mL tube. This tube was centrifuged at 6000 rpm for 5 minutes, decanted and the pellet resuspended in BH media, centrifuged a second time, and the pellet resuspended in PAH culture media to avoid diluting the PAH culture when adding the cell suspension to it.

Based on our OD_{600} calibration curve (Appendix B), 10^8 cells/mL were added to 100-120 mL of the PAH media in a sterile 250 mL shaker flask. The mouth of the flask was filled with sterile cotton to allow air exchange (Fig 21) and the flasks were placed immediately onto a shaking incubator at 28 ºC and 120 rpm, uncovered, to permit any potential photo-transformations that occur in an environmental degradation context. Abiotic control flasks were also placed on the shaking incubator and contained all of the same preparations except the addition of the cells was excluded. The shaking incubator was located next to a south-facing window. At the predetermined time points, a small aliquot (< 5 mL) was extracted from the flask and the flask was immediately

returned to the shaking incubator. The aliquot was filtered using a sterile syringe and a 0.1-0.2 µm filter (based on available supplies) to remove cells and any suspended solids, such as undissolved PAH or small clumps of the BH media. The filtrate was analyzed by HPLC and used for carrying out the Ames Test.

An Agilent™ 1200 series High-Performance Liquid Chromatograph (HPLC) with a UV detector was used to monitor the PAH degradation. A 4.6 mm x 150 mm Agilent™ C18 Column was used with a gradient of acetonitrile and de-ionized water. Individual TPs were not identified as this has already been heavily studied and metabolite identification was not the focus of this study. Peak height at 254 nm was used to measure the PAH. Details on the HPLC configuration are provided in Appendix B, and selected HPLC data is provided in Appendix D.

4.2.3 Ames Fluctuation Test

The Ames Fluctuation Test[261,274,275] (Ames FT) was selected to measure the occurrence of mutagenicity in the PAH degradation cultures. The Ames FT is a variation on the well-established traditional Ames Test for mutagenicity, which is a pour-plate bioassay.[21] In the traditional Ames Test, tester strains of *Salmonella typhimurium* containing a reverse mutation in the histidine coding gene are mixed into a top agar and poured over a base agar. Each Ames tester strain contains a specific type of reverse mutation (e.i. frameshift, base-pair substitution). The substance to be analyzed for mutagenicity is mixed into the top agar at a dilution gradient in order to assess dose-response. If a mutagen is present the mutation will reverse and colonies will form. The number of colonies are counted and evaluated against a negative control to correct for spontaneous mutations.[276] In the Ames FT the tester strains are cultured in multi-well plates and a colorimetric indicator is used to evaluate bacterial growth indicating the presence of a mutagen. The Ames FT

is more facile and efficient (with both time and material) to conduct than the traditional pour plate design. The Ames FT also has a similar level of sensitivity to the pour plate test,[22,277,278] and has been applied successfully to similar projects.[261,274,275] Due to the high volume of samples required for this project, the FT test was selected. An adjustment to the standard Ames Test procedure for this experiment was the omission of S9 replicates, which are typically used for enzymatic activation, designed to reflect higher organism's internal metabolism.[115] Since this project was intended to reflect only the mutagenicity that would result from environmental microbial degradation, we did not include the S9 activation component. The tester strains TA98 (frameshift mutation) and TA100 (base-pair substitution mutation) were selected as these provide two types of mutations, and are commonly used to evaluate the mutagenicity of water or wastewater,[279–281] and are the strains specified in ISO 11350.[203] ControlChem™ positive controls were used, 2-nitrofluorene for TA98 and 4-nitroquinoline-N-oxide for TA100. The controls were dissolved in 1 mL dimethyl sulfoxide for a dosing solution of 50 µg/mL. The Ames FT has a 1/25 dilution when the test substance is added to the media, so the final positive control dose was 2 µg/mL. All Ames Test specific materials were purchased from Molecular Toxicology (MOLTOX™), Inc, Boone, North Carolina.

4.2.3.1 Ames Fluctuation Test Process

To conduct an Ames FT, the TA98 and TA100 strains were inoculated with a 20 µL injection of a freshly thawed freezer stock (preparation described below) into 10mL of MolTox™ growth media in a sterile 50 mL Erlenmeyer flask. The exact volume of this culture was flexible as the culture would be required to meet a specific OD_{600} before being passaged for exposure to the sample. Extra growth media was used in the experiments with abiotic controls or if additional

OD_{600} checks were needed. This flask was incubated (stationary) overnight at 37 ºC and then placed on a shaking incubator on the morning of the day of the assay at 37 ºC and 150 rpm until it reached an OD_{600} of at least 1.0. 800uL of this overnight culture was added to 7.2 mL of MolTox™ exposure media. For abiotic controls 1.5x as much bacteria and exposure media were prepared. A 24 well plate was prepared with 10 uL of each sample (positive and negative controls and serial dilutions) and 240 uL of the exposure media mixture was added to each well. 3 dilution replicates with 6 serial dilutions of 50% were used for final dilutions of 1, 1/2, 1/4, 1/8, 1/16, and 1/32 of the original aliquot from the Ames degradation culture, with 3 positive and 3 negative controls for each sample. Autoclaved de-ionized water was used for dilution. Serial dilutions of 1, 1/4, and 1/16 were used for the abiotic controls. This plate was then placed on a shaking incubator at 37 ºC for 90 minutes. If an overnight culture reached OD_{600} of 1.0 and could not be used within 15 minutes due to the 24 well plate for the other strain being prepared it was placed in a 4 ºC refrigerator for no more than 2 hrs, and then used as described above. After the 90-minute incubation time the 24 well plate was removed from the shaking incubator and 2.75 mL of MolTox™ reversion indicator media was added to each well. Each of the 24 wells was then transferred to 48 wells on a 384 well plate, 50 µL per well. Each set of dilutions with positive and negative controls fills one 384 well plate. The 384 well plates were placed in a resealable plastic bag (i.e. Ziploc®), and placed in a stationary incubator at 37 ºC for 72 hrs. The Reversion Indicator Media changes from a deep purple color to a bright yellow if there is cell growth, indicating reverse mutations. (Fig 17) The full instructions for this Ames FT can be obtained from MOLTOX™.[282]

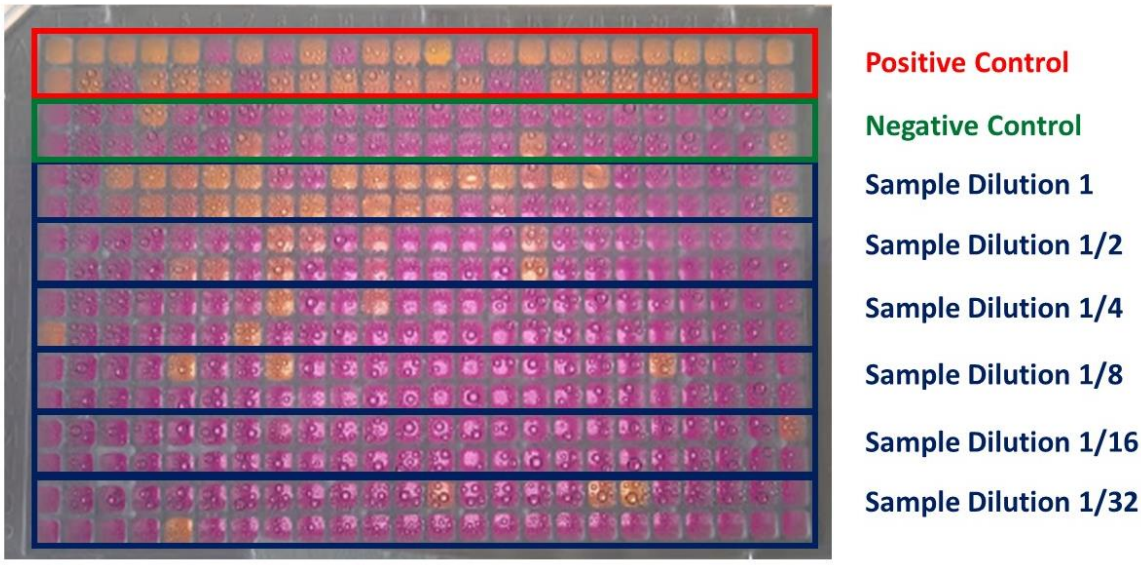


Figure 17 Example Ames Test 384 well plate. Each Ames Test time point has 3 such plates per strain, and the mean of the counts of the positive wells are used to determine the increase over the mean of the negative control.

4.2.3.2 Ames Test Freezer Stock Preparation

Freezer stock for the Ames FT was prepared by incubating the TA98 and TA100 *Salmonella typhimurium* overnight at 37 ºC and shaking to an OD_{600} of at least 1.0 at 37 ºC and 150 rpm the next day. 0.5 mL of 50% (v/v) glycerol was added in equal volumes to the culture to yield a final mixture of cell suspension in growth media with 25% glycerol (v/v) in 1.5 mL tubes, which were preserved at –80 ºC. 10 tubes of frozen Ames Test stock were prepared in a batch and each batch of frozen stock was tested to ensure responsiveness to the positive control (≥25 positive wells) and an acceptably low background rate of naturally occurring revertant positive wells (≤10), as defined by ISO 11350, "Water quality — Determination of the genotoxicity of water and waste water —Salmonella/microsome fluctuation test (Ames fluctuation test)". Freeze-thaw cycles can

cause high background reversion or loss of response to mutagens in Ames tester strains, [115,283] so each individually prepared frozen vial of Ames tester strain bacteria was only used once. To start an overnight Ames culture, a single microcentrifuge tube for each strain was thawed and 20 µLs of the contents was immediately used to inoculate an overnight culture.[284,285]

4.2.3.3 Ames Fluctuation Test Statistical Interpretation

The Ames FT is typically considered positive for mutagenicity when a statistically significant 2-fold increase over the negative control is observed.[93,203,274,277] For the purposes of this experiment, criteria similar to that used by Flückiger-Isler and Kamber, 2012[277] were applied to the results. In order to be positive a test value must satisfy the following criteria (Table 10):

1. At least a 1.5-fold increase for a weak positive and at least a 2-fold increase for a positive.

2. Significance (t-test) at $\alpha < 0.05$ or $\alpha < 0.1$, for a weak positive. The well counts were arcsine–square-root transformed as recommended by ISO 11350.[203]

3. A dose-dependent response. A linear response is not necessarily anticipated, but in order to be considered a positive, the mutagenic response should eventually cease to be observed at the lower doses.

4. If the mean value of the negative control replicates was less than 1, the value 1 was assigned for evaluating the fold response.[277] This ensures that a false positive is not observed due to background rates of spontaneous reversion.

Additionally, the validity criteria from ISO 11350[203] was applied. In order to be valid, the controls in a test must satisfy the following criteria:

A. the mean number of positive wells for negative controls must be ≥0 and ≤10.

B. the mean number of positive wells for positive controls must be ≥25.

The Ames FT can be read between 48 and 72 hrs of incubation.[203,282] All assays were read at approximately 72 hrs due to the anticipated weak response of the PAH TPs.

Table 10 Summary of positive and weak positive Ames FT result criteria. The fold-increase and the α level of significant pertain to individual dilutions with 3 replicates in an Ames FT. The overall Ames FT result

Fold Increase of the mean positive wells of a set of dilution replicates relative to negative control from that time point	α level of significance	Overall Ames FT Result
>=2	<0.05	**Positive**
>=2	<0.1	**Weak Positive**
>=1.5	<0.05	**Weak Positive**

In all cultures, the same batch of PAH culture media as was used for the degradation experiment, (sterile, never exposed to bacteria or significant amounts of sunlight) was run on the Ames Test for the day 0 time point. No batch of PAH culture media of either fluorene or phenanthrene was ever positive for mutagenicity on the Ames FT Test. Some of the phenanthrene cultures did not successfully inoculate on the first attempt and required a second inoculation of washed cells from the enrichment culture.

4.3 Results and Discussion

See attached file Chap4_PAH_culture_data.xlsx for raw data

4.3.1 Mutagenicity predictions from networks model and QSAR

Pilot experiments confirmed that the *Pseudomonas Putida* strain was capable of degrading both phenanthrene and fluorene. These pilot experiments were conducted as described in section 4.2, using the same PAH culture media preparation, inoculation procedure and HPLC methods, but not including the Ames Test.

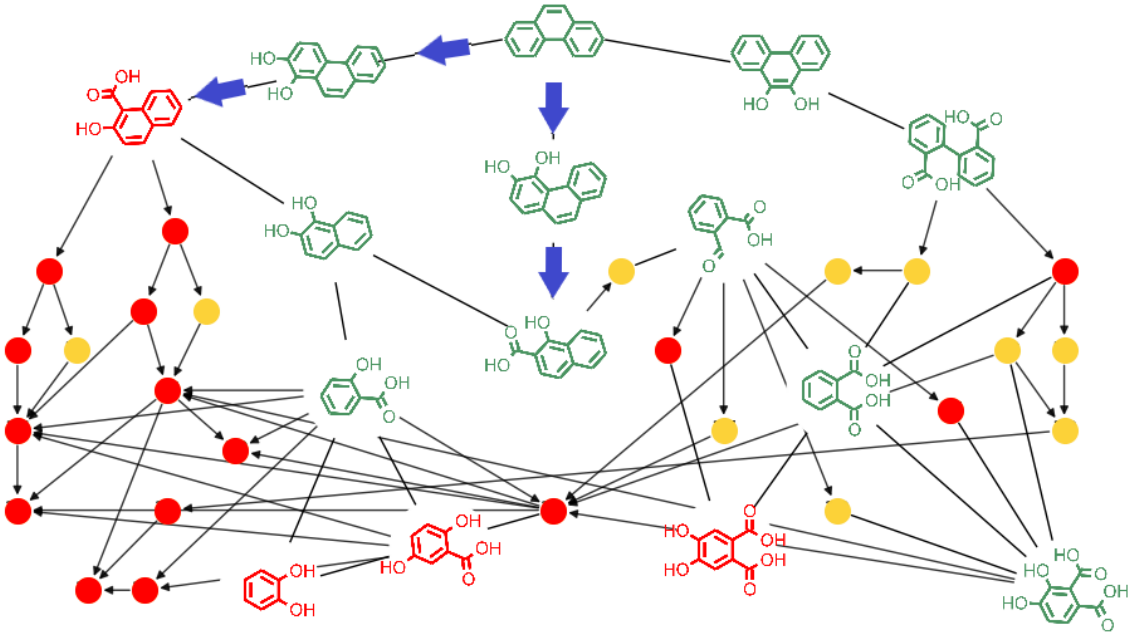


Figure 18 High Node Throughput (See Chap 2) portion of the predicted phenanthrene degradation network (refer to Chapter 2). Compounds that are a perfect match to those identified in an empirical study are shown as structures. Predicted TPs identified by the networks model (Chapter 2) are shown as circles. Red circles and structures indicate a potential mutagen as predicted by the QSAR (Chapter 3). The blue arrows indicate the pathways documented in literature by *Pseudomonas Putida* strains. (see Chap 2 SI)

Based on our mutagenicity QSAR described in Chapter 3, the number of predicted mutagenic products and the number of TPs before a mutagen was predicted from the original source PAH varied significantly between phenanthrene and fluorene. These differences made these two PAHs good candidates for testing the ability of the networks model and the QSAR to predict the potential occurrence of mutagenic TPs in a PAH degradation network. The QSAR was applied to the high node throughput compounds for phenanthrene (Fig 18) and fluorene (Fig 19). All input data was prepared in the same manner as the training data for the QSAR (Chap 3).

Based on previous degradation studies from literature using *Pseudomonas Putida* strains, the phenanthrene degradation culture was anticipated to proceed down either the left[286,287] or center pathways[113,288,289] (Figure 18, indicated by blue arrows), or both. The left-most pathway generates more predicted mutagens (denoted in red) than any other portion of the network.

It was anticipated that fluorene would degrade through the 9-hydroxyfluorene group, the most commonly documented pathway in empirical literature,[136,290–293] (blue arrows in Fig 19, also see literature review summary in Chapter 2 SI). This pathway does not generate many predicted mutagens until the very bottom of the network. Although this is only an approximation as the kinetics of each transition cannot be known, the TPs lower in the network are only a few transitions away from opening the last aromatic ring. Fluorene was therefore hypothesized to be less likely to generate mutagens than phenanthrene.

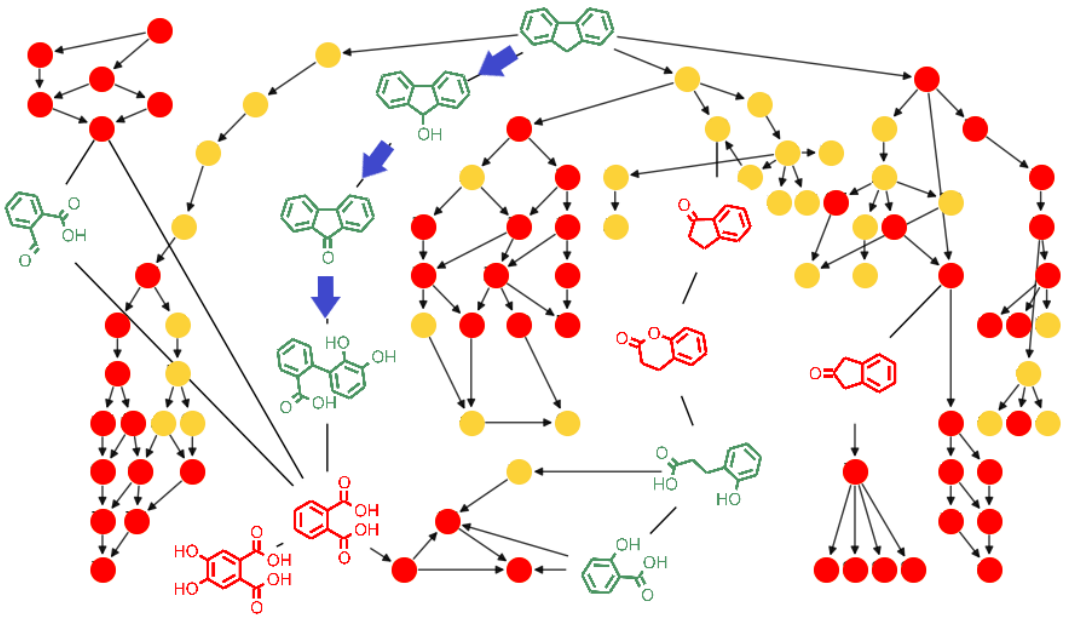


Figure 19 High Node Throughput (see Chap 2) portion of the predicted fluorene degradation network (refer to Chapter 2). Compounds that are a perfect match to those identified in an empirical study are shown as structures. Predicted TPs identified by the networks model (Chapter 2) are shown as circles. Red circles and structures indicate a potential mutagen as predicted by the QSAR (Chapter 3). The group of compounds in the upper left corner indicates a point where several lower throughput pathways joined that resulted in a high node throughput region. The blue arrows indicate the most commonly documented degradation pathway from literature (See Chap 2 SI).

4.3.2 Biodegradation Culture Ames Tests

Two degradation culture runs were conducted with phenanthrene and one with fluorene. The occurrence of direct-acting mutagenicity in a live biodegradation culture, without the use of a detailed preparation chain to isolate and/or concentrate specific metabolites, was hypothesized to

be transient in the degradation culture. All raw data is provided in the attached PAH_culture_data.xlsx file.

4.3.3 Phenanthrene Degradation

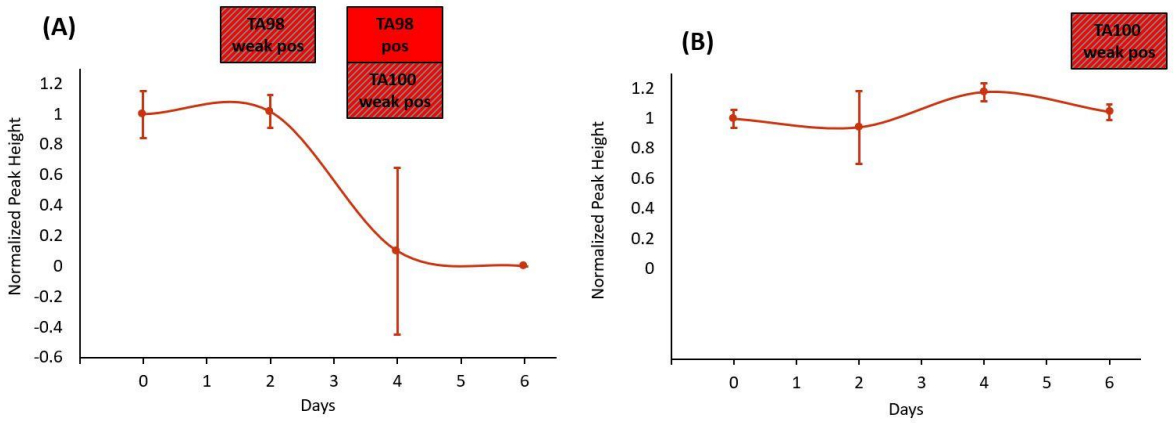


Figure 20 Normalized HPLC peak height (absorbance at 254 nm) (A) Phenanthrene biodegradation culture with Ames FT results. (B) Phenanthrene Abiotic Control. The error bars show one standard deviation, (n=3, some time points had extra replicates, see Chap4_PAH_culture_data.xlsx) The subsequent days were normalized to the day 0 mean peak height. Ames FT results (Tables 11 and 12) are shown in the shaded boxes. The TA100 weak pos on day 6 was the only positive result from the abiotic control.

The phenanthrene peak was almost undetectable on the HPLC by 6 days. (Fig 20) On Day 2, a weak positive was detected in the TA98 strain, at the start of the culture degradation, and on Day 4 a positive was detected in the TA98 strain and a weak positive in the TA100 strain. By Day 6, no further positives were detected, confirming the hypothesis that mutagenicity would occur while the culture was partially degraded. Several days did not meet the criteria for valid samples due to the positive or negative Ames FT controls being outside of the acceptable range (see 4.2.3.3,

and Table 11). It is anticipated that invalid controls would be most likely to impact the test's ability to show positives, as a low positive control would indicate a weak response for that plate and a high negative control would make it less likely to see enough of an increase in positive wells (relative to the negative control) to detect mutagenicity. Therefore, time points that still met the criteria to be considered positive may be of interest even if the controls were out of the desired range.

Table 11 Phenanthrene Ames FT Data. The values for each day/dilution are the fold increase of the mean number of positive wells (n = 3 replicates) relative to the mean number of positive wells of the negative control (n=3 replicates) for that day. Highlighting indicates the level of significance and magnitude of fold change; fold changes may be greater than 2 and not be significant at least α < 0.1. *The TA98 Day 2 positive control was low at 21 mean positive wells. *TA100 Day 4 negative control was high at 16 mean positive wells. See attached file PAH_culture_data.xlsx for full details

TA98	Day 0	Day 2*	Day 4	Day 6
1	1.50	3.00	1.33	1.17
1/2	1.00	1.67	0.67	0.67
1/4	1.50	0.67	1.00	0.63
1/8	0.50	0.67	2.33	0.79
1/16	1.25	1.00	2.67	0.58
1/32	0.25	0.67	1.00	0.50
	neg	weak pos	pos	neg

TA100	Day 0	Day 2	Day 4*	Day 6
1	1.07	1.06	1.70	0.87
1/2	0.81	1.28	1.60	0.63
1/4	1.22	1.22	1.47	0.87
1/8	1.11	1.28	1.47	0.80
1/16	0.78	1.33	1.51	0.90
1/32	1.11	1.17	0.87	0.80
	neg	neg	weak pos	neg

Individual Dilution Result	Ames FT Result
α < 0.05, fold increase > 1.5 =	weak positive
α < 0.1, fold increase >= 2.0 =	weak positive
α < 0.05, fold increase >= 2.0 =	overall positive

The abiotic control had only one Ames FT positive sample (Table 12). The peak of the phenanthrene peak remained stable throughout the experiment, suggesting that there was little to no photodegradation, although there may have been enough to cause a low level of mutagenicity.[294] (Fig 20, Table 12)

Table 12 Phenanthrene Ames FT Abiotic Control Data. The values for each day/dilution are the fold increase of the mean number of positive wells (n = 3 replicates) relative to the mean number of positive wells of the negative control (n=3 replicates) for that day. The Day 0 values are the same data as Table 11. The TA98 strain showed a weak positive on day 6, but otherwise there were no positive results. *TA 100 Day 4 negative control was high at 24 mean positive wells.

TA98	Day 0	Day 2	Day 4	Day 6
1	1.5	0.5	0.25	1.94
1/4	1.5	1	0.75	1.19
1/16	1.25	1.5	2.50	1.31
	neg	neg	neg	weak pos

TA100	Day 0	Day 2	Day 4*	Day 6
1	1.07	2.05	1.17	0.70
1/4	1.22	1.62	0.83	0.90
1/16	0.78	1.48	0.68	0.97
	neg	neg	neg	neg

Individual Dilution Result	Ames FT Result
α < 0.05, fold increase > 1.5 =	weak positive
α < 0.1, fold increase >= 2.0 =	weak positive
α < 0.05, fold increase >= 2.0 =	overall positive

Mutagenicity was detected in both strains in the phenanthrene biodegradation culture on day 4, (positive for TA98 and weak positive for TA100) and a weak positive for TA98 was detected on day 2.

4.3.4 Fluorene Degradation

Fluorene was selected as a PAH that was expected to be less likely than phenanthrene to generate mutagenic TPs. The fluorene culture was prepared in exactly the same manner as the phenanthrene cultures with Bushnell Haas media and 0.5% Tween® 80. The fluorene likely did

not fully solubilize as a fine white precipitate was visible, so this culture was diluted with equal volumes of ACN for HPLC analysis to ensure that the HPLC measured the full concentration of fluorene and no suspended fluorene was removed in the filtering step. The culture was run for 8 days and sampled every 2 days as with the phenanthrene culture. Turbidity indicating the growth of bacteria was observed in the degradation culture flask. A slightly green tint was observed in the abiotic control, indicating the presence of the photo-metabolites. (Fig 21)

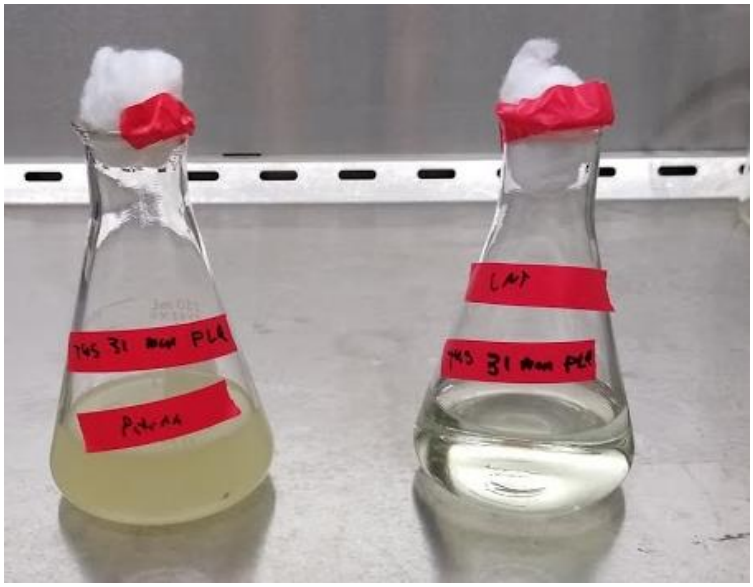


Figure 21 Fluorene degradation culture (left) and abiotic control (right) on day 4 of the degradation run. Turbidity can be seen in the degradation culture, and a faint green hue in the abiotic control.

Both fluorene cultures appeared to photodegrade from fluorene to a photo-metabolite. (Appendix D) This photo-metabolite is most likely 9-hydroxyfluorene or 9-fluorenone, the most common photo-metabolite of fluorene.[5,295] The *Pseudomonas Putida* degradation strain was less adapted to degrading fluorene than phenanthrene, and only reduced the peak to about 70% of the original peak height in 8 days. This reduction may have been due to photodegradation. However,

comparison with the abiotic control suggested that fluorene's first photo-metabolite was degraded by the bacteria. (Fig 22) The photo-metabolite showed a downward trend over the degradation culture, whereas the photo-metabolite showed an upward trend in the abiotic control, demonstrating that the *Pseudomonas Putida* were at least degrading the photo-product, if not also degrading fluorene.[296]

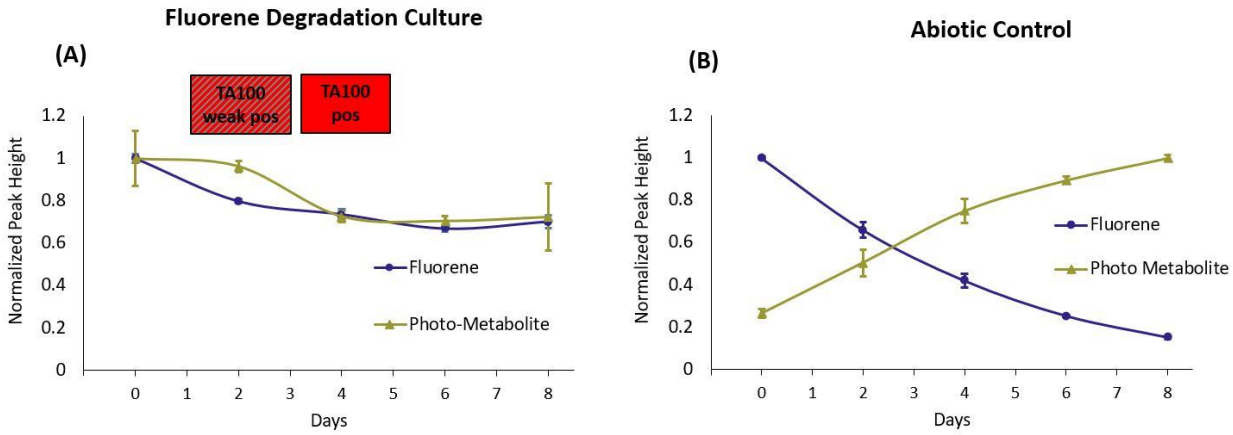


Figure 22 Fluorene degradation culture. Normalized HPLC peak height (absobance at 254 nm) (A) Phenanthrene biodegradation culture with Ames FT results. (B) Fluorene Abiotic Control. The error bars show one standard deviation in both panels (n=3). The subsequent days were normalized to the day 0 mean peak height. Ames FT results are shown in the shaded boxes. No positive Ames FT results were found in the abiotic control. (see attached file PAH_culture_data.xlsx)

Mutagenicity was observed at days 2 and 4 in the TA100 strain, but not in the TA98 strain. (Table 13) This may be indicative of base pair substitution mutations, rather than frameshift mutations, which are shown by TA98.[203]

Table 13 Fluorene Degradation Ames FT Data. The values for each day/dilution are the fold increase of the mean number of positive wells (n = 3 replicates) relative to the mean number of positive wells of the negative control (n=3 replicates) for that day. Highlighting indicates the level of significance and magnitude of fold change; fold changes may be greater than 2 and not be significant at least α < 0.1. *The TA98 degradation culture day 4 positive control was low at 7 mean positive wells. (C) The TA100 degradation culture day 0 negative control was slightly high at 11 wells. (D) Normalized HPLC peak height (254 nm) for the abiotic control and its first photo product. (E) TA100 results for the abiotic control. *The TA100 degradation culture day 8 postitive control was low at 17 mean positive wells. *The TA100 abiotic control day 4 negative control, may have died due to old media. (see attached file Chap4_PAH_culture_data.xlsx)

TA98 Data	Day 0	Day 2	Day 4*	Day 6	Day 8
1	0.57	0.67	0.00	0.5	0.00
1/2	0.00	1.33	0.33	0.25	0.33
1/4	0.71	0.33	0.33	0.25	0.00
1/8	0.00	1.00	1.67	1.25	0.67
1/16	1.14	0.00	0.00	0.25	1.00
1/32	0.71	1.67	0.67	0.5	1.33
	neg	neg	neg	neg	neg

TA100 Data	Day 0*	Day 2	Day 4	Day 6	Day 8*
1	0.88	1.32	2.31	1.41	1.13
1/2	1.03	1.54	1.81	0.94	0.93
1/4	0.61	1.00	2.00	1.00	0.73
1/8	0.76	1.29	1.19	1.12	1.20
1/16	0.76	1.18	1.63	1.00	1.00
1/32	1.09	0.79	1.75	1.29	0.93
	neg	weak pos	pos	neg	neg

Individual Dilution Result	Ames FT Result
α < 0.05, fold increase > 1.5 =	weak positive
α < 0.1, fold increase >= 2.0 =	weak positive
α < 0.05, fold increase >= 2.0 =	overall positive

Mutagenicity was not observed in the abiotic control, indicating that although fluorene was photo-transformed due to the ambient sunlight, the photo transformation products were not mutagenic.[295]

Table 14 Fluorene Ames Ft Abiotic Control Data. The values for each day/dilution are the fold increase of the mean number of positive wells (n = 3 replicates) relative to the mean number of positive wells of the negative control (n=3 replicates) for that day. The Day 0 values are the same data as Table 13. No Ames FT results of either strain met the criteria for positivity. *The TA100 abiotic control day 4 negative control, may have died due to old media. (see attached file Chap4_PAH_culture_data.xlsx)

TA98 Data	Day 0	Day 2	Day 4	Day 6	Day 8
1	0.57	1.00	0.00	0.33	1.00
1/4	0.71	0.40	0.25	0.83	1.33
1/16	1.14	0.60	0.25	0.33	1.33
	neg	neg	neg	neg	neg

TA100 Data	Day 0	Day 2	Day 4*	Day 6	Day 8
1	0.88	1.27	1.80	1.24	1.24
1/4	0.61	0.92	1.20	1.44	1.44
1/16	0.76	1.04	2.40	1.00	1.00
	neg	neg	neg	neg	neg

Individual Dilution Result	Ames FT Result
α < 0.05, fold increase > 1.5 =	weak positive
α < 0.1, fold increase >= 2.0 =	weak positive
α < 0.05, fold increase >= 2.0 =	overall positive

Mutagenicity was only detected in the TA100 strain in the fluorene degradation culture. As with the phenanthrene culture, mutagenicity was not detected (in either strain) at the beginning or end of the culture.

4.3.5 Limitations

There were several limiting factors to this experiment. Because the biodegradation culture is always actively degrading and a reliable method for preserving culture aliquots had not been established, there was no opportunity to repeat Ames FT runs if the positive or negative controls were outside of tolerances. Due to this limitation, all data was retained, but noted as not meeting the validity criteria (see section 4.2.3.3).

Additionally, because no two biodegradation cultures can ever be perfect replicates of each other, even if the preparation methods and time points were standardized, even standard time points are not perfectly identical. Supply and scheduling issues were encountered due to operating during the COVID-19 pandemic and adjustments had to be made which limited the number of pilot experiments that could be conducted to fine tune parameters. Due to the unknown nature of the exact chemical makeup of the samples from the degradation cultures and concerns about stability if the samples were stored, samples were always analyzed immediately rather than preserved for future analysis, which prevented any re-analyzing of samples.

4.4 Conclusions

This experiment required the use of three independent bacterial cultures; one to degrade the PAH (*P. Putida*), and two to monitor the potential occurrence of mutagenicity (TA98 and TA100). As anticipated, this created a high degree of variability and complexity. To guide the experiment, the networks model was used to predict the TPs most likely to be generated by a degradation culture and the QSAR was used to classify them as likely mutagens or likely non-

mutagens. These models provide only a rough estimate of the relative likelihood of mutagenicity based on all possible TPs that may be generated by a given PAH. By using a single degrading bacterium strain only some of the potential degradation pathways will be active. Furthermore, even if the degradation pathway is known or can be estimated, the kinetics are impossible to predict. Some steps may proceed much faster than others and the bacteria may be unable to degrade the TPs after a certain point. Despite these limitations, the phenanthrene culture samples (Tables 11 and 12) showed positive for mutagens with the Ames FT in both strains, whereas the fluorene culture (Tables 13 and 14) only showed positive for mutagens in the TA100 strain, and all cultures had mutagenicity occur in the middle of the degradation time, but not at the beginning or the end, which fits with our hypothesis that phenanthrene would generate more mutagens than fluorene and that they would occur part way through the degradation chain. The prediction of the likelihood of detecting mutagenic TPs from the combination of the networks model, based on a pathway predictor, and the QSAR, based on empirical training data, may prove a valuable tool for future experiments evaluating the toxic impact of dynamic and partially unknown degradation chains. In the case of phenanthrene available literature suggests that some of the biodegradation TPs might be mutagenic and this is also supported by the work herein. Fluorene may form TPs that are mutagenic to the TA100 strain that are not currently identified in the literature. As the field of bioremediation for PAHs and even other compounds based on carbon backbones continues to improve and more degraders are identified, this same approach could be applied to identify the portions of highest concern for mutagenicity in the degradation network. QSARs for other toxicity endpoints could also be applied to expand the hazard analysis. Based on the estimate of where in the network the most hazardous compounds would be expected, appropriate hazard monitoring and mitigation measures could be applied.

5.0 Conclusions and Future Work

Risk is a function of the likelihood of a negative event occurring and the severity of the event should it occur. In the real world under time and resource constraints, the risk assessor aims to identify the highest severity and highest likelihood negative outcomes to then design and enact appropriate mitigation measures against these possibilities. This work developed tools to predict each component of risk for PAH biodegradation TP mutagenicity and evaluated them in a benchtop experiment.

5.1 Summary and Conclusions

The hazard assessment of PAH environmental degradation products is a complex and widespread problem. Biological degradation is the most effective mechanism of mineralization and ultimate removal of these contaminants from the environment.[59,297] However, bacteria degrade contaminants in stages, creating a vast array of potential TPs,[175] which have their own unique toxicological properties.[37,95] In general, PAHs are transformed in surface soil and water through ring-opening reactions, functionalization with various oxygen groups, and cleavage of carbon chains. These degradation processes result in TPs with lower narcotic toxicity, which can be estimated based on K_{ow}.[75–77,298] Although the polar oxygen groups lower the K_{ow} value, they also provide opportunities for DNA adduction, creating the potential for mutagenicity.[20,85,94,221] However, not all TPs are mutagenic and some that are mutagenic may emerge and diminish over the course of the degradation chain. This results in a constantly changing hazard profile of a

multitude of TPs. The variability in the nature of the toxicity of the various TPs causes changing impacts on the environment and makes it difficult to develop an accurate hazard profile of the partially degraded contamination.

Tools have been developed to address the challenge of predicting environmental transformations. However, comprehensive evaluation generates an intractable number of possible TPs.[144,175] The networks model developed in this work provides a tool for refining the thousands of predicted potential intermediates from a given PAH to a manageable dataset with the goal of narrowing the compounds to focus experimental or computational efforts. While the network model does not explicitly provide information on kinetics, it does provide the order in which the TPs would be anticipated to occur along each individual pathway. The ability to predict the most likely TPs to result from biodegradation provides a first step towards understanding their potential toxic impact. Once the likely structures are known, they can be evaluated through targeted empirical or computational approaches for their potential toxic impacts.

As PAH's degrade into TPs, their narcotic toxicity tends to decrease, but their mutagenic potential increases.[89,133] Several mutagenicity classifiers exist;[190,191] however, they are designed to reflect both direct-acting mutagenicity and mutagenicity that only occurs after metabolic transformation in higher organisms, which are not reflective of the mutagenicity that can occur due to environmental microbial biodegradation. Additionally, many existing tools are designed to reflect a broad range of compounds and are not sensitive to the subtle variations between different PAH TPs. For example, structural-alert based mutagenicity classifiers are often not designed for PAHs and may contain only a few relevant groups such as the epoxide group to alert for mutagenicity.[299,300] This gap in the available tools motivated the development of the mutagenicity QSAR. The QSAR's performance was further improved by sub-dividing the dataset based on

unsupervised clustering using features of the molecular structure, which suggested a dominant mechanism within each cluster.

The QSAR developed in Chapter 3 provides an example of a useful approach for a specific hazard assessment application (environmental PAH TPs). The performance improvement of this QSAR, designed for non-enzymatically activated mutagenicity (direct-acting mutagenicity), over QSARs predicting both enzymatically activated and direct-acting mutagenicity when evaluating just for the latter shows that there is a difference in the structural characteristics of direct-acting mutagens and enzymatically activated mutagens. This highlights the importance of studying direct-acting mutagens, and additional QSARs covering broader applicability domains are needed for other environmentally relevant, potentially mutagenic compounds. Future Ames test databases should take care to separate mutagenicity induced by S9 activation from direct-acting mutagenicity. This will enable better computational tools for all types of environmental transformations that can result in mutagenic TPs, including atmospheric compounds. Advanced machine learning tools based on these datasets may consider providing an option for the user to specify if the tool should predict mutagenicity with or without S9 activation, rather than combining the two into one result.

The networks model and the QSAR were designed to provide insight into the potential occurrence of mutagenicity in PAH environmental degradation. Degrading a PAH in a controlled manner and measuring for mutagenicity enabled evaluation of their potential utility. The batch degradation culture experiment provides initial insight into evaluating mutagenicity over time in an actively degrading culture, using the predictions from the computational tools to target the sampling time points. The mutagenicity of PAH degradation has garnered interest in the

environmental toxicity field over the last few years[20,89,93–95] and improved methods of evaluating mutagenicity in PAH degradation will continue to be needed.

5.2 Future work

PAHs are major current and historical environmental contaminants. Substantial work remains to more fully comprehend their transformations in the environment, appropriately evaluate their hazards, including mutagenicity that can result from PAH biodegradation. Ultimately, the collective findings are intended to inform appropriate mitigation approaches to control the array of potential hazards. This research has contributed several tools towards this goal. Furthermore, the overall approach is transferable to the environmental transformation networks of numerous other environmental containments for which (1) the degradation pathways are known or can be predicted with a tool such as EAWAG-PPS, and (2) an appropriate QSAR for the toxic mechanism of interest exists or can be developed from available empirical data.

One valuable expansion of this work would be to evaluate the mutagenic potential from the biodegradation of larger 4-6 ring PAHs. The pilot experiments herein included attempting to degrade pyrene, fluoranthene, and chrysene with the *Pseudomonas Putida* degradation strain. However, the degradation ability for this bacterial strain appears to be limited to 3-ring PAHs and smaller. In general, larger PAHs are more resistant to photo and biological degradation,[301–303] and these PAHs are also an important component of PAH environmental contamination. For example, the PAH of primary carcinogenic concern to humans, benzo[a]pyrene,[304] contains 5 rings, and one third of mutagenic compounds from the training data for the QSAR developed in Chapter 3 contain 4 or 5 rings. Experiments to degrade the larger PAHs would likely be difficult and might take

months to complete rather than 1-2 weeks as with flask cultures degrading the smaller PAHs. The degradation networks are also proportionally larger, but still well within the capabilities of the computational tools developed in Chapters 2 and 3.

Another expansion of this work would be to analyze PAHs beyond the 16 PAHs on the EPA's Priority Pollutant list, as there may be PAHs which degrade into mutagenic TPs that have not been identified yet. The 16 priority PAHs are representative structures, but are by no means an exhaustive list.[170] The networks model and the QSAR provide powerful tools for estimating the potential hazards of any compound that can be predicted by EAWAG-PPS. There are numerous potential PAH structures beyond those 16 priority PAHs, and a theoretical study based on the computational tools developed in this work could provide an expanded dataset of PAHs likely to degrade into mutagens, even if no empirical data exist yet for these other PAHs.

Estimates of the active pathways in a particular experiment could be potentially improved as more TPs are identified. If TPs were found that were unique to a particular section of a network, these could be used as "checkpoints" and threshold quantities relative to the original PAH concentration could be identified along a degradation chain. Checkpoints would be unique identifiable compounds which would indicate that the degradation had taken a certain pathway (such as the pathways indicated by the blue arrows in Figures 18 and 19). This would allow rough estimates of which pathways are active and how quickly the culture proceeds to each checkpoint, even if the steps between them cannot always be known.

An additional improvement would be to use the pour-plate Ames Test. This test is more difficult and resource intensive (cost and time) to conduct than the FT method. While the pour-plate design is not necessarily more sensitive, it would be easier to evaluate whether a result that was close to the threshold for a positive mutagenic response was due to the sample being only

weakly mutagenic or to a low response of the indicator strain itself. The pour-plate method provides a numerical result rather than simply a positive or negative result as the with the FT test, which would be beneficial in trying to detect more subtle increases in mutagenicity. It is also easier to evaluate cytotoxicity in the pour-plate design, which would make it possible to work at higher sample concentrations as it would be possible to determine if a negative reading was the result of cytotoxicity rather than non-mutagenicity.

Samples were always analyzed immediately in this work due to the concerns of auto-oxidation and interaction with other TPs. It would be beneficial in future work to develop a suitable method of preserving samples such that they could be analyzed at the completion of the degradation experiment, rather than in real-time. This would allow a higher number of samples to be analyzed than are feasible analyzing purely in real-time, as was done under the current design. The Ames Test requires 3 days before results can be read, so results are not available in real time under any design, and performing the analysis using preserved samples would allow repeating the Ames Test for samples that resulted in inconclusive results due to either the positive or negative controls being outside of tolerances. The Ames FT could also be used as a screening assay and the pour-plate version could be used as a follow up for confirmation. The first samples to be analyzed could still be targeted based on the networks model and the QSAR, however more samples could also be collected than would initially be analyzed. Samples before or after one of the initial samples that resulted in a positive on the Ames FT could also be analyzed to determine if nearby time points were also mutagenic. This would allow a more detailed analysis on the portion of the degradation timeline believed to contain the mutagens with the benefit of hindsight from the initial round of analysis, rather than only trying to predict the point in the degradation timeline where the mutagens would occur. Preserving samples would likely also improve the repeatability of

biological replicates. Biodegradation cultures do not have consistent kinetics, so it is difficult to sample the same time point in biological replicates based off of pre-selected time points. A preservation method would need to preserve the exact structure of the TPs, avoiding auto-oxidation that could change the structure, ensuring that the culture would retain the same degree of mutagenicity (see Chapter 4). The success of a preservation method would be difficult to evaluate as the Ames Test contains some intrinsic variation in its repeatability, so many replicates would likely be needed.[305]

With numerous possible sources and substantial historical contamination, PAH environmental contamination will be an ongoing problem for many years to come. The potential hazards of TPs are a key component of PAH impacts on the environment. The combination of computational predictive power and experimental verification undertaken in this work will enable better understanding of the ecological hazards of contaminated environments, the impacts of natural and guided biodegradation, and better overall environmental stewardship.

Disclaimer:

The views expressed in this dissertation are those of the author and do not reflect the official policy or position of the United States Air Force, Department of Defense, or the U.S. Government

Appendix A Summary of Attached Files

The following files accompany this dissertation as attachments, as noted, some are reproduced in accordance with the permissions granted in the American Chemical Society's Journal Publishing Agreement

Appendix A.1.1 Supporting Information for Chapter 2

Reproduced with permission from Environmental Science and Technology, Sleight, T.; Khanna, V.; Gilbertson, L.; Ng, C. Network Analysis for Identifying High Impact Biodegradation Metabolites: A PAH Case Study. Environ. Sci. Technol. Rev. 2020. Copyright 2021, American Chemical Society

- Chap2_networks_analysis_pathmatches.xlsx: SMILES codes of the Pathway Matches
- Chap2_networks_analysis_SI.docx: Networks Analysis Supporting Information; Additional details and similar plots for PAHs not shown in the primary text, full empirical literature review for each PAH including all references, images of structures, and SMILES codes, and summary of the python scripts and the process for data analysis and the sensitivity analysis

Appendix A.1.2 Supporting Information for Chapter 3

-Chap3_classification_model_SI.docx: Detailed plots of clustering (S1.1-S1.6) Expanded plots of descriptor selection, correlation plots, confusion matrices and variance inflation factor iterative removal (S2.1-S2.15) Swarm plots of differences between clusters (S3) PaDEL descriptor details, weightings and equations (S4-S5) Representative structures for each cluster (S6)

- Chap3_assigned_priority.xlsx: Assigned Priority for Closely Correlated Descriptors

Appendix A.1.3 Supporting Information for Chapter 4

- Chap4_PAH_culture_data.xlsx Raw data for Ames Tests and HPLC measurements

Appendix B Detailed Biodegradation Culture Protocol

Appendix B.1.1 PAH Degradation Strain

Pseudomonas Putida (Trevisan) ATCC 17484, a known naphthalene, anthracene, and phenanthrene degrader, was purchased from ATCC. The bacteria were rehydrated according to the suppliers in instructions and freezer stock was prepared with 25% glycerol. The stock was maintained at -80 ºC, and never thawed except to scrape ice off the surface as described below.

Appendix B.1.2 Freezer Extraction Procedure

When a new culture of *Pseudomonas Putida* was needed, a tube of frozen cell strain was quickly removed from the freezer and held in a freezer container filled with frozen ethanol to prevent the tube from thawing. A small amount of frozen culture was scrapped off of the top with a sterile pipette and then swirled in a 250 mL shaker flask of BD Difco 234000 media, inside a sterile hood. The freezer container was then immediately returned to the -80 ºC freezer. The full extraction took approximately 30 seconds. The shaker flask was then placed on the shaking incubator at 28 ºC and approximately 120 rpm.

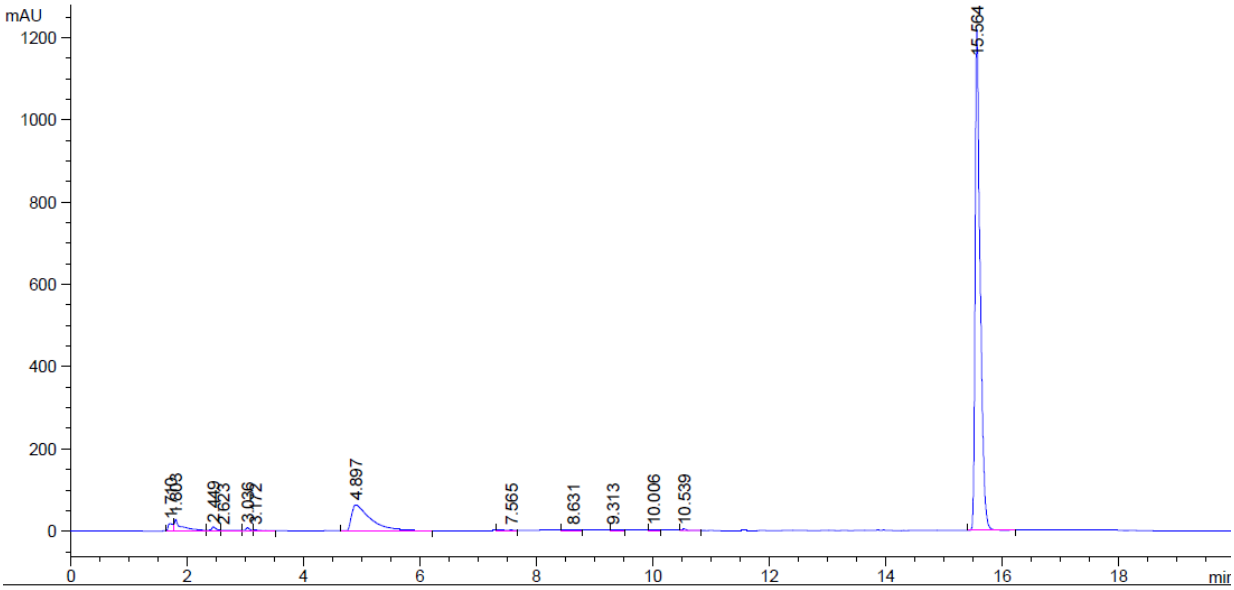
Appendix B.1.3 High Performance Liquid Chromatography Procedure

HPLC details: The reverse phase HPLC used a 20-minute custom method with a mixture of acetonitrile (ACN) and DI water. The method started with a 3-minute hydrophilic hold at 90%

water/10% ACN to flush out the salts from the BH media. Next a 10-minute gradient slowly transitioned to 20% water/80% ACN. Some TPs would typically elute during this gradient. Once at 20% water/80% ACN, this was held for 4 minutes The PAHs would elute during this time. The final step was a gradient back to 90% water/10% ACN to prepare for the next run and clean out any residual TPs.

Blanks were run at the beginning of a set of analyses and a conditioning sequence was run at the end in order to maintain the health of the column.

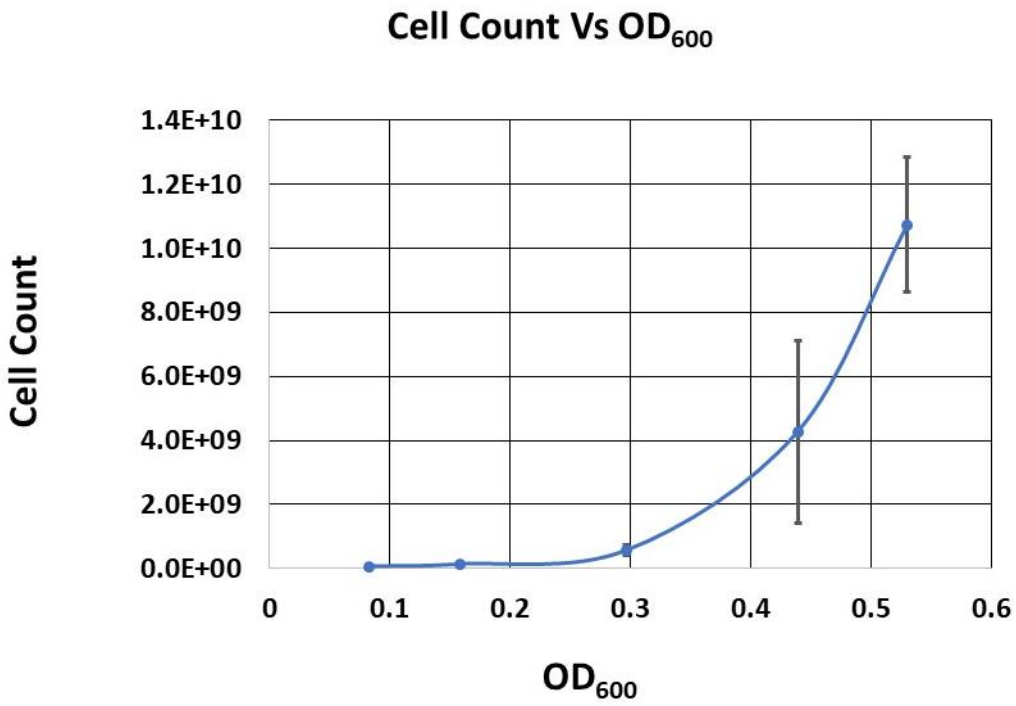
Ames FT supplies were purchased from Molecular Toxicology™, Inc.



Appendix Figure 1 HPLC Sample Phenanthrene Chromatogram. The peak at 15.564 minutes is phenanthrene and the peak at 4.897 is a transformation product, most likely either 2-hdyroxy-1-naphthoic acid or 1-hydrox-2- naphthoic acid.

Appendix B.1.4 Pseudomonas Putida Enrichment Culture

In order to be able to inoculate the PAH degradation cultures at an appropriate initial cell concentration, the *Pseudomonas Putida (P. Putida)* was precultured in BD Difco Media.

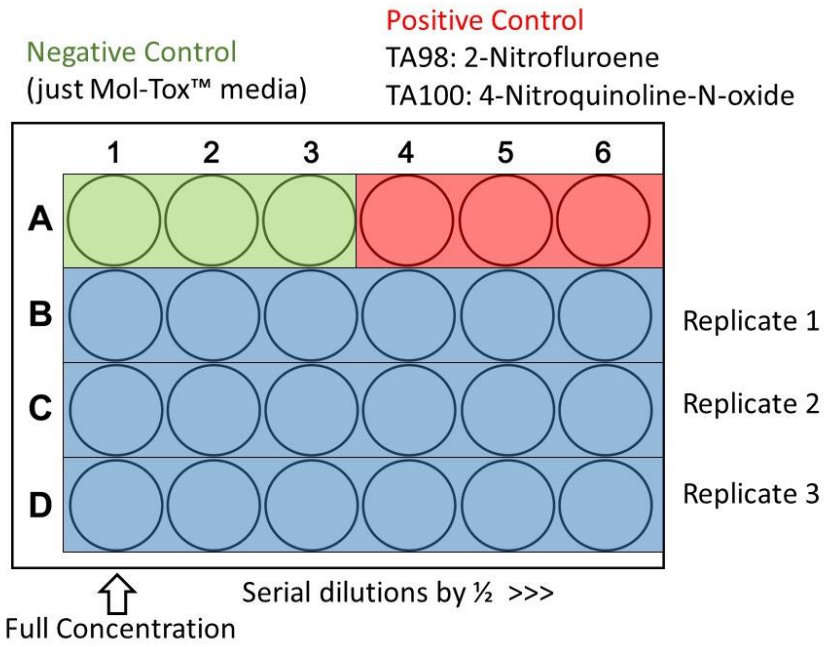


Appendix Figure 2 OD_{600} calibration curve

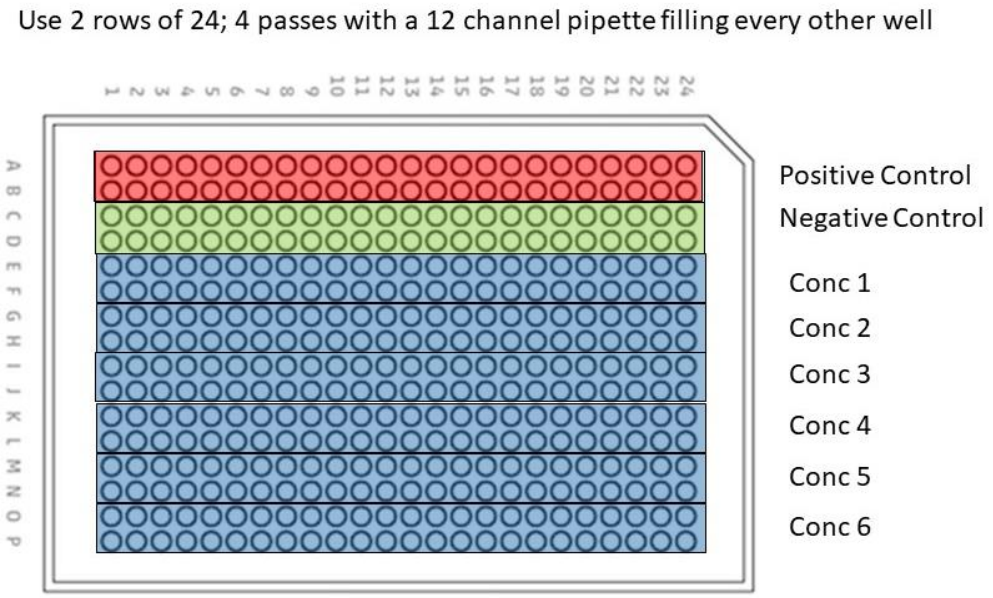
Appendix B.1.5 Ames Test Procedure.

A 16-channel pipette is preferred for pipetting into a 384 well plate. However, the largest multi-channel pipette available in the University of Pittsburgh Civil and Environmental Engineering laboratory at the time of this experiment was a 12-channel pipette. Alternate-well pipette was used

in order to overcome this limitation. In order to minimize the potential for errors, a different well mapping was adopted from the mapping suggested by the Ames Test supplier. All other procedures were as described in the Mol-Tox™ manual *Ames FT™ Mutagenicity Test Kit Instruction Manual 31-300*. S9 activation was not used. Pipette tips and reservoirs were re-used on the same replicate, starting with the most dilute (1/32) and proceeding to the least dilute (1), so that any carry-over would only result in a very small amount of media from a lower dilution being mixed with a higher dilution. A fresh reservoir and pipette were used for the negative control and tips and reservoir were also re-used from the negative control to the positive control; any carry over would have resulted in a negligible dilution of the positive control.



Appendix Figure 3 Ames Test 24-Well Plate Mapping



Appendix Figure 4 Ames Test 384 Well Plate Mapping

Appendix B.1.6 Safety

All procedures were conducted in accordance with the University of Pittsburgh's Chemical Hygiene safety protocols. Aseptic procedure was maintained throughout. Biosafety hoods were sterilized with 20 minutes of UV and 70% ethanol. A Biosafety Level 2 room and dedicated equipment was used for the Salmonella strains of the Ames Test.

Appendix B.1.7 Supply List

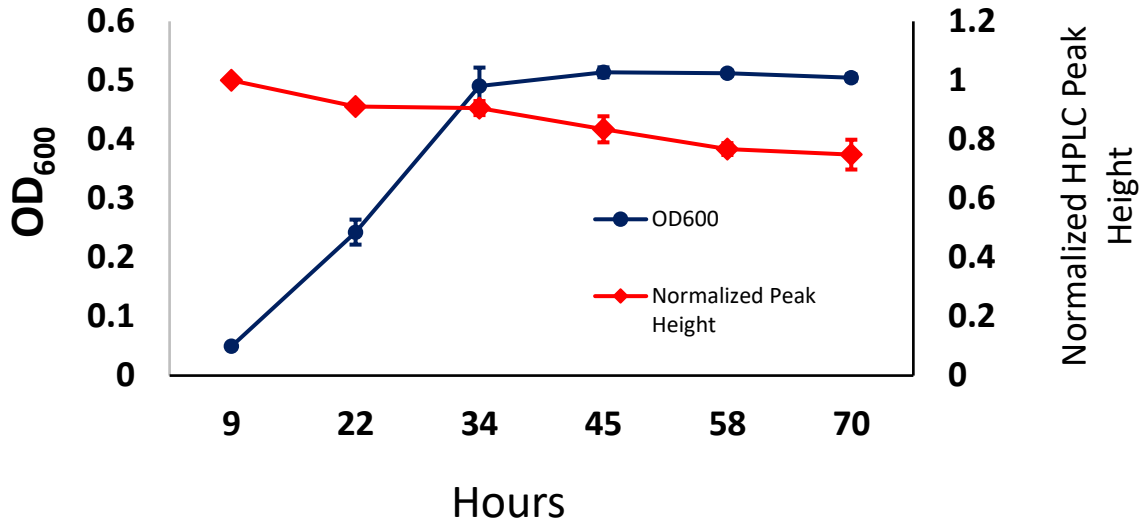
Appendix Table 1 Appendix Media Products and Suppliers

Product	**Supplier**
Becton, Dickinson and Company (BD) Difco™ 234000 Media	Fisher Scientific
Bushnell-Haas Media	Fisher Scientific
Tween 80	Fisher Scientific
Phenanthrene and Fluorene	Sigma-Aldrich
Acetonitrile	Fisher Scientific

Appendix C Phenanthrene Pilot Work

Appendix C.1 Phenanthrene Degradation Pilot Experiments

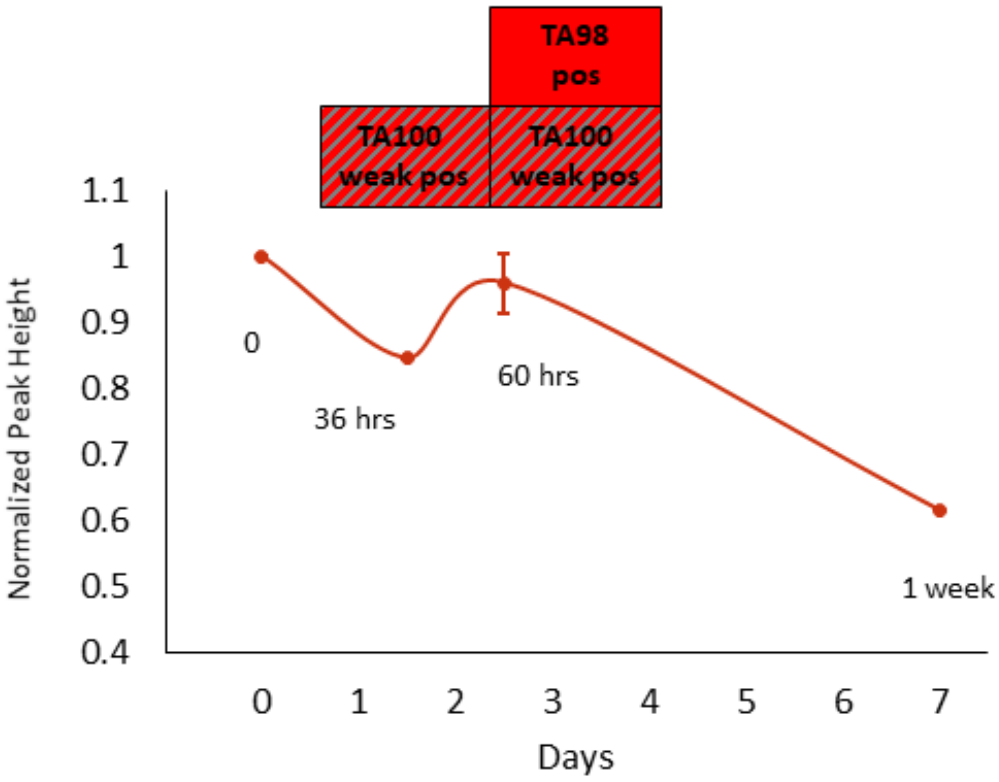
Several pilot experiments were conducted with phenanthrene to establish the protocol for biodegradation and conducting Ames FTs on an actively degrading culture. One of the early studies was to compare OD_{600} readings to the HPLC data. Three identical cultures were started simultaneously as biological replicates. The media and inoculation procedures were as described in 4.2.2, except that the PAH culture flasks were not placed on the shaker ahead of inoculation. This experiment was part of what determined the need to place the flasks on the incubator ahead of inoculation to fully dissolve the PAH. An existing supply of phenanthrene manufactured by Kodak from the Civil and Environmental Engineering Chemical Stock room was used for these early experiments.



Appendix Figure 5 OD_{600} readings and phenanthrene peak height. 3 biological replicates were used. The error bars represent 1 standard deviation. The HPLC values are normalized to the 9 hr time point. A small increase in the phenanthrene peak height was seen at 9 hrs relative to the media baseline, (data not shown) most likely due to phenanthrene solubilizing due to the shaking motion of the incubator.

The OD_{600} of a freshly inoculated phenanthrene culture climbs for 1-2 days and the phenanthrene does not begin to degrade significantly until the OD_{600} reaches the stationary phase and the Pseudomonas Putida begin to degrade the phenanthrene. (Appendix Figure 5)

Appendix C.2 Phenanthrene Mutagenicity Pilot Experiment



Appendix Figure 6 Normalized HPLC peak height (254 UV) for phenanthrene in the degradation culture. The error bars show one standard deviation.

In the TA100 strain, weak positive responses were observed at 36 hrs and 60 hrs. In the TA98 strain, a positive response was observed at the 60-hour time point. (Appendix Table 2) Some fluctuation is observed in the HPLC data due to the uncertainty in sampling a live culture. It is possible that the sample for the HPLC 36-hr time point was allowed to sit for too long off of the shaking incubator, resulting in an artificially low HPLC reading. Replicates for the HPLC data were only collected at the 60 hr time point. Based on the occurrence positive Ames Test are 36 and 60 hrs in this pilot study, the 2-day interval was selected for further experiments.

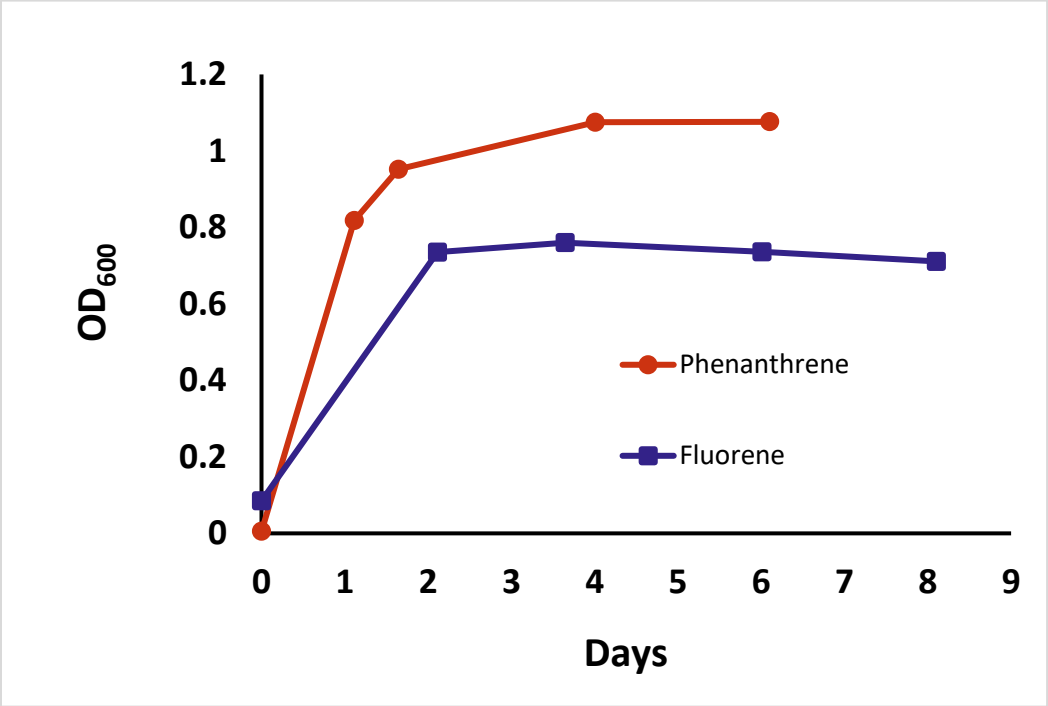
Appendix Table 2 TA98 and TA100 results from the phenanthrene degradataion pilot study. *The TA100 Day 0 had a high negative control baseline of a mean of 28 positive wells which would have made it difficult to detect a mutagenic response.

TA98 Data	0	36 hrs	60 hrs	1 week
1	1.75	0.90	5.00	1.29
1/2	1.25	0.40	2.33	0.53
1/4	1.25	0.60	1.17	0.59
1/8	1.50	0.50	1.00	0.35
1/16	1.13	0.50	0.67	1.24
1/32	0.75	1.00	1.17	0.59
	neg	**neg**	**pos**	**neg**

TA100 Data	0*	36 hrs	60 hrs	1 week
1	0.72	2.65	1.83	1.17
1/2	0.58	1.95	1.70	0.98
1/4	0.55	1.5	1.78	0.94
1/8	0.36	1.1	1.30	1.06
1/16	0.52	1.75	1.30	0.87
1/32	0.42	1.55	1.04	0.66
	neg	**weak pos**	**weak pos**	neg

Individual Dilution Result	Ames FT Result
α < 0.05, fold increase > 1.5 =	**weak positive**
α < 0.1, fold increase >= 2.0 =	**weak positive**
α < 0.05, fold increase >= 2.0 =	**overall positive**

Appendix C.3 OD600 Plot of phenanthrene and fluorene experiment

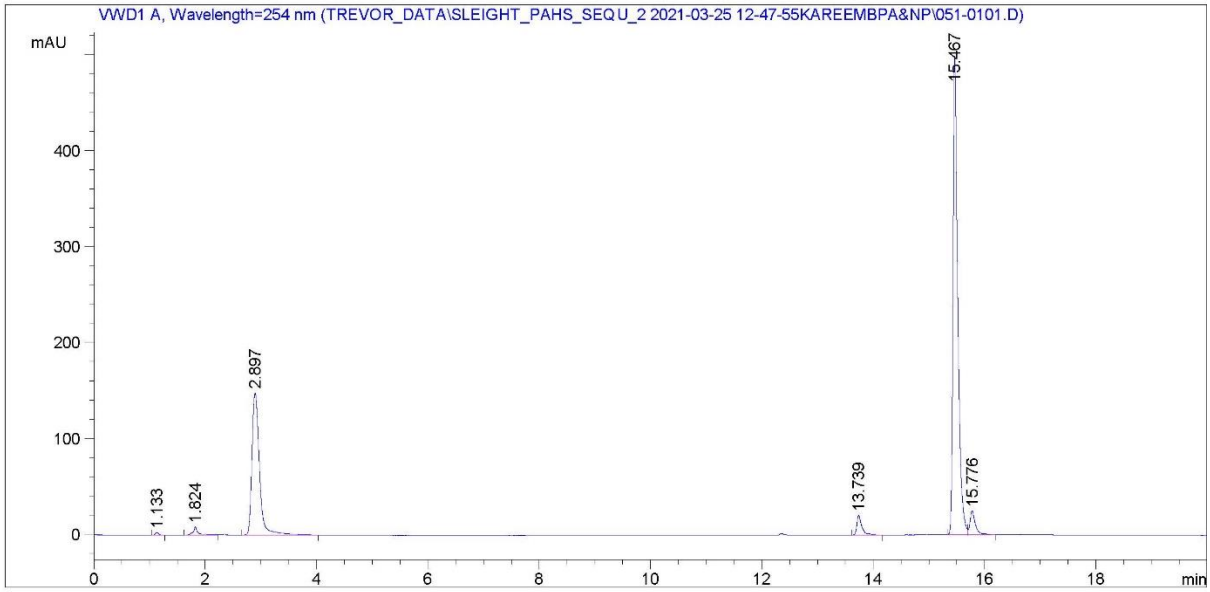


Appendix Figure 7 OD_{600} timelines of phenanthrene and fluorene degradation cultures.

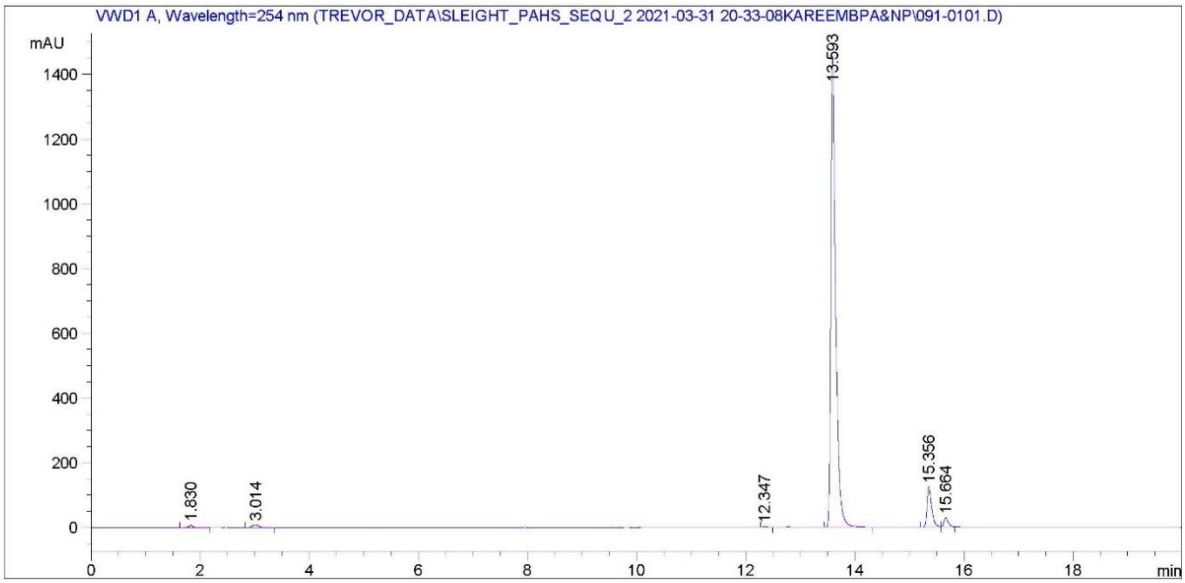
The OD_{600} timelines of the Phenanthrene and Fluorene Cultures presented in section 4.3 follow a similar pattern to the pilot studies, reaching the terminal OD_{600} at about 2 days. Thus the 2 and 4 day time points should be in the early and mid parts of the PAH degradation curve, respectively.

Appendix D Selected HPLC Data

Appendix D.1 Fluorene



Appendix Figure 8 Fluorene Baseline. The peak at 2.897 is a media artifact. The highest peak at 15.467 is the fluorene peak. The peak at 13.739 was tenatively identified as 9-hydroxy fluroene.



Appendix Figure 9 Fluorene abiotic control after 6 days. The 13.593 peak is now the highest peak on the chromatogram, and the phenanthrene peak at 15.664 is significantly reduced.

Bibliography

(1) Zhang, Y.; Tao, S. Global Atmospheric Emission Inventory of Polycyclic Aromatic Hydrocarbons (PAHs) for 2004. *Atmos. Environ.* **2009**, *43* (4), 812–819. https://doi.org/10.1016/j.atmosenv.2008.10.050.

(2) Shen, H.; Huang, Y.; Wang, R.; Zhu, D.; Li, W.; Shen, G.; Wang, B.; Zhang, Y.; Chen, Y.; Lu, Y.; Chen, H.; Li, T.; Sun, K.; Li, B.; Liu, W.; Liu, J.; Tao, S. Global Atmospheric Emissions of Polycyclic Aromatic Hydrocarbons from 1960 to 2008 and Future Predictions. *Environ. Sci. Technol.* **2013**, *47* (12), 6415–6424. https://doi.org/10.1021/es400857z.

(3) Taranina, O. A.; Burkat, V. S.; Volkodaeva, M. V. Analysis of the Concentration of Gas-Phase and Solid-Phase Polyaromatic Hydrocarbons in Industrial Emissions from Aluminum Production. *Metallurgist* **2020**, *63* (11), 1227–1236. https://doi.org/10.1007/s11015-020-00945-6.

(4) Li, R.; Hua, P.; Zhang, J.; Krebs, P. Characterizing and Predicting the Impact of Vehicular Emissions on the Transport and Fate of Polycyclic Aromatic Hydrocarbons in Environmental Multimedia. *J. Clean. Prod.* **2020**, *271*, 122591. https://doi.org/10.1016/j.jclepro.2020.122591.

(5) Gbeddy, G.; Goonetilleke, A.; Ayoko, G. A.; Egodawatta, P. Transformation and Degradation of Polycyclic Aromatic Hydrocarbons (PAHs) in Urban Road Surfaces: Influential Factors, Implications and Recommendations. *Environ. Pollut.* **2020**, *257*. https://doi.org/10.1016/j.envpol.2019.113510.

(6) Mitra, T.; Chu, C.; Naseri, A.; Thomson, M. J. Polycyclic Aromatic Hydrocarbon Formation in a Flame of the Alkylated Aromatic Trimethylbenzene Compared to Those of the Alkane Dodecane. *Combust. Flame* **2021**, *223*, 495–510. https://doi.org/10.1016/j.combustflame.2020.10.015.

(7) Salat, A. P. J.; Eickmeyer, D. C.; Kimpe, L. E.; Hall, R. I.; Wolfe, B. B.; Mundy, L. J.; Trudeau, V. L.; Blais, J. M. Integrated Analysis of Petroleum Biomarkers and Polycyclic Aromatic Compounds in Lake Sediment Cores from an Oil Sands Region. *Environ. Pollut.* **2021**, *270*, 116060. https://doi.org/10.1016/j.envpol.2020.116060.

(8) Chen, A.; Wu, X.; Simonich, S. L. M.; Kang, H.; Xie, Z. Volatilization of Polycyclic Aromatic Hydrocarbons (PAHs) over the North Pacific and Adjacent Arctic Ocean: The Impact of Offshore Oil Drilling. *Environ. Pollut.* **2021**, *268*, 115963. https://doi.org/10.1016/j.envpol.2020.115963.

(9) Kim, M.; Kennicutt, M. C.; Qian, Y. Source Characterization Using Compound Composition and Stable Carbon Isotope Ratio of PAHs in Sediments from Lakes, Harbor,

and Shipping Waterway. *Sci. Total Environ.* **2008**, *389* (2), 367–377. https://doi.org/10.1016/j.scitotenv.2007.08.045.

(10) Morillo, E.; Romero, A. S.; Madrid, L.; Villaverde, J.; Maqueda, C. Characterization and Sources of PAHs and Potentially Toxic Metals in Urban Environments of Sevilla (Southern Spain). *Water. Air. Soil Pollut.* **2008**, *187* (1), 41–51. https://doi.org/10.1007/s11270-007-9495-9.

(11) Yan, B.; Abrajano, T. A.; Bopp, R. F.; Benedict, L. A.; Chaky, D. A.; Perry, E.; Song, J.; Keane, D. P. Combined Application of Δ13C and Molecular Ratios in Sediment Cores for PAH Source Apportionment in the New York/New Jersey Harbor Complex. *Org. Geochem.* **2006**, *37* (6), 674–687. https://doi.org/10.1016/j.orggeochem.2006.01.013.

(12) Zakaria, M. P.; Takada, H.; Tsutsumi, S.; Ohno, K.; Yamada, J.; Kouno, E.; Kumata, H. Distribution of Polycyclic Aromatic Hydrocarbons (PAHs) in Rivers and Estuaries in Malaysia: A Widespread Input of Petrogenic PAHs. *Environ. Sci. Technol.* **2002**, *36* (9), 1907–1918. https://doi.org/10.1021/es011278+.

(13) Miura, K.; Shimada, K.; Sugiyama, T.; Sato, K.; Takami, A.; Chan, C. K.; Kim, I. S.; Kim, Y. P.; Lin, N.-H.; Hatakeyama, S. Seasonal and Annual Changes in PAH Concentrations in a Remote Site in the Pacific Ocean. *Sci. Rep.* **2019**, *9* (1), 12591. https://doi.org/10.1038/s41598-019-47409-9.

(14) Cao, Y.; Lin, C.; Zhang, X.; Liu, X.; He, M.; Ouyang, W. Distribution, Source, and Ecological Risks of Polycyclic Aromatic Hydrocarbons in Lake Qinghai, China. *Environ. Pollut.* **2020**, *266*, 115401. https://doi.org/10.1016/j.envpol.2020.115401.

(15) Roslund, M. I.; Gronroos, M.; Rantalainen, A. L.; Jumpponen, A.; Romantschuk, M.; Parajuli, A.; Hyoty, H.; Laitinen, O.; Sinkkone, A. Half-Lives of PAHs and Temporal Microbiota Changes in Commonly Used Urban Landscaping Materials. *Peerj* **2018**, *6*. https://doi.org/10.7717/peerj.4508.

(16) Oleszczuk, P.; Baran, S. Degradation of Individual Polycyclic Aromatic Hydrocarbons (PAHs) in Soil Polluted with Aircraft Fuel. *Pol. J. Environ. Stud.* **2003**, *12* (4), 431–437.

(17) Hussar, E.; Richards, S.; Lin, Z.-Q.; Dixon, R. P.; Johnson, K. A. Human Health Risk Assessment of 16 Priority Polycyclic Aromatic Hydrocarbons in Soils of Chattanooga, Tennessee, USA. *Water. Air. Soil Pollut.* **2012**, *223* (9), 5535–5548. https://doi.org/10.1007/s11270-012-1265-7.

(18) Syed, J. H.; Iqbal, M.; Zhong, G.; Katsoyiannis, A.; Yadav, I. C.; Li, J.; Zhang, G. Polycyclic Aromatic Hydrocarbons (PAHs) in Chinese Forest Soils: Profile Composition, Spatial Variations and Source Apportionment. *Sci. Rep.* **2017**, *7* (1), 2692. https://doi.org/10.1038/s41598-017-02999-0.

(19) Liu, L.-Y.; Wang, J.-Z.; Wei, G.-L.; Guan, Y.-F.; Wong, C. S.; Zeng, E. Y. Sediment Records of Polycyclic Aromatic Hydrocarbons (PAHs) in the Continental Shelf of China:

Implications for Evolving Anthropogenic Impacts. *Environ. Sci. Technol.* **2012**, *46* (12), 6497–6504. https://doi.org/10.1021/es300474z.

(20) de Souza Pohren, R.; Rocha, J. A. V.; Horn, K. A.; Vargas, V. M. F. Bioremediation of Soils Contaminated by PAHs: Mutagenicity as a Tool to Validate Environmental Quality. *Chemosphere* **2019**, *214*, 659–668. https://doi.org/10.1016/j.chemosphere.2018.08.020.

(21) Ames, B. N.; Lee, F. D.; Durston, W. E. An Improved Bacterial Test System for the Detection and Classification of Mutagens and Carcinogens. *Proc. Natl. Acad. Sci. U. S. A.* **1973**, *70* (3), 782–786. https://doi.org/10.1073/pnas.70.3.782.

(22) Proudlock, R.; Evans, K. The Micro-Ames Test: A Direct Comparison of the Performance and Sensitivities of the Standard and 24-Well Plate Versions of the Bacterial Mutation Test. *Environ. Mol. Mutagen.* **2016**, *57* (9), 687–705. https://doi.org/10.1002/em.22065.

(23) Simcik, M. F.; Eisenreich, S. J.; Lioy, P. J. Source Apportionment and Source/Sink Relationships of PAHs in the Coastal Atmosphere of Chicago and Lake Michigan. *Atmos. Environ.* **1999**, *33* (30), 5071–5079. https://doi.org/10.1016/S1352-2310(99)00233-2.

(24) Hwang, H.-M.; Wade, T. L.; Sericano, J. L. Concentrations and Source Characterization of Polycyclic Aromatic Hydrocarbons in Pine Needles from Korea, Mexico, and United States. *Atmos. Environ.* **2003**, *37* (16), 2259–2267. https://doi.org/10.1016/S1352-2310(03)00090-6.

(25) Singh, D. P.; Gadi, R.; Mandal, T. K.; Saud, T.; Saxena, M.; Sharma, S. K. Emissions Estimates of PAH from Biomass Fuels Used in Rural Sector of Indo-Gangetic Plains of India. *Atmos. Environ.* **2013**, *68*, 120–126. https://doi.org/10.1016/j.atmosenv.2012.11.042.

(26) Tobiszewski, M.; Namieśnik, J. PAH Diagnostic Ratios for the Identification of Pollution Emission Sources. *Environ. Pollut.* **2012**, *162*, 110–119. https://doi.org/10.1016/j.envpol.2011.10.025.

(27) Lin, Y.; Ma, Y.; Qiu, X.; Li, R.; Fang, Y.; Wang, J.; Zhu, Y.; Hu, D. Sources, Transformation, and Health Implications of PAHs and Their Nitrated, Hydroxylated, and Oxygenated Derivatives in PM2.5 in Beijing. *J. Geophys. Res.* **2015**, *120* (14), 7219–7228. https://doi.org/10.1002/2015JD023628.

(28) Tomaz, S.; Shahpoury, P.; Jaffrezo, J. L.; Lammel, G.; Perraudin, E.; Villenave, E.; Albinet, A. One-Year Study of Polycyclic Aromatic Compounds at an Urban Site in Grenoble (France): Seasonal Variations, Gas/Particle Partitioning and Cancer Risk Estimation. *Sci. Total Environ.* **2016**, *565*, 1071–1083. https://doi.org/10.1016/j.scitotenv.2016.05.137.

(29) Zhang, L.; Morisaki, H.; Wei, Y.; Li, Z.; Yang, L.; Zhou, Q.; Zhang, X.; Xing, W.; Hu, M.; Shima, M.; Toriba, A.; Hayakawa, K.; Tang, N. PM2.5-Bound Polycyclic Aromatic Hydrocarbons and Nitro-Polycyclic Aromatic Hydrocarbons inside and Outside a Primary

School Classroom in Beijing: Concentration, Composition, and Inhalation Cancer Risk. *Sci. Total Environ.* **2020**, *705*, 135840. https://doi.org/10.1016/j.scitotenv.2019.135840.

(30) Santos, R. R.; Cardeal, Z. de L.; Menezes, H. C. Phase Distribution of Polycyclic Aromatic Hydrocarbons and Their Oxygenated and Nitrated Derivatives in the Ambient Air of a Brazilian Urban Area☆. *Chemosphere* **2020**, *250*, 126223. https://doi.org/10.1016/j.chemosphere.2020.126223.

(31) Atkinson, R.; Arey, J. Atmospheric Chemistry of Gas-Phase Polycyclic Aromatic Hydrocarbons: Formation of Atmospheric Mutagens. *Environ. Health Perspect.* **1994**, *102 Suppl 4*, 117–126. https://doi.org/10.1289/ehp.94102s4117.

(32) Gbeddy, G.; Egodawatta, P.; Goonetilleke, A.; Akortia, E.; Glover, E. T. Influence of Photolysis on Source Characterization and Health Risk of Polycyclic Aromatic Hydrocarbons (PAHs), and Carbonyl-, Nitro-, Hydroxy- PAHs in Urban Road Dust. *Environ. Pollut.* **2021**, *269*, 116103. https://doi.org/10.1016/j.envpol.2020.116103.

(33) Zhang, J.; Yang, L.; Mellouki, A.; Chen, J.; Chen, X.; Gao, Y.; Jiang, P.; Li, Y.; Yu, H.; Wang, W. Atmospheric PAHs, NPAHs, and OPAHs at an Urban, Mountainous, and Marine Sites in Northern China: Molecular Composition, Sources, and Ageing. *Atmos. Environ.* **2018**, *173*, 256–264. https://doi.org/10.1016/j.atmosenv.2017.11.002.

(34) Dreij, K.; Lundin, L.; Le Bihanic, F.; Lundstedt, S. Polycyclic Aromatic Compounds in Urban Soils of Stockholm City: Occurrence, Sources and Human Health Risk Assessment. *Environ. Res.* **2020**, *182*, 108989. https://doi.org/10.1016/j.envres.2019.108989.

(35) Walgraeve, C.; Demeestere, K.; Dewulf, J.; Zimmermann, R.; Van Langenhove, H. Oxygenated Polycyclic Aromatic Hydrocarbons in Atmospheric Particulate Matter: Molecular Characterization and Occurrence. *Atmos. Environ.* **2010**, *44* (15), 1831–1846. https://doi.org/10.1016/j.atmosenv.2009.12.004.

(36) Idowu, O.; Semple, K. T.; Ramadass, K.; O'Connor, W.; Hansbro, P.; Thavamani, P. Beyond the Obvious: Environmental Health Implications of Polar Polycyclic Aromatic Hydrocarbons. *Environ. Int.* **2019**, *123*, 543–557. https://doi.org/10.1016/j.envint.2018.12.051.

(37) Clergé, A.; Le Goff, J.; Lopez, C.; Ledauphin, J.; Delépée, R. Oxy-PAHs: Occurrence in the Environment and Potential Genotoxic/Mutagenic Risk Assessment for Human Health. *Crit. Rev. Toxicol.* **2019**, *49* (4), 302–328. https://doi.org/10.1080/10408444.2019.1605333.

(38) Friedman, C. L.; Selin, N. E. Long-Range Atmospheric Transport of Polycyclic Aromatic Hydrocarbons: A Global 3-D Model Analysis Including Evaluation of Arctic Sources. *Environ. Sci. Technol.* **2012**, *46* (17), 9501–9510. https://doi.org/10.1021/es301904d.

(39) Halsall, C. J.; Sweetman, A. J.; Barrie, L. A.; Jones, K. C. Modelling the Behaviour of PAHs during Atmospheric Transport from the UK to the Arctic. *Atmos. Environ.* **2001**, *35* (2), 255–267. https://doi.org/10.1016/S1352-2310(00)00195-3.

(40) Arellano, L.; Fernández, P.; van Drooge, B. L.; Rose, N. L.; Nickus, U.; Thies, H.; Stuchlík, E.; Camarero, L.; Catalan, J.; Grimalt, J. O. Drivers of Atmospheric Deposition of Polycyclic Aromatic Hydrocarbons at European High-Altitude Sites. *Atmospheric Chem. Phys.* **2018**, *18* (21), 16081–16097. https://doi.org/10.5194/acp-18-16081-2018.

(41) Wang, Q.; Liu, M.; Li, Y.; Liu, Y.; Li, S.; Ge, R. Dry and Wet Deposition of Polycyclic Aromatic Hydrocarbons and Comparison with Typical Media in Urban System of Shanghai, China. *Atmos. Environ.* **2016**, *144*, 175–181. https://doi.org/10.1016/j.atmosenv.2016.08.079.

(42) Niu, L.; Gelder, P. H. A. J. M. van. Distribution and Source Assessment of Polycyclic Aromatic Hydrocarbons Levels from Lake IJssel (the Netherlands) and Their Responses to Hydrology. *Water Environ. Res.* **2020**, *92* (8), 1214–1229. https://doi.org/10.1002/wer.1317.

(43) Duttagupta, S.; Mukherjee, A.; Routh, J.; Devi, L. G.; Bhattacharya, A.; Bhattacharya, J. Role of Aquifer Media in Determining the Fate of Polycyclic Aromatic Hydrocarbons in the Natural Water and Sediments along the Lower Ganges River Basin. *J. Environ. Sci. Health Part A* **2020**, *55* (4), 354–373. https://doi.org/10.1080/10934529.2019.1696617.

(44) Andrade, M. V. F.; Santos, F. R.; Oliveira, A. H. B.; Nascimento, R. F.; Cavalcante, R. M. Influence of Sediment Parameters on the Distribution and Fate of PAHs in an Estuarine Tropical Region Located in the Brazilian Semi-Arid (Jaguaribe River, Ceará Coast). *Mar. Pollut. Bull.* **2019**, *146*, 703–710. https://doi.org/10.1016/j.marpolbul.2019.07.027.

(45) EPA, U. S. *Priority Pollutant List*; Vol. 40 CFR Part 423, Appendix A.

(46) Zoppini, A.; Ademollo, N.; Amalfitano, S.; Capri, S.; Casella, P.; Fazi, S.; Marxsen, J.; Patrolecco, L. Microbial Responses to Polycyclic Aromatic Hydrocarbon Contamination in Temporary River Sediments: Experimental Insights. *Sci. Total Environ.* **2016**, *541*, 1364–1371. https://doi.org/10.1016/j.scitotenv.2015.09.144.

(47) Wu, J.; Li, H.; Zhang, J.; Gu, Y.; Zhou, X.; Zhang, D.; Ma, Y.; Wang, S.; Nian, X.; Jin, W.; Li, R.; Xu, Z. Microbial Diversity and Function in Response to Occurrence and Source Apportionment of Polycyclic Aromatic Hydrocarbons in Combined Sewer Overflows. *J. Clean. Prod.* **2021**, *279*, 123723. https://doi.org/10.1016/j.jclepro.2020.123723.

(48) Maurya, A. P.; Rajkumari, J.; Pandey, P. Enrichment of Antibiotic Resistance Genes (ARGs) in Polyaromatic Hydrocarbon–Contaminated Soils: A Major Challenge for Environmental Health. *Environ. Sci. Pollut. Res.* **2021**. https://doi.org/10.1007/s11356-020-12171-3.

(49) Sverdrup, L. E.; Jensen, J.; Krogh, P. H.; Stenersen, J. Studies on the Effect of Soil Aging on the Toxicity of Pyrene and Phenanthrene to a Soil-Dwelling Springtail. *Environ. Toxicol. Chem.* **2002**, *21* (3), 489–492. https://doi.org/10.1002/etc.5620210303.

(50) Sverdrup, L. E.; Jensen, J.; Kelley, A. E.; Krogh, P. H.; Stenersen, J. Effects of Eight Polycyclic Aromatic Compounds on the Survival and Reproduction of Enchytraeus

Crypticus (Oligochaeta, Clitellata). *Environ. Toxicol. Chem.* **2002**, *21* (1), 109–114. https://doi.org/10.1002/etc.5620210116.

(51) Carls, M. G.; Holland, L.; Larsen, M.; Collier, T. K.; Scholz, N. L.; Incardona, J. P. Fish Embryos Are Damaged by Dissolved PAHs, Not Oil Particles. *Aquat. Toxicol.* **2008**, *88* (2), 121–127. https://doi.org/10.1016/j.aquatox.2008.03.014.

(52) Logan, D. T. Perspective on Ecotoxicology of PAHs to Fish. *Hum. Ecol. Risk Assess. Int. J.* **2007**, *13* (2), 302–316. https://doi.org/10.1080/10807030701226749.

(53) Shen, H.; Grist, S.; Nugegoda, D. The PAH Body Burdens and Biomarkers of Wild Mussels in Port Phillip Bay, Australia and Their Food Safety Implications. *Environ. Res.* **2020**, *188*, 109827. https://doi.org/10.1016/j.envres.2020.109827.

(54) Speciale, A.; Zena, R.; Calabrò, C.; Bertuccio, C.; Aragona, M.; Saija, A.; Trombetta, D.; Cimino, F.; Lo Cascio, P. Experimental Exposure of Blue Mussels (Mytilus Galloprovincialis) to High Levels of Benzo[a]Pyrene and Possible Implications for Human Health. *Ecotoxicol. Environ. Saf.* **2018**, *150*, 96–103. https://doi.org/10.1016/j.ecoenv.2017.12.038.

(55) Gardner, G. R.; Yevich, P. P.; Harshbarger, J. C.; Malcolm, A. R. Carcinogenicity of Black Rock Harbor Sediment to the Eastern Oyster and Trophic Transfer of Black Rock Harbor Carcinogens from the Blue Mussel to the Winter Flounder. *Environ. Health Perspect.* **1991**, *90*, 53–66. https://doi.org/10.2307/3430845.

(56) Meador, J. P.; Stein, J. E.; Reichert, W. L.; Varanasi, U. Bioaccumulation of Polycyclic Aromatic Hydrocarbons by Marine Organisms. **1995**, *143*, 79–165. https://doi.org/10.1007/978-1-4612-2542-3_4.

(57) Shinya, M.; Tsuchinaga, T.; Kitano, M.; Yamada, Y.; Ishikawa, M. Characterization of Heavy Metals and Polycyclic Aromatic Hydrocarbons in Urban Highway Runoff. *Water Sci. Technol.* **2000**, *42* (7–8), 201–208. https://doi.org/10.2166/wst.2000.0570.

(58) Eom, I. C.; Rast, C.; Veber, A. M.; Vasseur, P. Ecotoxicity of a Polycyclic Aromatic Hydrocarbon (PAH)-Contaminated Soil. *Ecotoxicol. Environ. Saf.* **2007**, *67* (2), 190–205. https://doi.org/10.1016/j.ecoenv.2006.12.020.

(59) Cerniglia, C. E. Microbial Metabolism of Polycyclic Aromatic Hydrocarbons. *Adv. Appl. Microbiol.* **1984**, *30*, 31–71. https://doi.org/10.1016/S0065-2164(08)70052-2.

(60) Ghosal, D.; Ghosh, S.; Dutta, T. K.; Ahn, Y. Current State of Knowledge in Microbial Degradation of Polycyclic Aromatic Hydrocarbons (PAHs): A Review. *Front Microbiol* **2016**, *7*, 1369. https://doi.org/10.3389/fmicb.2016.01369.

(61) Qin, Z.; Zhao, Z.; Jiao, W.; Han, Z.; Xia, L.; Fang, Y.; Wang, S.; Ji, L.; Jiang, Y. Phenanthrene Removal and Response of Bacterial Community in the Combined System of Photocatalysis and PAH-Degrading Microbial Consortium in Laboratory System. *Bioresour. Technol.* **2020**, *301*, 122736. https://doi.org/10.1016/j.biortech.2020.122736.

(62) Haritash, A. K.; Kaushik, C. P. Biodegradation Aspects of Polycyclic Aromatic Hydrocarbons (PAHs): A Review. *J Hazard Mater* **2009**, *169* (1–3), 1–15. https://doi.org/10.1016/j.jhazmat.2009.03.137.

(63) Genthner, B. R. S.; Townsend, G. T.; Lantz, S. E.; Mueller, J. G. Persistence of Polycyclic Aromatic Hydrocarbon Components of Creosote Under Anaerobic Enrichment Conditions. *Arch. Environ. Contam. Toxicol.* **1997**, *32* (1), 99–105. https://doi.org/10.1007/s002449900160.

(64) Analogues, N. R. C. (US) C. on P. and S. Polycyclic Aromatic Hydrocarbons: Evaluation of Sources and Effects; National Academies Press (US): Washington (DC), 1983.

(65) Huang, X.; Hu, B.; Wang, P.; Chen, X.; Xu, B. Microbial Diversity in Lake–River Ecotone of Poyang Lake, China. *Environ. Earth Sci.* **2016**, *75* (11), 965. https://doi.org/10.1007/s12665-016-5473-0.

(66) Miller, R. M.; Singer, G. M.; Rosen, J. D.; Bartha, R. Photolysis Primes Biodegradation of Benzo[a]Pyrene. *Appl. Environ. Microbiol.* **1988**, *54* (7), 1724–1730. https://doi.org/10.1128/aem.54.7.1724-1730.1988.

(67) Peng, R. H.; Xiong, A. S.; Xue, Y.; Fu, X. Y.; Gao, F.; Zhao, W.; Tian, Y. S.; Yao, Q. H. Microbial Biodegradation of Polyaromatic Hydrocarbons. *FEMS Microbiol Rev* **2008**, *32* (6), 927–955. https://doi.org/10.1111/j.1574-6976.2008.00127.x.

(68) Sleight, T.; Khanna, V.; Gilbertson, L.; Ng, C. Network Analysis for Prioritizing Biodegradation Metabolites of Polycyclic Aromatic Hydrocarbons. *Environ. Sci. Technol.* **2020**. https://doi.org/10.1021/acs.est.0c02217.

(69) Lawal, A. T.; Fantke, P. Polycyclic Aromatic Hydrocarbons. A Review. *Cogent Environ. Sci.* **2017**, *3* (1). https://doi.org/10.1080/23311843.2017.1339841.

(70) Chibwe, L.; Titaley, I. A.; Hoh, E.; Simonich, S. L. M. Integrated Framework for Identifying Toxic Transformation Products in Complex Environmental Mixtures. *Environ. Sci. Technol. Lett.* **2017**, *4* (2), 32–43. https://doi.org/10.1021/acs.estlett.6b00455.

(71) Titaley, I. A.; Simonich, S. L. M.; Larsson, M. Recent Advances in the Study of the Remediation of Polycyclic Aromatic Compound (PAC)-Contaminated Soils: Transformation Products, Toxicity, and Bioavailability Analyses. *Environ. Sci. Technol. Lett.* **2020**, *7* (12), 873–882. https://doi.org/10.1021/acs.estlett.0c00677.

(72) Schemeth, D.; Nielsen, N. J.; Andersson, J. T.; Christensen, J. H. A Tiered Analytical Approach for Target, Non-Target and Suspect Screening Analysis of Polar Transformation Products of Polycyclic Aromatic Compounds. *Chemosphere* **2019**, *235*, 175–184. https://doi.org/10.1016/j.chemosphere.2019.06.149.

(73) Carls, M. G.; Meador, J. P. A Perspective on the Toxicity of Petrogenic PAHs to Developing Fish Embryos Related to Environmental Chemistry. *Hum. Ecol. Risk Assess. Int. J.* **2009**, *15* (6), 1084–1098. https://doi.org/10.1080/10807030903304708.

(74) Toro, D. M. D.; McGrath, J. A. Technical Basis for Narcotic Chemicals and Polycyclic Aromatic Hydrocarbon Criteria. II. Mixtures and Sediments. *Environ. Toxicol. Chem.* **2000**, *19* (8), 1971–1982. https://doi.org/10.1002/etc.5620190804.

(75) EPA, U. S. *Ecological Structure - Activity Relationships Program (ECOSAR) Operation Manual v2.0*; 2017.

(76) Escher, B. I.; Baumer, A.; Bittermann, K.; Henneberger, L.; König, M.; Kühnert, C.; Klüver, N. General Baseline Toxicity QSAR for Nonpolar, Polar and Ionisable Chemicals and Their Mixtures in the Bioluminescence Inhibition Assay with Aliivibrio Fischeri. *Environ. Sci. Process. Impacts* **2017**, *19* (3), 414–428. https://doi.org/10.1039/c6em00692b.

(77) Dearden, J. C. Prediction of Environmental Toxicity and Fate Using Quantitative Structure-Activity Relationships (QSARs). *J. Braz. Chem. Soc.* **2002**, *13* (6), 754–762. https://doi.org/10.1590/S0103-50532002000600005.

(78) Wood, A. W.; Chang, R. L.; Levin, W.; Ryan, D. E.; Thomas, P. E.; Crolsy-Delcey, M. Mutagenicity of the Dihydrodiols and Bay-Region Diol-Epoxides of Benzo(c)Phenanthrene in Bacterial and Mammalian Cells. 8.

(79) Newcomb, K. O.; Sangaiah, R.; Gold, A.; Ball, L. M. Activation and Metabolism of Benz[j]Aceanthrylene-9,10-Dihydrodiol, the Precursor to Bay-Region Metabolism of the Genotoxic Cyclopenta-PAH Benz[j]Aceanthrylene. *Mutat. Res. Mol. Mech. Mutagen.* **1993**, *287* (2), 181–190. https://doi.org/10.1016/0027-5107(93)90011-4.

(80) Jerina, D. M.; Sayer, J. M.; Agarwal, S. K.; Yagi, H.; Levin, W.; Wood, A. W.; Conney, A. H.; Pruess-Schwartz, D.; Baird, W. M.; Pigott, M. A. Reactivity and Tumorigenicity of Bay-Region Diol Epoxides Derived from Polycyclic Aromatic Hydrocarbons. *Adv. Exp. Med. Biol.* **1986**, *197*, 11–30. https://doi.org/10.1007/978-1-4684-5134-4_2.

(81) Moorthy, B.; Chu, C.; Carlin, D. J. Polycyclic Aromatic Hydrocarbons: From Metabolism to Lung Cancer. *Toxicol. Sci.* **2015**, *145* (1), 5–15. https://doi.org/10.1093/toxsci/kfv040.

(82) Gao, P.; da Silva, E.; Hou, L.; Denslow, N. D.; Xiang, P.; Ma, L. Q. Human Exposure to Polycyclic Aromatic Hydrocarbons: Metabolomics Perspective. *Environ. Int.* **2018**, *119*, 466–477. https://doi.org/10.1016/j.envint.2018.07.017.

(83) Lehr, R. E.; Jerina, D. M. Metabolic Activations of Polycyclic Hydrocarbons - Structure-Activity Relationships. *Arch. Toxicol.* **1977**, *39* (1–2), 1–6. https://doi.org/10.1007/BF00343269.

(84) Wood, A. W.; Chang, R. L.; Levin, W.; Ryan, D. E.; Thomas, P. E.; Mah, H. D.; Karle, J. M.; Yagi, H.; Jerina, D. M.; Conney, A. H. Mutagenicity and Tumorigenicity of Phenanthrene and Chrysene Epoxides and Diol Epoxides. **1979**, *39*, 9.

(85) Arey, J.; Harger, W. P.; Helmig, D.; Atkinson, R. Bioassay-Directed Fractionation of Mutagenic PAH Atmospheric Photooxidation Products and Ambient Particulate Extracts. *Mutat. Res. Lett.* **1992**, *281* (1), 67–76. https://doi.org/10.1016/0165-7992(92)90038-J.

(86) Fernandez, P.; Grifoll, M.; Solanas, A. M.; Bayona, J. M.; Albaiges, J. Bioassay-Directed Chemical Analysis of Genotoxic Components in Coastal Sediments. *Environ. Sci. Technol.* **1992**, *26* (4), 817–829. https://doi.org/10.1021/es00028a024.

(87) Topinka, J.; Schwarz, L. R.; Kiefer, F.; Wiebel, F. J.; Gajdoš, O.; Vidová, P.; Dobiáš, L.; Fried, M.; Šrám, R. J.; Wolff, T. DNA Adduct Formation in Mammalian Cell Cultures by Polycyclic Aromatic Hydrocarbons (PAH) and Nitro-PAH in Coke Oven Emission Extract. *Mutat. Res. Toxicol. Environ. Mutagen.* **1998**, *419* (1), 91–105. https://doi.org/10.1016/S1383-5718(98)00127-2.

(88) Topinka, J.; Schwarz, L. R.; Wiebel, F. J.; Černá, M.; Wolff, T. Genotoxicity of Urban Air Pollutants in the Czech Republic: Part II. DNA Adduct Formation in Mammalian Cells by Extractable Organic Matter. *Mutat. Res. Toxicol. Environ. Mutagen.* **2000**, *469* (1), 83–93. https://doi.org/10.1016/S1383-5718(00)00061-9.

(89) Schrlau, J. E.; Kramer, A. L.; Chlebowski, A.; Truong, L.; Tanguay, R. L.; Simonich, S. L. M.; Semprini, L. Formation of Developmentally Toxic Phenanthrene Metabolite Mixtures by Mycobacterium Sp. ELW1. *Env. Sci Technol* **2017**, *51* (15), 8569–8578. https://doi.org/10.1021/acs.est.7b01377.

(90) Umbuzeiro, G. A.; Franco, A.; Martins, M. H.; Kummrow, F.; Carvalho, L.; Schmeiser, H. H.; Leykauf, J.; Stiborova, M.; Claxton, L. D. Mutagenicity and DNA Adduct Formation of PAH, Nitro-PAH, and Oxy-PAH Fractions of Atmospheric Particulate Matter from São Paulo, Brazil. *Mutat. Res. - Genet. Toxicol. Environ. Mutagen.* **2008**, *652* (1), 72–80. https://doi.org/10.1016/j.mrgentox.2007.12.007.

(91) Mattsson, Å.; Lundstedt, S.; Stenius, U. Exposure of HepG2 Cells to Low Levels of PAH-Containing Extracts from Contaminated Soils Results in Unpredictable Genotoxic Stress Responses. *Environ. Mol. Mutagen.* **2009**, *50* (4), 337–348. https://doi.org/10.1002/em.20486.

(92) Wang, W. T.; Jariyasopit, N.; Schrlau, J.; Jia, Y. L.; Tao, S.; Yu, T. W.; Dashwood, R. H.; Zhang, W.; Wang, X. J.; Simonich, S. L. M. Concentration and Photochemistry of PAHs, NPAHs, and OPAHs and Toxicity of PM2.5 during the Beijing Olympic Games. *Environ. Sci. Technol.* **2011**, *45* (16), 6887–6895. https://doi.org/10.1021/es201443z.

(93) Meyer, W.; Seiler, T.-B.; Schwarzbauer, J.; Püttmann, W.; Hollert, H.; Achten, C. Polar Polycyclic Aromatic Compounds from Different Coal Types Show Varying Mutagenic Potential, EROD Induction and Bioavailability Depending on Coal Rank. *Sci. Total Environ.* **2014**, *494–495*, 320–328. https://doi.org/10.1016/j.scitotenv.2014.06.140.

(94) Chibwe, L.; Geier, M. C.; Nakamura, J.; Tanguay, R. L.; Aitken, M. D.; Simonich, S. L. Aerobic Bioremediation of PAH Contaminated Soil Results in Increased Genotoxicity and

Developmental Toxicity. *Env. Sci Technol* **2015**, *49* (23), 13889–13898. https://doi.org/10.1021/acs.est.5b00499.

(95) Wincent, E.; Jönsson, M. E.; Bottai, M.; Lundstedt, S.; Dreij, K. Aryl Hydrocarbon Receptor Activation and Developmental Toxicity in Zebrafish in Response to Soil Extracts Containing Unsubstituted and Oxygenated PAHs. *Environ. Sci. Technol.* **2015**, *49* (6), 3869–3877. https://doi.org/10.1021/es505588s.

(96) Di Giorgio, C.; Malleret, L.; Gueydon-Morin, C.; Rigaud, S.; De Méo, M. Comparison of Two Extraction Procedures for the Assessment of Sediment Genotoxicity: Implication of Polar Organic Compounds. *Mutat. Res. Toxicol. Environ. Mutagen.* **2011**, *725* (1), 1–12. https://doi.org/10.1016/j.mrgentox.2011.05.012.

(97) Park, J.; Ball, L. M.; Richardson, S. D.; Zhu, H. B.; Aitken, M. D. Oxidative Mutagenicity of Polar Fractions from Polycyclic Aromatic Hydrocarbon-Contaminated Soils. *Environ. Toxicol. Chem.* **2008**, *27* (11), 2207–2215. https://doi.org/10.1897/07-572.1.

(98) Dasgupta, S.; Cao, A.; Mauer, B.; Yan, B.; Uno, S.; McElroy, A. Genotoxicity of Oxy-PAHs to Japanese Medaka (Oryzias Latipes) Embryos Assessed Using the Comet Assay. *Environ. Sci. Pollut. Res.* **2014**, *21* (24), 13867–13876. https://doi.org/10.1007/s11356-014-2586-4.

(99) Saeed, M.; Higginbotham, S.; Rogan, E.; Cavalieri, E. Formation of Depurinating N3adenine and N7guanine Adducts after Reaction of 1,2-Naphthoquinone or Enzyme-Activated 1,2-Dihydroxynaphthalene with DNA: Implications for the Mechanism of Tumor Initiation by Naphthalene. *Chem. Biol. Interact.* **2007**, *165* (3), 175–188. https://doi.org/10.1016/j.cbi.2006.12.007.

(100) McCoull, K. D.; Rindgen, D.; Blair, I. A.; Penning, T. M. Synthesis and Characterization of Polycyclic Aromatic Hydrocarbon O-Quinone Depurinating N7-Guanine Adducts. *Chem. Res. Toxicol.* **1999**, *12* (3), 237–246. https://doi.org/10.1021/tx980182z.

(101) Euler, L. Solutio Problematis Ad Geometriam Situs Pertinentis. *Comment. Acad. Sci. Imp. Petropolitanae* **1736**, *8*, 128–140.

(102) Oh, P.; Monge, P. Network Theory and Models. In *The International Encyclopedia of Communication Theory and Philosophy*; American Cancer Society, 2016; pp 1–15. https://doi.org/10.1002/9781118766804.wbiect246.

(103) Krause, J.; Croft, D. P.; James, R. Social Network Theory in the Behavioural Sciences: Potential Applications. *Behav. Ecol. Sociobiol.* **2007**, *62* (1), 15–27. https://doi.org/10.1007/s00265-007-0445-8.

(104) Vora, N.; Fath, B. D.; Khanna, V. A Systems Approach To Assess Trade Dependencies in U.S. Food–Energy–Water Nexus. *Environ. Sci. Technol.* **2019**, *53* (18), 10941–10950. https://doi.org/10.1021/acs.est.8b07288.

(105) Newman, M. *Networks: An Introduction*; Oxford University Press, 2010.

(106) Dunne, J. A.; Williams, R. J.; Martinez, N. D. Food-Web Structure and Network Theory: The Role of Connectance and Size. *Proc. Natl. Acad. Sci.* **2002**, *99* (20), 12917–12922. https://doi.org/10.1073/pnas.192407699.

(107) Cesare, C. Graph-Theory Breakthrough Tantalizes Mathematicians. *Nat. News*. https://doi.org/10.1038/nature.2015.18801.

(108) Aittokallio, T.; Schwikowski, B. Graph-Based Methods for Analysing Networks in Cell Biology. *Brief. Bioinform.* **2006**, *7* (3), 243–255. https://doi.org/10.1093/bib/bbl022.

(109) Cros A. Action de l'alcool Amylique Sur l'organisme.; 1863.

(110) Gramatica, P. A Short History of QSAR Evolution.

(111) Dearden, J. The History and Development of Quantitative Structure-Activity Relationships (QSARs). *Int. J. Quant. Struct.-Prop. Relatsh.* **2016**, *1*, 1–44. https://doi.org/10.4018/IJQSPR.2016010101.

(112) Perkins, James A. A History of Molecular Representation, Part One: 1800 to the 1960s http://www.jbiocommunication.org/issues/31-1/features3.html (accessed 2021 -02 -04).

(113) Pantsyrnaya, T.; Delaunay, S.; Goergen, J. L.; Guedon, E.; Paris, C.; Poupin, P.; Guseva, E.; Boudrant, J. Biodegradation of Phenanthrene by Pseudomonas Putida and a Bacterial Consortium in the Presence and in the Absence of a Surfactant. *Indian J. Microbiol.* **2012**, *52* (3), 420–426. https://doi.org/10.1007/s12088-012-0265-z.

(114) Balashova, N. V.; Kosheleva, I. A.; Filonov, A. E.; Gayazov, R. R.; Boronin, A. M. Phenanthrene-and Naphthalene-Degrading Strains of Pseudomonas Putida. *Microbiology* **1997**, *66* (4), 408–412.

(115) Mortelmans, K.; Zeiger, E. The Ames Salmonella/Microsome Mutagenicity Assay. *Mutat. Res. - Fundam. Mol. Mech. Mutagen.* **2000**, *455* (1–2), 29–60. https://doi.org/10.1016/S0027-5107(00)00064-6.

(116) EChA. *Guidance on Information Requirements and Chemical Safety Assessment, Part D: Framework for Exposure Assessment Https://Echa.Europa.Eu/Documents/10162/13632/Information_requirements_part_d_en. Pdf (Accessed 2019)*.

(117) U.S. EPA. Guidance for Reporting on the Environmental Fate and Transport of the Stressors of Concern in Problem Formulations.

(118) Wang, Z. D.; Fingas, M.; Shu, Y. Y.; Sigouin, L.; Landriault, M.; Lambert, P.; Turpin, R.; Campagna, P.; Mullin, J. Quantitative Characterization of PAHs In Burn Residue And Soot Samples And Differentiation Of Pyrogenic PAHs From Petrogenic PAHs - The 1994 Mobile Burn Study. *Environ. Sci. Technol.* **1999**, *33* (18), 3100–3109. https://doi.org/10.1021/es990031y.

(119) Abdel-Shafy, H. I.; Mansour, M. S. M. A Review on Polycyclic Aromatic Hydrocarbons: Source, Environmental Impact, Effect on Human Health and Remediation. *Egypt. J. Pet.* **2016**, *25* (1), 107–123. https://doi.org/10.1016/j.ejpe.2015.03.011.

(120) Yu, H. T. Environmental Carcinogenic Polycyclic Aromatic Hydrocarbons: Photochemistry and Phototoxicity. *J. Environ. Sci. Health Part C-Environ. Carcinog. Ecotoxicol. Rev.* **2002**, *20* (2), 149–183. https://doi.org/10.1081/gnc-120016203.

(121) Keyte, I. J.; Harrison, R. M.; Lammel, G. Chemical Reactivity and Long-Range Transport Potential of Polycyclic Aromatic Hydrocarbons--a Review. *Chem Soc Rev* **2013**, *42* (24), 9333–9391. https://doi.org/10.1039/c3cs60147a.

(122) Ligaray, M.; Baek, S. S.; Kwon, H. O.; Choi, S. D.; Cho, K. H. Watershed-Scale Modeling on the Fate and Transport of Polycyclic Aromatic Hydrocarbons (PAHs). *J Hazard Mater* **2016**, *320*, 442–457. https://doi.org/10.1016/j.jhazmat.2016.08.063.

(123) Guieysse, B.; Viklund, G.; Toes, A.-C.; Mattiasson, B. Combined UV-Biological Degradation of PAHs. *Chemosphere* **2004**, *55* (11), 1493–1499. https://doi.org/10.1016/j.chemosphere.2004.01.021.

(124) National Research Council (US) Committee on Pyrene and Selected Analogues. *Polycyclic Aromatic Hydrocarbons: Evaluation of Sources and Effects*; National Academies Press (US): Washington (DC), 1983.

(125) Srogi, K. Monitoring of Environmental Exposure to Polycyclic Aromatic Hydrocarbons: A Review. *Env. Chem Lett* **2007**, *5* (4), 169–195. https://doi.org/10.1007/s10311-007-0095-0.

(126) Wood, A. W.; Chang, R. L.; Levin, W.; Ryan, D. E.; Thomas, P. E.; Mah, H. D.; Karle, J. M.; Yagi, H.; Jerina, D. M.; Conney, A. H. Mutagenicity and Tumorigenicity of Phenanthrene and Chrysene Epoxides and Diol Epoxides. *Cancer Res.* **1979**, *39* (10), 4069–4077.

(127) Guo, J.; Sinclair, C. J.; Selby, K.; Boxall, A. B. A. Toxicological and Ecotoxicological Risk-Based Prioritization of Pharmaceuticals in the Natural Environment. *Environ. Toxicol. Chem.* **2016**, *35* (6), 1550–1559. https://doi.org/10.1002/etc.3319.

(128) EPA, U. S. *SAB Review of EPA's "Development of a Relative Potency Factor (RPF) Approach for Polycyclic Aromatic Hydrocarbon (PAH) Mixtures (February 2010 Draft)"*; 2011.

(129) Escher, B. I.; Eggen, R. I. L.; Schreiber, U.; Schreiber, Z.; Vye, E.; Wisner, B.; Schwarzenbach, R. P. Baseline Toxicity (Narcosis) of Organic Chemicals Determined by In Vitro Membrane Potential Measurements in Energy-Transducing Membranes. *Environ. Sci. Technol.* **2002**, *36* (9), 1971–1979. https://doi.org/10.1021/es015844c.

(130) Bruce, E. D.; Autenrieth, R. L.; Burghardt, R. C.; Donnelly, K. C.; McDonald, T. J. Using Quantitative Structure–Activity Relationships (QSAR) to Predict Toxic Endpoints for

Polycyclic Aromatic Hydrocarbons (PAH). *J. Toxicol. Environ. Health A* **2008**, *71* (16), 1073–1084. https://doi.org/10.1080/15287390802114337.

(131) Engraff, M.; Solere, C.; Smith, K. E.; Mayer, P.; Dahllof, I. Aquatic Toxicity of PAHs and PAH Mixtures at Saturation to Benthic Amphipods: Linking Toxic Effects to Chemical Activity. *Aquat Toxicol* **2011**, *102* (3–4), 142–149. https://doi.org/10.1016/j.aquatox.2011.01.009.

(132) Staal, Y. C.; Hebels, D. G.; van Herwijnen, M. H.; Gottschalk, R. W.; van Schooten, F. J.; van Delft, J. H. Binary PAH Mixtures Cause Additive or Antagonistic Effects on Gene Expression but Synergistic Effects on DNA Adduct Formation. *Carcinogenesis* **2007**, *28* (12), 2632–2640. https://doi.org/10.1093/carcin/bgm182.

(133) Hu, J.; Nakamura, J.; Richardson, S. D.; Aitken, M. D. Evaluating the Effects of Bioremediation on Genotoxicity of Polycyclic Aromatic Hydrocarbon-Contaminated Soil Using Genetically Engineered, Higher Eukaryotic Cell Lines. *Env. Sci Technol* **2012**, *46* (8), 4607–4613. https://doi.org/10.1021/es300020e.

(134) Traczewska, T. M. Changes of Toxicological Properties of Biodegradation Products of Anthracene and Phenanthrene. *Water Sci. Technol.* **2000**, *41* (12), 31–38.

(135) Roy, M.; Khara, P.; Dutta, T. K. Meta-Cleavage of Hydroxynaphthoic Acids in the Degradation of Phenanthrene by Sphingobium Sp. Strain PNB. *Microbiology* **2012**, *158* (Pt 3), 685–695. https://doi.org/10.1099/mic.0.053363-0.

(136) Luan, T. G.; Yu, K. S.; Zhong, Y.; Zhou, H. W.; Lan, C. Y.; Tam, N. F. Study of Metabolites from the Degradation of Polycyclic Aromatic Hydrocarbons (PAHs) by Bacterial Consortium Enriched from Mangrove Sediments. *Chemosphere* **2006**, *65* (11), 2289–2296. https://doi.org/10.1016/j.chemosphere.2006.05.013.

(137) Gustafson, K. E.; Dickhut, R. M. Gaseous Exchange of Polycyclic Aromatic Hydrocarbons across the Air−Water Interface of Southern Chesapeake Bay. *Environ. Sci. Technol.* **1997**, *31* (6), 1623–1629. https://doi.org/10.1021/es960377y.

(138) Wickliffe, J.; Overton, E.; Frickel, S.; Howard, J.; Wilson, M.; Simon, B.; Echsner, S.; Nguyen, D.; Gauthe, D.; Blake, D.; Miller, C.; Elferink, C.; Ansari, S.; Fernando, H.; Trapido, E.; Kane, A. Evaluation of Polycyclic Aromatic Hydrocarbons Using Analytical Methods, Toxicology, and Risk Assessment Research: Seafood Safety after a Petroleum Spill as an Example. *Env. Health Perspect* **2014**, *122* (1), 6–9. https://doi.org/10.1289/ehp.1306724.

(139) Cajthaml T; Möder M; Kacer P; Sasek V; P, P. Study of Fungal Degradation Products of Polycyclic Aromatic Hydrocarbons Using Gas Chromatography with Ion Trap Mass Spectrometry Detection. *J. Chromatogr.* **2002**, *974*, 213–222.

(140) Ye, J. S.; Yin, H.; Qiang, J.; Peng, H.; Qin, H. M.; Zhang, N.; He, B. Y. Biodegradation of Anthracene by Aspergillus Fumigatus. *J Hazard Mater* **2011**, *185* (1), 174–181. https://doi.org/10.1016/j.jhazmat.2010.09.015.

(141) Gao, S.; Seo, J. S.; Wang, J.; Keum, Y. S.; Li, J.; Li, Q. X. Multiple Degradation Pathways of Phenanthrene by Stenotrophomonas Maltophilia C6. *Int Biodeterior Biodegrad.* **2013**, *79*, 98–104. https://doi.org/10.1016/j.ibiod.2013.01.012.

(142) Hadibarata, T.; Zubir, M.; Rubiyatno; Chuang, T. Z. Microbial Transformation and Sorption of Anthracene in Liquid Culture. *Bioprocess Biosyst. Eng.* **2013**, *36* (9), 1229–1233. https://doi.org/10.1007/s00449-012-0850-x.

(143) Mallick, S.; Chatterjee, S.; Dutta, T. K. A Novel Degradation Pathway in the Assimilation of Phenanthrene by Staphylococcus Sp. Strain PN/Y via Meta-Cleavage of 2-Hydroxy-1-Naphthoic Acid: Formation of Trans-2,3-Dioxo-5-(2'-Hydroxyphenyl)-Pent-4-Enoic Acid. *Microbiology* **2007**, *153* (Pt 7), 2104–2115. https://doi.org/10.1099/mic.0.2006/004218-0.

(144) Gao, J. F.; Ellis, L. B. M.; Wackett, L. P. The University of Minnesota Biocatalysis/Biodegradation Database: Improving Public Access. *Nucleic Acids Res.* **2010**, *38*, D488–D491. https://doi.org/10.1093/nar/gkp771.

(145) Kirkley, A.; Barbosa, H.; Barthelemy, M.; Ghoshal, G. From the Betweenness Centrality in Street Networks to Structural Invariants in Random Planar Graphs. *Nat. Commun.* **2018**, *9*. https://doi.org/10.1038/s41467-018-04978-z.

(146) Finley, S. D.; Broadbelt, L. J.; Hatzimanikatis, V. Computational Framework for Predictive Biodegradation. *Biotechnol. Bioeng.* **2009**, *104* (6), 1086–1097. https://doi.org/10.1002/bit.22489.

(147) Ravikrishnan, A.; Nasre, M.; Raman, K. Enumerating All Possible Biosynthetic Pathways in Metabolic Networks. *Sci. Rep.* **2018**, *8* (1), 9932. https://doi.org/10.1038/s41598-018-28007-7.

(148) Shinar, G.; Feinberg, M. Structural Sources of Robustness in Biochemical Reaction Networks. *Science* **2010**, *327* (5971), 1389–1391. https://doi.org/10.1126/science.1183372.

(149) Shinar, G.; Alon, U.; Feinberg, M. Sensitivity and Robustness in Chemical Reaction Networks. *SIAM J. Appl. Math.* **2009**, *69* (4), 977–998.

(150) Blum, T.; Kohlbacher, O. Using Atom Mapping Rules for an Improved Detection of Relevant Routes in Weighted Metabolic Networks. *J. Comput. Biol. J. Comput. Mol. Cell Biol.* **2008**, *15* (6), 565–576. https://doi.org/10.1089/cmb.2008.0044.

(151) RDKit, O.-S. Cheminformatics. http://www.rdkit.org. .

(152) EPA, U. Estimation Programs Interface Suite™ for Microsoft® Windows, v 4.11. *U. S. Environ. Prot. Agency Wash. DC USA* **2019**.

(153) Cheng, T. J.; Zhao, Y.; Li, X.; Lin, F.; Xu, Y.; Zhang, X. L.; Li, Y.; Wang, R. X.; Lai, L. H. Computation of Octanol-Water Partition Coefficients by Guiding an Additive Model

with Knowledge. *J. Chem. Inf. Model.* **2007**, *47* (6), 2140–2148. https://doi.org/10.1021/ci700257y.

(154) Plante, J.; Werner, S. JPlogP: An Improved LogP Predictor Trained Using Predicted Data. *J. Cheminformatics* **2018**, *10*. https://doi.org/10.1186/s13321-018-0316-5.

(155) Eros, D.; Kovesdi, I.; Orfi, L.; Takacs-Novak, K.; Acsady, G.; Keri, G. Reliability of LogP Predictions Based on Calculated Molecular Descriptors: A Critical Review. *Curr. Med. Chem.* **2002**, *9* (20), 1819–1829. https://doi.org/10.2174/0929867023369042.

(156) Sakuratani, Y.; Kasai, K.; Noguchi, Y.; Yamada, J. Comparison of Predictivities of Log P Calculation Models Based on Experimental Data for 134 Simple Organic Compounds. *Qsar Comb. Sci.* **2007**, *26* (1), 109–116. https://doi.org/10.1002/qsar.200630019.

(157) McKinney, W. Data Structures for Statistical Computing in Python; 2010; pp 51–56.

(158) van der Walt, S.; Colbert, S. C.; Varoquaux, G. The NumPy Array: A Structure for Efficient Numerical Computation. *Comput. Sci. Eng.* **2011**, *13* (2), 22–30. https://doi.org/10.1109/MCSE.2011.37.

(159) Hunter, J. D. Matplotlib: A 2D Graphics Environment. *Comput. Sci. Eng.* **2007**, *9* (3). https://doi.org/10.1109/MCSE.2007.55.

(160) Morgan, H. L. The Generation of a Unique Machine Description for Chemical Structures-A Technique Developed at Chemical Abstracts Service. *J. Chem. Doc.* **1965**, *5* (2), 107–113. https://doi.org/10.1021/c160017a018.

(161) Tanimoto, T. T. *An Elementary Mathematical Theory of Classification and Prediction*; International Business Machines Corporation: New York, 1958.

(162) Butina, D. Unsupervised Data Base Clustering Based on Daylight's Fingerprint and Tanimoto Similarity: A Fast and Automated Way to Cluster Small and Large Data Sets. *J. Chem. Inf. Comput. Sci.* **1999**, *39* (4), 747–750. https://doi.org/10.1021/ci9803381.

(163) Hagberg, A. A.; Schul, D. A.; Swart, P. J. Exploring Network Structure, Dynamics, and Function Using NetworkX; G Varoquaux, T. V., J. Millman, Series Ed.; pp 11–15.

(164) Voutchkova, A. M.; Kostal, J.; Steinfeld, J. B.; Emerson, J. W.; Brooks, B. W.; Anastas, P.; Zimmerman, B. Towards Rational Molecular Design: Derivation of Property Guidelines for Reduced Acute Aquatic Toxicity. *Green Chem.* **2011**, *13* (9), 2373–2379. https://doi.org/10.1039/c1gc15651a.

(165) Ewa, B.; Danuta, M. S. Polycyclic Aromatic Hydrocarbons and PAH-Related DNA Adducts. *J Appl Genet* **2017**, *58* (3), 321–330. https://doi.org/10.1007/s13353-016-0380-3.

(166) Schultz, T. W.; Yarbrough, J. W. Trends in Structure-Toxicity Relationships for Carbonyl-Containing Alpha,Beta-Unsaturated Compounds. *SAR QSAR Environ. Res.* **2004**, *15* (2), 139–146. https://doi.org/10.1080/10629360410001665839.

(167) Landrum, G. The RDKit Documentation https://www.rdkit.org/docs/index.html.

(168) Tian, L.; Ma, P.; Zhong, J.-J. Impact of the Presence of Salicylate or Glucose on Enzyme Activity and Phenanthrene Degradation by Pseudomonas Mendocina. *Process Biochem.* **2003**, *38* (8), 1125–1132. https://doi.org/10.1016/s0032-9592(02)00245-5.

(169) Sponza, D. T.; Gok, O. Aerobic Biodegradation and Inhibition Kinetics of Poly-Aromatic Hydrocarbons (PAHs) in a Petrochemical Industry Wastewater in the Presence of Biosurfactants. *J. Chem. Technol. Biotechnol.* **2012**, *87* (5), 658–672. https://doi.org/10.1002/jctb.2762.

(170) Andersson, J. T.; Achten, C. Time to Say Goodbye to the 16 EPA PAHs? Toward an Up-to-Date Use of PACs for Environmental Purposes. *Polycycl Aromat Compd* **2015**, *35* (2–4), 330–354. https://doi.org/10.1080/10406638.2014.991042.

(171) Abdel-Shafy, H. I.; Mansour, M. S. M. A Review on Polycyclic Aromatic Hydrocarbons: Source, Environmental Impact, Effect on Human Health and Remediation. *Egypt. J. Pet.* **2016**, *25* (1), 107–123. https://doi.org/10.1016/j.ejpe.2015.03.011.

(172) Zhang, Y.; Tao, S. Global Atmospheric Emission Inventory of Polycyclic Aromatic Hydrocarbons (PAHs) for 2004. *Atmos. Environ.* **2009**, *43* (4), 812–819. https://doi.org/10.1016/j.atmosenv.2008.10.050.

(173) Liu, J.; Liu, Y. J.; Liu, Z.; Zhang, A.; Liu, Y. Source Apportionment of Soil PAHs and Human Health Exposure Risks Quantification from Sources: The Yulin National Energy and Chemical Industry Base, China as Case Study. *Environ. Geochem. Health* **2019**, *41* (2), 617–632. https://doi.org/10.1007/s10653-018-0155-3.

(174) Ghosal, D.; Ghosh, S.; Dutta, T. K.; Ahn, Y. Current State of Knowledge in Microbial Degradation of Polycyclic Aromatic Hydrocarbons (PAHs): A Review. *Front. Microbiol.* **2016**, *7*. https://doi.org/10.3389/fmicb.2016.01369.

(175) Sleight, T. W.; Khanna, V.; Gilbertson, L. M.; Ng, C. A. Network Analysis for Prioritizing Biodegradation Metabolites of Polycyclic Aromatic Hydrocarbons. *Environ. Sci. Technol.* **2020**, *54* (17), 10735–10744. https://doi.org/10.1021/acs.est.0c02217.

(176) EPA, U. S. *Priority Pollutant List*; Vol. 40 CFR Part 423, Appendix A.

(177) Hu, J.; Nakamura, J.; Richardson, S. D.; Aitken, M. D. Evaluating the Effects of Bioremediation on Genotoxicity of Polycyclic Aromatic Hydrocarbon-Contaminated Soil Using Genetically Engineered, Higher Eukaryotic Cell Lines. *Env. Sci Technol* **2012**, *46* (8), 4607–4613. https://doi.org/10.1021/es300020e.

(178) Lundstedt, S.; White, P. A.; Lemieux, C. L.; Lynes, K. D.; Lambert, I. B.; Öberg, L.; Haglund, P.; Tysklind, M. Sources, Fate, and Toxic Hazards of Oxygenated Polycyclic Aromatic Hydrocarbons (PAHs) at PAH-Contaminated Sites. *Ambio* **2007**, *36* (6), 475–485. https://doi.org/10.1579/0044-7447(2007)36[475:SFATHO]2.0.CO;2.

(179) Belkin, S.; Stieber, M.; Tiehm, A.; Frimmel, F. H.; Abeliovich, A.; Werner, P.; Ulitzur, S. Toxicity and Genotoxicity Enhancement during Polycyclic Aromatic Hydrocarbons' Biodegradation. *Environ. Toxicol. Water Qual.* **1994**, *9* (4), 303–309. https://doi.org/10.1002/tox.2530090409.

(180) Traczewska, T. M. Evaluation of Mutagenicity of Anthracene and Phenanthrene and Products of Their Microbiological Transformation by the Ames Test. **January 1**, *7*, 63–78.

(181) Haile, T. M.; Misik, M.; Grummt, T.; Halh, A. S.; Pichler, C.; Knasmueller, S.; Fuerhacker, M. Cytotoxic and Genotoxic Activities of Waters and Sediments from Highway and Parking Lot Runoffs. *Water Sci. Technol.* **2016**, *73* (11), 2772–2780. https://doi.org/10.2166/wst.2016.137.

(182) Hughes, T. J.; Claxton, L. D.; Brooks, L.; Warren, S.; Brenner, R.; Kremer, F. Genotoxicity of Bioremediated Soils from the Reilly Tar Site, St. Louis Park, Minnesota. *Environ. Health Perspect.* **1998**, *106*, 1427–1433. https://doi.org/10.1289/ehp.98106s61427.

(183) US EPA, O. EPI Suite™-Estimation Program Interface https://www.epa.gov/tsca-screening-tools/epi-suitetm-estimation-program-interface (accessed 2020 -06 -22).

(184) Haritash, A. K.; Kaushik, C. P. Biodegradation Aspects of Polycyclic Aromatic Hydrocarbons (PAHs): A Review. *J Hazard Mater* **2009**, *169* (1–3), 1–15. https://doi.org/10.1016/j.jhazmat.2009.03.137.

(185) McCarrick, S.; Cunha, V.; Zapletal, O.; Vondráček, J.; Dreij, K. In Vitro and in Vivo Genotoxicity of Oxygenated Polycyclic Aromatic Hydrocarbons. *Environ. Pollut.* **2019**, *246*, 678–687. https://doi.org/10.1016/j.envpol.2018.12.092.

(186) Umbuzeiro, G. A.; Franco, A.; Martins, M. H.; Kummrow, F.; Carvalho, L.; Schmeiser, H. H.; Leykauf, J.; Stiborova, M.; Claxton, L. D. Mutagenicity and DNA Adduct Formation of PAH, Nitro-PAH, and Oxy-PAH Fractions of Atmospheric Particulate Matter from São Paulo, Brazil. *Mutat. Res. - Genet. Toxicol. Environ. Mutagen.* **2008**, *652* (1), 72–80. https://doi.org/10.1016/j.mrgentox.2007.12.007.

(187) Neves, B. J.; Braga, R. C.; Melo-Filho, C. C.; Moreira-Filho, J. T.; Muratov, E. N.; Andrade, C. H. QSAR-Based Virtual Screening: Advances and Applications in Drug Discovery. *Front. Pharmacol.* **2018**, *9*. https://doi.org/10.3389/fphar.2018.01275.

(188) Piir Geven; Kahn Iiris; García-Sosa Alfonso T.; Sild Sulev; Ahte Priit; Maran Uko. Best Practices for QSAR Model Reporting: Physical and Chemical Properties, Ecotoxicity, Environmental Fate, Human Health, and Toxicokinetics Endpoints. *Environ. Health Perspect. 126* (12), 126001. https://doi.org/10.1289/EHP3264.

(189) Honma, M.; Kitazawa, A.; Cayley, A.; Williams, R. V.; Barber, C.; Hanser, T.; Saiakhov, R.; Chakravarti, S.; Myatt, G. J.; Cross, K. P.; Benfenati, E.; Raitano, G.; Mekenyan, O.; Petkov, P.; Bossa, C.; Benigni, R.; Battistelli, C. L.; Giuliani, A.; Tcheremenskaia, O.; DeMeo, C.; Norinder, U.; Koga, H.; Jose, C.; Jeliazkova, N.; Kochev, N.; Paskaleva, V.; Yang, C.; Daga, P. R.; Clark, R. D.; Rathman, J. Improvement of Quantitative Structure-Activity Relationship (QSAR) Tools for Predicting Ames Mutagenicity: Outcomes of the Ames/QSAR International Challenge Project. *Mutagenesis* **2019**, *34* (1), 41–48. https://doi.org/10.1093/mutage/gey031.

(190) EPA, U. S. *Toxicity Estimation Software Tool (TEST)*; 2016.

(191) Benfenati E, Manganaro A, Gini G.,. VEGA-QSAR: AI inside a Platform for Predictive Toxicology. Proceedings of the Workshop "Popularize Artificial Intelligence 2013"; CEUR Workshop Proceedings: Turin, Italy, 2013; Vol. 1107.

(192) Ames, B. N.; Lee, F. D.; Durston, W. E. An Improved Bacterial Test System for the Detection and Classification of Mutagens and Carcinogens. *Proc. Natl. Acad. Sci. U. S. A.* **1973**, *70* (3), 782–786. https://doi.org/10.1073/pnas.70.3.782.

(193) Lemieux, C. L.; Lambert, I. B.; Lundstedt, S.; Tysklind, M.; White, P. A. Mutagenic Hazards of Complex Polycyclic Aromatic Hydrocarbon Mixtures in Contaminated Soil. *Environ. Toxicol. Chem.* **2008**, *27* (4), 978–990. https://doi.org/10.1897/07-157.1.

(194) Wang, Y.; Chen, J.; Li, F.; Qin, H.; Qiao, X.; Hao, C. Modeling Photoinduced Toxicity of PAHs Based on DFT-Calculated Descriptors. *Chemosphere* **2009**, *76* (7), 999–1005. https://doi.org/10.1016/j.chemosphere.2009.04.010.

(195) Al-Fahemi, J. H. The Use of Quantum-Chemical Descriptors for Predicting the Photoinduced Toxicity of PAHs. *J. Mol. Model.* **2012**, *18* (9), 4121–4129. https://doi.org/10.1007/s00894-012-1417-0.

(196) Debnath, A. K.; Lopez Compadre, R. L.; Shusterman, A. J.; Hansch, C. Quantitative Structure-activity Relationship Investigation of the Role of Hydrophobicity in Regulating Mutagenicity in the Ames Test: 2. Mutagenicity of Aromatic and Heteroaromatic Nitro Compounds in Salmonella Typhimurium TA100. *Environ. Mol. Mutagen.* **1992**, *19* (1), 53–70. https://doi.org/10.1002/em.2850190108.

(197) Debnath, A. K.; Debnath, G.; Hansch, C.; de Compadre, R. L. L.; Shusterman, A. J. Structure-Activity Relationship of Mutagenic Aromatic and Heteroaromatic Nitro Compounds. Correlation with Molecular Orbital Energies and Hydrophobicity. *J. Med. Chem.* **1991**, *34* (2), 786–797. https://doi.org/10.1021/jm00106a046.

(198) Cronin, M. T. D.; Walker, J. D.; Jaworska, J. S.; Comber, M. H. I.; Watts, C. D.; Worth, A. P. Use of QSARs in International Decision-Making Frameworks to Predict Ecologic Effects and Environmental Fate of Chemical Substances. *Environ. Health Perspect.* **2003**, *111* (10), 1376–1390. https://doi.org/10.1289/ehp.5759.

(199) Wang, L. R.; Wang, Y.; Chen, J. W.; Guo, L. H. A Structure-Based Investigation on the Binding Interaction of Hydroxylated Polycyclic Aromatic Hydrocarbons with DNA. *Toxicology* **2009**, *262* (3), 250–257. https://doi.org/10.1016/j.tox.2009.06.015.

(200) Pérez-Garrido, A.; Helguera, A. M.; López, G. C.; Cordeiro, M. N. D. S.; Escudero, A. G. A Topological Substructural Molecular Design Approach for Predicting Mutagenesis End-Points of α, β-Unsaturated Carbonyl Compounds. *Toxicology* **2010**, *268* (1–2), 64–77. https://doi.org/10.1016/j.tox.2009.11.023.

(201) Papa, E.; Pilutti, P.; Gramatica, P. Prediction of PAH Mutagenicity in Human Cells by QSAR Classification. *SAR QSAR Environ. Res.* **2008**, *19* (1–2), 115–127. https://doi.org/10.1080/10629360701843482.

(202) Gramatica, P.; Papa, E.; Marrocchi, A.; Minuti, L.; Taticchi, A. Quantitative Structure–Activity Relationship Modeling of Polycyclic Aromatic Hydrocarbon Mutagenicity by Classification Methods Based on Holistic Theoretical Molecular Descriptors. *Ecotoxicol. Environ. Saf.* **2007**, *66* (3), 353–361. https://doi.org/10.1016/j.ecoenv.2006.02.005.

(203) International Organization for Standardization. ISO 11350 Water quality — Determination of the genotoxicity of water and waste water — Salmonella/microsome fluctuation test (Ames fluctuation test) https://www.iso.org/cms/render/live/en/sites/isoorg/contents/data/standard/05/03/50393.html (accessed 2021 -01 -29).

(204) Chemical Carcinogenesis Research Information System https://pubchem.ncbi.nlm.nih.gov/source/22070.

(205) Property Search https://www.echemportal.org/echemportal/property-search (accessed 2021 -04 -30).

(206) *EURL ECVAM Genotoxicity and Carcinogenicity Consolidated Database of Ames Positive Chemicals. European Commission, Joint Research Centre (JRC)*.

(207) *Carcinogenic Potency Database (CPDB) Data*.

(208) Hansen, K.; Mika, S.; Schroeter, T.; Sutter, A.; ter Laak, A.; Steger-Hartmann, T.; Heinrich, N.; Müller, K.-R. Benchmark Data Set for in Silico Prediction of Ames Mutagenicity. *J. Chem. Inf. Model.* **2009**, *49* (9), 2077–2081. https://doi.org/10.1021/ci900161g.

(209) Hao, Y.; Sun, G.; Fan, T.; Sun, X.; Liu, Y.; Zhang, N.; Zhao, L.; Zhong, R.; Peng, Y. Prediction on the Mutagenicity of Nitroaromatic Compounds Using Quantum Chemistry Descriptors Based QSAR and Machine Learning Derived Classification Methods. *Ecotoxicol. Environ. Saf.* **2019**, *186*, 109822. https://doi.org/10.1016/j.ecoenv.2019.109822.

(210) Reenu, Vikas. Electron-Correlation Based Externally Predictive QSARs for Mutagenicity of Nitrated-PAHs in Salmonella Typhimurium TA100. *Ecotoxicol. Environ. Saf.* **2014**, *101* (1), 42–50. https://doi.org/10.1016/j.ecoenv.2013.11.020.

(211) Kim, D.; Hong, S. I.; Lee, D. S. The Quantitative Structure-Mutagenicity Relationship of Polycyclic Aromatic Hydrocarbon Metabolites. *Int. J. Mol. Sci.* **2006**, *7* (12), 556–570. https://doi.org/10.3390/i7120556.

(212) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; Li, X.; Caricato, M.; Marenich, A.; Bloino, J.; Janesko, B. G.; Gomperts, R.; Mennucci, B.; Hratchian, H. P.; Ortiz, J. V.; Izmaylov, A. F.; Sonnenberg, J. L.; Williams-Young, D.; Ding, F.; Lipparini, F.; Egidi, F.; Goings, J.; Peng, B.; Petrone, A.; Henderson, T.; Ranasinghe, D.; Zakrzewski, V. G.; Gao, J.; Rega, N.; Zheng, G.; Liang, W.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Throssell, K.; Montgomery, Jr., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Keith, T.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Millam, J. M.; Klene, M.; Adamo, C.; Cammi, R.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Farkas, O.; Foresman, J. B.; Fox, D. J. *Gaussian 09*; Gaussian, Inc.: Wallingford CT, 2013.

(213) Zhao, Y.; Truhlar, D. G. The M06 Suite of Density Functionals for Main Group Thermochemistry, Thermochemical Kinetics, Noncovalent Interactions, Excited States, and Transition Elements: Two New Functionals and Systematic Testing of Four M06-Class Functionals and 12 Other Functionals. *Theor. Chem. Acc.* **2008**, *120* (1), 215–241. https://doi.org/10.1007/s00214-007-0310-x.

(214) Krishnan, R.; Binkley, J. S.; Seeger, R.; Pople, J. A. Self-consistent Molecular Orbital Methods. XX. A Basis Set for Correlated Wave Functions. *J. Chem. Phys.* **1980**, *72* (1), 650–654. https://doi.org/10.1063/1.438955.

(215) Ong, B. K.; Woon, K. L.; Ariffin, A. Evaluation of Various Density Functionals for Predicting the Electrophosphorescent Host HOMO, LUMO and Triplet Energies. *Synth. Met.* **2014**, *195*, 54–60. https://doi.org/10.1016/j.synthmet.2014.05.015.

(216) Ramesh, P.; Lydia Caroline, M.; Muthu, S.; Narayana, B.; Raja, M.; Aayisha, S. Spectroscopic and DFT Studies, Structural Determination, Chemical Properties and Molecular Docking of 1-(3-Bromo-2-Thienyl)-3-[4-(Dimethylamino)-Phenyl]Prop-2-En-1-One. *J. Mol. Struct.* **2020**, *1200*, 127123. https://doi.org/10.1016/j.molstruc.2019.127123.

(217) Thamarai, A.; Vadamalar, R.; Raja, M.; Muthu, S.; Narayana, B.; Ramesh, P.; Muhamed, R. R.; Sevvanthi, S.; Aayisha, S. Molecular Structure Interpretation, Spectroscopic (FT-IR, FT-Raman), Electronic Solvation (UV–Vis, HOMO-LUMO and NLO) Properties and Biological Evaluation of (2E)-3-(Biphenyl-4-Yl)-1-(4-Bromophenyl)Prop-2-En-1-One: Experimental and Computational Modeling Approach. *Spectrochim. Acta. A. Mol. Biomol. Spectrosc.* **2020**, *226*, 117609. https://doi.org/10.1016/j.saa.2019.117609.

(218) Yap, C. W. PaDEL-Descriptor: An Open Source Software to Calculate Molecular Descriptors and Fingerprints. *J. Comput. Chem.* **2011**, *32* (7), 1466–1474. https://doi.org/10.1002/jcc.21707.

(219) Bolton, J. L.; Trush, M. A.; Penning, T. M.; Dryhurst, G.; Monks, T. J. Role of Quinones in Toxicology. *Chem. Res. Toxicol.* **2000**, *13* (3), 135–160. https://doi.org/10.1021/tx9902082.

(220) Shen, Y.-M.; Troxel, A. B.; Vedantam, S.; Penning, T. M.; Field, J. Comparison of P53 Mutations Induced by PAH O-Quinones with Those Caused by Anti-Benzo[a]Pyrene Diol Epoxide in Vitro: Role of Reactive Oxygen and Biological Selection. *Chem. Res. Toxicol.* **2006**, *19* (11), 1441–1450. https://doi.org/10.1021/tx0601206.

(221) Park, J.; Ball, L. M.; Richardson, S. D.; Zhu, H.-B.; Aitken, M. D. Oxidative Mutagenicity of Polar Fractions from Polycyclic Aromatic Hydrocarbon–Contaminated Soils. *Environ. Toxicol. Chem.* **2008**, *27* (11), 2207–2215. https://doi.org/10.1897/07-572.1.

(222) Penning, T. M.; Ohnishi, S. T.; Ohnishi, T.; Harvey, R. G. Generation of Reactive Oxygen Species during the Enzymatic Oxidation of Polycyclic Aromatic Hydrocarbon Trans-Dihydrodiols Catalyzed by Dihydrodiol Dehydrogenase. *Chem. Res. Toxicol.* **1996**, *9* (1), 84–92. https://doi.org/10.1021/tx950055s.

(223) Szeliga, J.; Amin, S.; Zhang, F.-J.; Harvey, R. G. Reactions of Dihydrodiol Epoxides of 5-Methylchrysene and 5,6-Dimethylchrysene with DNA and Deoxyribonucleotides. *Chem. Res. Toxicol.* **1999**, *12* (4), 347–352. https://doi.org/10.1021/tx980228o.

(224) Newcomb, K. O.; Sangaiah, R.; Gold, A.; Ball, L. M. Activation and Metabolism of Benz[j]Aceanthrylene-9,10-Dihydrodiol, the Precursor to Bay-Region Metabolism of the Genotoxic Cyclopenta-PAH Benz[j]Aceanthrylene. *Mutat. Res. Mol. Mech. Mutagen.* **1993**, *287* (2), 181–190. https://doi.org/10.1016/0027-5107(93)90011-4.

(225) Lever, J.; Krzywinski, M.; Altman, N. Classification Evaluation. *Nat. Methods* **2016**, *13* (8), 603–604. https://doi.org/10.1038/nmeth.3945.

(226) Rajabi, M.; Shafiei, F. Structure–Property Relationships of Aliphatic Esters Using Topological Descriptors and Backward-Multiple Linear Regression Method. *J. Chin. Chem. Soc.* **2020**, *67* (8), 1338–1347. https://doi.org/10.1002/jccs.201900528.

(227) Marcoulides, K. M.; Raykov, T. Evaluation of Variance Inflation Factors in Regression Models Using Latent Variable Modeling Methods. *Educ. Psychol. Meas.* **2019**, *79* (5), 874–882. https://doi.org/10.1177/0013164418817803.

(228) Consonni, V.; Todeschini, R. Molecular Descriptors. In *Recent Advances in QSAR Studies: Methods and Applications*; Puzyn, T., Leszczynski, J., Cronin, M. T., Eds.; Challenges and Advances in Computational Chemistry and Physics; Springer Netherlands: Dordrecht, 2010; pp 29–102. https://doi.org/10.1007/978-1-4020-9783-6_3.

(229) OECD. Test No. 471: Bacterial Reverse Mutation Test, 2020.

(230) Dearden, J. C.; Cronin, M. T. D.; Kaiser, K. L. E. How Not to Develop a Quantitative Structure–Activity or Structure–Property Relationship (QSAR/QSPR). *SAR QSAR Environ. Res.* **2009**, *20* (3–4), 241–266. https://doi.org/10.1080/10629360902949567.

(231) Matsui, T.; Yamada, N.; Kuno, H.; Kanaly, R. A. Formation of Bulky DNA Adducts by Non-Enzymatic Production of 1,2-Naphthoquinone-Epoxide from 1,2-Naphthoquinone under Physiological Conditions. *Chem. Res. Toxicol.* **2019**, *32* (9), 1760–1771. https://doi.org/10.1021/acs.chemrestox.9b00088.

(232) Park, J.-H.; Troxel, A. B.; Harvey, R. G.; Penning, T. M. Polycyclic Aromatic Hydrocarbon (PAH) o-Quinones Produced by the Aldo-Keto-Reductases (AKRs) Generate Abasic Sites, Oxidized Pyrimidines, and 8-Oxo-DGuo via Reactive Oxygen Species. *Chem. Res. Toxicol.* **2006**, *19* (5), 719–728. https://doi.org/10.1021/tx0600245.

(233) Cavalieri, E. L.; Li, K.-M.; Balu, N.; Saeed, M.; Devanesan, P.; Higginbotham, S.; Zhao, J.; Gross, M. L.; Rogan, E. G. Catechol Ortho-Quinones: The Electrophilic Compounds That Form Depurinating DNA Adducts and Could Initiate Cancer and Other Diseases. *Carcinogenesis* **2002**, *23* (6), 1071–1077. https://doi.org/10.1093/carcin/23.6.1071.

(234) Jerina, D. M.; Chadha, A.; Cheh, A. M.; Schurdak, M. E.; Wood, A. W.; Sayer, J. M. Covalent Bonding of Bay-Region Diol Epoxides to Nucleic Acids. *Adv. Exp. Med. Biol.* **1991**, *283*, 533–553. https://doi.org/10.1007/978-1-4684-5877-0_70.

(235) Nordqvist, M.; Thakker, D. R.; Vyas, K. P.; Yagi, H.; Levin, W.; Ryan, D. E.; Thomas, P. E.; Conney, A. H.; Jerina, D. M. Metabolism of Chrysene and Phenanthrene to Bay-Region Diol Epoxides by Rat Liver Enzymes. *Mol. Pharmacol.* **1981**, *19* (1), 168–178.

(236) Kazius, J.; McGuire, R.; Bursi, R. Derivation and Validation of Toxicophores for Mutagenicity Prediction. *J. Med. Chem.* **2005**, *48* (1), 312–320. https://doi.org/10.1021/jm040835a.

(237) Borosky, G. L.; Laali, K. K. Carbocations from Oxidized Metabolites of Benzo[a]Anthracene: A Computational Study of Their Methylated and Fluorinated Derivatives and Guanine Adducts. *Chem. Res. Toxicol.* **2006**, *19* (7), 899–907. https://doi.org/10.1021/tx060067l.

(238) LaVoie, E. J.; Tulley-Freiler, L.; Bedenko, V.; Hoffmann, D. Mutagenicity, Tumor-Initiating Activity, and Metabolism of Methylphenanthrenes. *Cancer Res.* **1981**, *41* (9 Part 1), 3441–3447.

(239) LaVoie, E. J.; Tulley-Freiler, L.; Bedenko, V.; Hoffmann, D. Mutagenicity of Substituted Phenanthrenes in Salmonella Typhimurium. *Mutat. Res. Toxicol.* **1983**, *116* (2), 91–102. https://doi.org/10.1016/0165-1218(83)90100-3.

(240) Sinsheimer, J. E.; Giri, A. K.; Hooberman, B. H.; Jung, K. Y.; Gopalaswamy, R.; Koreeda, M. Mutagenicity in Salmonella and Sister Chromatid Exchange in Mice for 1,4-, 1,3-, 2,4-, and 3,4-Dimethylphenanthrenes. *Environ. Mol. Mutagen.* **1991**, *17* (2), 93–97. https://doi.org/10.1002/em.2850170205.

(241) Todeschini, R.; Gramatica, P. New 3D Molecular Descriptors: The WHIM Theory and QSAR Applications. 26.

(242) 37th joint Meeting of the Chemical Committee and Working Part on Chemicals, Pesticides and Biotechnology. OECD Principles for The Validation, for Regulatory Purposes, of (Quantitative) Structure-Activity Relationship Models. November 2004.

(243) Lehr, R. E.; Kumar, S.; Levin, W.; Wood, A. W.; Chang, R. L.; Conney, A. H.; Yagi, H.; Sayer, J. M.; Jerina, D. M. The Bay Region Theory of Polycyclic Aromatic Hydrocarbon Carcinogenesis. *Acs Symp. Ser.* **1985**, *283*, 63–84.

(244) Bücker, M.; Glatt, H. R.; Platt, K. L.; Avnir, D.; Ittah, Y.; Blum, J.; Oesch, F. Mutagenicity of Phenanthrene and Phenanthrene K-Region Derivatives. *Mutat. Res. Toxicol.* **1979**, *66* (4), 337–348. https://doi.org/10.1016/0165-1218(79)90044-2.

(245) Vasilieva, S.; Tanirbergenov, B.; Abilev, S.; Migatchev, G.; Tapani Huttunen, M. A Comparative Study of Mutagenic and SOS-Inducing Activity of Biphenyls, Phenanthrenequinones and Fluorenones. *Mutat. Res. Lett.* **1990**, *244* (4), 321–329. https://doi.org/10.1016/0165-7992(90)90080-4.

(246) Kobets, T.; Duan, J.-D.; Brunnemann, K. D.; Vock, E.; Deschl, U.; Williams, G. M. DNA-Damaging Activities of Twenty-Four Structurally Diverse Unsubstituted and Substituted Cyclic Compounds in Embryo-Fetal Chicken Livers. *Mutat. Res. - Genet. Toxicol. Environ. Mutagen.* **2019**, *844*, 10–24. https://doi.org/10.1016/j.mrgentox.2019.06.004.

(247) Xue, W.; Warshawsky, D. Metabolic Activation of Polycyclic and Heterocyclic Aromatic Hydrocarbons and DNA Damage: A Review. *Toxicol. Appl. Pharmacol.* **2005**, *206* (1), 73–93. https://doi.org/10.1016/j.taap.2004.11.006.

(248) Flowers-Geary, L.; Bleczinski, W.; Harvey, R. G.; Penning, T. M. Cytotoxicity and Mutagenicity of Polycyclic Aromatic Hydrocarbon O-Quinones Produced by Dihydrodiol Dehydrogenase. *Chem. Biol. Interact.* **1996**, *99* (1), 55–72. https://doi.org/10.1016/0009-2797(95)03660-1.

(249) Huberman, E.; Chou, M. W.; Yang, S. K. Identification of 7,12-Dimethylbenz[a]Anthracene Metabolites That Lead to Mutagenesis in Mammalian Cells. *Proc. Natl. Acad. Sci. U. S. A.* **1979**, *76* (2), 862–866.

(250) Sen, S.; Field, J. M. Chapter Three - Genotoxicity of Polycyclic Aromatic Hydrocarbon Metabolites: Radical Cations and Ketones. In *Advances in Molecular Toxicology*; Fishbein, J. C., Heilman, J. M., Eds.; Elsevier, 2013; Vol. 7, pp 83–127. https://doi.org/10.1016/B978-0-444-62645-5.00003-1.

(251) Cavalieri, E.; Rogan, E. Role of Radical Cations in Aromatic Hydrocarbon Carcinogenesis. *Environ. Health Perspect.* **1985**, *64*, 69–84.

(252) Cavalieri, E. L.; Rogan, E. G. The Approach to Understanding Aromatic Hydrocarbon Carcinogenesis. The Central Role of Radical Cations in Metabolic Activation. *Pharmacol. Ther.* **1992**, *55* (2), 183–199. https://doi.org/10.1016/0163-7258(92)90015-R.

(253) Park, J.-H.; Gopishetty, S.; Szewczuk, L. M.; Troxel, A. B.; Harvey, R. G.; Penning, T. M. Formation of 8-Oxo-7,8-Dihydro-2¢-Deoxyguanosine (8-Oxo-DGuo) by PAH o-Quinones: Involvement of Reactive Oxygen Species and Copper(II)/Copper(I) Redox Cycling. *Chem. Res. Toxicol.* **2005**, *18* (6), 1026–1037. https://doi.org/10.1021/tx050001a.

(254) Flowers, L.; Ohnishi, S. T.; Penning, T. M. DNA Strand Scission by Polycyclic Aromatic Hydrocarbon *o* -Quinones: Role of Reactive Oxygen Species, Cu(II)/Cu(I) Redox Cycling, and *o* -Semiquinone Anion Radicals † , ‡. *Biochemistry* **1997**, *36* (28), 8640–8648. https://doi.org/10.1021/bi970367p.

(255) Park, J.-H.; Troxel, A. B.; Harvey, R. G.; Penning, T. M. Polycyclic Aromatic Hydrocarbon (PAH) o-Quinones Produced by the Aldo-Keto-Reductases (AKRs) Generate Abasic Sites, Oxidized Pyrimidines, and 8-Oxo-DGuo via Reactive Oxygen Species. 10.

(256) Jarabak, R.; Harvey, R. G.; Jarabak, J. Redox Cycling of Polycyclic Aromatic Hydrocarbon O-Quinones: Metal Ion-Catalyzed Oxidation of Catechols Bypasses Inhibition by Superoxide Dismutase. *Chem. Biol. Interact.* **1998**, *115* (3), 201–213. https://doi.org/10.1016/S0009-2797(98)00070-2.

(257) Kasai, H.; Nishimura, S. Hydroxylation of Deoxyguanosine at the C-8 Position by Ascorbic Acid and Other Reducing Agents. *Nucleic Acids Res.* **1984**, *12* (4), 2137–2145.

(258) Vlachogianni, T.; Fiotakis, C.; Valavanidis, A. 8-Hydroxy-2′ -Deoxyguanosine (8-OHdG): A Critical Biomarker of Oxidative Stress and Carcinogenesis. *J. Environ. Sci. Health Part C* **2009**, *27* (2), 120–139. https://doi.org/10.1080/10590500902885684.

(259) Guo, C.; Li, X.; Wang, R.; Yu, J.; Ye, M.; Mao, L.; Zhang, S.; Zheng, S. Association between Oxidative DNA Damage and Risk of Colorectal Cancer: Sensitive Determination of Urinary 8-Hydroxy-2′-Deoxyguanosine by UPLC-MS/MS Analysis. *Sci. Rep.* **2016**, *6* (1), 32581. https://doi.org/10.1038/srep32581.

(260) Wilk, A.; Rski, P. W.; Lassak, A.; Vashistha, H.; Lirette, D.; Tate, D.; Zea, A. H.; Koochekpour, S.; Rodriguez, P.; Meggs, L. G.; Estrada, J. J.; Ochoa, A.; Reiss, K. Polycyclic Aromatic Hydrocarbons—Induced ROS Accumulation Enhances Mutagenic Potential of T-Antigen From Human Polyomavirus JC. *J. Cell. Physiol.* **2013**, *228* (11), 2127–2138. https://doi.org/10.1002/jcp.24375.

(261) Johann, S.; Goßen, M.; Behnisch, P. A.; Hollert, H.; Seiler, T.-B. Combining Different In Vitro Bioassays to Evaluate Genotoxicity of Water-Accommodated Fractions from Petroleum Products. *Toxics* **2020**, *8* (2), 45. https://doi.org/10.3390/toxics8020045.

(262) Wulandari, R.; Lotrakul, P.; Punnapayak, H.; Amirta, R.; Kim, S. W.; Prasongsuk, S. Toxicity Evaluation and Biodegradation of Phenanthrene by Laccase from Trametes

Polyzona PBURU 12. *3 Biotech* **2021**, *11* (1), 32. https://doi.org/10.1007/s13205-020-02556-z.

(263) Drotikova, T.; Ali, A. M.; Halse, A. K.; Reinardy, H. C.; Kallenborn, R. Polycyclic Aromatic Hydrocarbons (PAHs) and Oxy- and Nitro-PAHs in Ambient Air of the Arctic Town Longyearbyen, Svalbard. *Atmospheric Chem. Phys.* **2020**, *20* (16), 9997–10014. https://doi.org/10.5194/acp-20-9997-2020.

(264) Huang, B.; Liu, M.; Bi, X.; Chaemfa, C.; Ren, Z.; Wang, X.; Sheng, G.; Fu, J. Phase Distribution, Sources and Risk Assessment of PAHs, NPAHs and OPAHs in a Rural Site of Pearl River Delta Region, China. *Atmospheric Pollut. Res.* **2014**, *5* (2), 210–218. https://doi.org/10.5094/APR.2014.026.

(265) Flocco, C. G.; Lobalbo, A.; Carranza, M. P.; Bassi, M.; Giulietti, A. M.; Mac Cormack, W. P. Some Physiological, Microbial, and Toxicological Aspects of the Removal of Phenanthrene by Hydroponic Cultures of Alfalfa (Medicago Sativa L.). *Int. J. Phytoremediation* **2002**, *4* (3), 169–186. https://doi.org/10.1080/15226510208500081.

(266) Heflich, R. H.; Neft, R. E. Genetic Toxicity of 2-Acetylaminofluorene, 2-Aminofluorene and Some of Their Metabolites and Model Metabolites. *Mutat. Res. Genet. Toxicol.* **1994**, *318* (2), 73–174. https://doi.org/10.1016/0165-1110(94)90025-6.

(267) Tang, M.; Lieberman, M. W. Quantification of Adducts Formed in DNA Treated with N-Acetoxy-2-Acetylaminofluorene or N-Hydroxy-2-Aminofluorene: Comparison of Trifluoroacetic Acid and Enzymatic Degradation. *Carcinogenesis* **1983**, *4* (8), 1001–1006. https://doi.org/10.1093/carcin/4.8.1001.

(268) Glover, S. A.; Schumacher, R. R. Mutagenicity of N-Acyloxy-N-Alkoxyamides as an Indicator of DNA Intercalation: The Role of Fluorene and Fluorenone Substituents as DNA Intercalators. *Mutat. Res. Toxicol. Environ. Mutagen.* **2021**, *863–864*, 503299. https://doi.org/10.1016/j.mrgentox.2020.503299.

(269) Rentz, J. A.; Alvarez, P. J. J.; Schnoor, J. L. Repression of Pseudomonas Putida Phenanthrene-Degrading Activity by Plant Root Extracts and Exudates. *Environ. Microbiol.* **2004**, *6* (6), 574–583. https://doi.org/10.1111/j.1462-2920.2004.00589.x.

(270) Rodrigues, A.; Nogueira, R.; Melo, L. F.; Brito, A. G. Effect of Low Concentrations of Synthetic Surfactants on Polycyclic Aromatic Hydrocarbons (PAH) Biodegradation. *Int. Biodeterior. Biodegrad.* **2013**, *83*, 48–55. https://doi.org/10.1016/j.ibiod.2013.04.001.

(271) Kwok, C.-K.; Loh, K.-C. Effects of Singapore Soil Type on Bioavailability of Nutrients in Soil Bioremediation. *Adv. Environ. Res.* **2003**, *7* (4), 889–900. https://doi.org/10.1016/S1093-0191(02)00084-9.

(272) Gottfried, A.; Singhal, N.; Elliot, R.; Swift, S. The Role of Salicylate and Biosurfactant in Inducing Phenanthrene Degradation in Batch Soil Slurries. *Appl. Microbiol. Biotechnol.* **2010**, *86* (5), 1563–1571. https://doi.org/10.1007/s00253-010-2453-2.

(273) Cho, H. Y.; Woo, S. H.; Park, J. M. Effects of Intermediate Metabolites on Phenanthrene Biodegradation. *J. Microbiol. Biotechnol.* **2006**, *16* (6), 969–973.

(274) Pérez, A. S.; Reifferscheid, G.; Eichhorn, P.; Barceló, D. Assessment of the Mutagenic Potency of Sewage Sludges Contaminated with Polycyclic Aromatic Hydrocarbons by an Ames Fluctuation Assay. *Environ. Toxicol. Chem.* **2003**, *22* (11), 2576–2584. https://doi.org/10.1897/02-416.

(275) Ubomba-Jaswa, E.; Fernández-Ibáñez, P.; McGuigan, K. G. A Preliminary Ames Fluctuation Assay Assessment of the Genotoxicity of Drinking Water That Has Been Solar Disinfected in Polyethylene Terephthalate (PET) Bottles. *J. Water Health* **2010**, *8* (4), 712–719. https://doi.org/10.2166/wh.2010.136.

(276) Ames, B. N.; Lee, F. D.; Durston, W. E. An Improved Bacterial Test System for the Detection and Classification of Mutagens and Carcinogens. *Proc. Natl. Acad. Sci. U. S. A.* **1973**, *70* (3), 782–786. https://doi.org/10.1073/pnas.70.3.782.

(277) Flückiger-Isler, S.; Kamber, M. Direct Comparison of the Ames Microplate Format (MPF) Test in Liquid Medium with the Standard Ames Pre-Incubation Assay on Agar Plates by Use of Equivocal to Weakly Positive Test Compounds. *Mutat. Res. Toxicol. Environ. Mutagen.* **2012**, *747* (1), 36–45. https://doi.org/10.1016/j.mrgentox.2012.03.014.

(278) Kamber, M.; Flückiger-Isler, S.; Engelhardt, G.; Jaeckh, R.; Zeiger, E. Comparison of the Ames II and Traditional Ames Test Responses with Respect to Mutagenicity, Strain Specificities, Need for Metabolism and Correlation with Rodent Carcinogenicity. *Mutagenesis* **2009**, *24* (4), 359–366. https://doi.org/10.1093/mutage/gep017.

(279) Cevik, M.; Dartan, G.; Ulker, M.; Bezci, K.; Deliorman, G.; Cagatay, P.; Lacin, T.; Cinel, I. H.; Aksu, B.; Keskin, Y.; Can, Z. S.; Yurdun, T.; Susleyici, B. Evaluation of Cytotoxicity and Mutagenicity of Wastewater from Istanbul: Data from Hospitals and Advanced Wastewater Treatment Plant. *Bull. Environ. Contam. Toxicol.* **2020**, *104* (6), 852–857. https://doi.org/10.1007/s00128-020-02853-6.

(280) Zwart, N.; Jonker, W.; Broek, R. ten; de Boer, J.; Somsen, G.; Kool, J.; Hamers, T.; Houtman, C. J.; Lamoree, M. H. Identification of Mutagenic and Endocrine Disrupting Compounds in Surface Water and Wastewater Treatment Plant Effluents Using High-Resolution Effect-Directed Analysis. *Water Res.* **2020**, *168*, 115204. https://doi.org/10.1016/j.watres.2019.115204.

(281) Abbas, A.; Schneider, I.; Bollmann, A.; Funke, J.; Oehlmann, J.; Prasse, C.; Schulte-Oehlmann, U.; Seitz, W.; Ternes, T.; Weber, M.; Wesely, H.; Wagner, M. What You Extract Is What You See: Optimising the Preparation of Water and Wastewater Samples for in Vitro Bioassays. *Water Res.* **2019**, *152*, 47–60. https://doi.org/10.1016/j.watres.2018.12.049.

(282) MOLTOX. Ames FT Mutagneicity Test Kit Instruction Manual. July 2, 2019.

(283) Maron, D. M.; Ames, B. N. Revised Methods for the Salmonella Mutagenicity Test. *Mutat. Res. Mutagen. Relat. Subj.* **1983**, *113* (3–4), 173–215. https://doi.org/10.1016/0165-1161(83)90010-9.

(284) Flückiger-Isler, S.; Baumeister, M.; Braun, K.; Gervais, V.; Hasler-Nguyen, N.; Reimann, R.; Van Gompel, J.; Wunderlich, H.-G.; Engelhardt, G. Assessment of the Performance of the Ames II™ Assay: A Collaborative Study with 19 Coded Compounds. *Mutat. Res. Toxicol. Environ. Mutagen.* **2004**, *558* (1–2), 181–197. https://doi.org/10.1016/j.mrgentox.2003.12.001.

(285) U.S. Environmental Protection Agency. Interim Procedures for Conducting the Salmonella/Microsomal Mutagenicity Assay (Ames Test), EPA-600 4-82-068. Environmental Monitoring Systems Laboratory, Las Vegas, NV, 89114 1983.

(286) Balashova, N. V.; Kosheleva, I. A.; Golovchenko, N. P.; Boronin, A. M. Phenanthrene Metabolism by Pseudomonas and Burkholderia Strains. *Process Biochem.* **1999**, *35* (3), 291–296. https://doi.org/10.1016/S0032-9592(99)00069-2.

(287) Puntus, I. F.; Filonov, A. E.; Akhmetov, L. I.; Karpov, A. V.; Boronin, A. M. Phenanthrene Degradation by Bacteria of the Genera Pseudomonas and Burkholderia in Model Soil Systems. *Microbiology* **2008**, *77* (1), 7–15. https://doi.org/10.1134/s0026261708010025.

(288) Ovchinnikova, A. A.; Vetrova, A. A.; Filonov, A. E.; Boronin, A. M. Phenanthrene Biodegradation and the Interaction of Pseudomonas Putida BS3701 and Burkholderia Sp BS3702 in Plant Rhizosphere. *Microbiology* **2009**, *78* (4), 433–439. https://doi.org/10.1134/s0026261709040067.

(289) Balashova, N. V.; Stolz, A.; Knackmuss, H. J.; Kosheleva, I. A.; Naumov, A. V.; Boronin, A. M. Purification and Characterization of a Salicylate Hydroxylase Involved in 1-Hydroxy-2-Naphthoic Acid Hydroxylation from the Naphthalene and Phenanthrene-Degrading Bacterial Strain Pseudomonas Putida BS202-P1. *Biodegradation* **2001**, *12* (3), 179–188. https://doi.org/10.1023/a:1013126723719.

(290) Grifoll, M.; Selifonov, S. A.; Chapman, P. J. Evidence for a Novel Pathway in the Degradation of Fluorene by Pseudomonas Sp Strain F274. *Appl. Environ. Microbiol.* **1994**, *60* (7), 2438–2449.

(291) Zhang, Z. Z.; Hou, Z. W.; Yang, C. Y.; Ma, C. Q.; Tao, F.; Xu, P. Degradation of N-Alkanes and Polycyclic Aromatic Hydrocarbons in Petroleum by a Newly Isolated Pseudomonas Aeruginosa DQ8. *Bioresour. Technol.* **2011**, *102* (5), 4111–4116. https://doi.org/10.1016/j.biortech.2010.12.064.

(292) Chupungars, K.; Rerngsamran, P.; Thaniyavarn, S. Polycyclic Aromatic Hydrocarbons Degradation by Agrocybe Sp CU-43 and Its Fluorene Transformation. *Int. Biodeterior. Biodegrad.* **2009**, *63* (1), 93–99. https://doi.org/10.1016/j.ibiod.2008.06.006.

(293) Juckpech, K.; Pinyakong, O.; Rerngsamran, P. Degradation of Polycyclic Aromatic Hydrocarbons by Newly Isolated Curvularia Sp F18, Lentinus Sp S5, and Phanerochaete

Sp T20. *Scienceasia* **2012**, *38* (2), 147–156. https://doi.org/10.2306/scienceasia1513-1874.2012.38.147.

(294) Yan, J.; Wang, L.; Fu, P. P.; Yu, H. Photomutagenicity of 16 Polycyclic Aromatic Hydrocarbons from the US EPA Priority Pollutant List. *Mutat. Res. Toxicol. Environ. Mutagen.* **2004**, *557* (1), 99–108. https://doi.org/10.1016/j.mrgentox.2003.10.004.

(295) Kinani, S.; Souissi, Y.; Kinani, A.; Vujović, S.; Aït-Aïssa, S.; Bouchonnet, S. Photodegradation of Fluorene in Aqueous Solution: Identification and Biological Activity Testing of Degradation Products. *J. Chromatogr. A* **2016**, *1442*, 118–128. https://doi.org/10.1016/j.chroma.2016.03.012.

(296) Rodrigues, A. C.; Wuertz, S.; Brito, A. G.; Melo, L. F. Fluorene and Phenanthrene Uptake by Pseudomonas Putida ATCC 17514: Kinetics and Physiological Aspects. *Biotechnol. Bioeng.* **2005**, *90* (3), 281–289. https://doi.org/10.1002/bit.20377.

(297) Mueller, J. G.; Cerniglia, C. E.; Pritchard, P. H. Bioremediation of Environments Contaminated by Polycyclic Aromatic Hydrocarbons. *Biotechnol. Res. Ser. Bioremediation Princ. Appl.* **1996**, *6*, 125–194. https://doi.org/10.1017/cbo9780511608414.007.

(298) Verhaar, H. J. M.; Ramos, E. U.; Hermens, J. L. M. Classifying Environmental Pollutants. 2: Separation of Class 1 (Baseline Toxicity) and Class 2 ('polar Narcosis') Type Compounds Based on Chemical Descriptors. *J. Chemom.* **1996**, *10* (2), 149–162. https://doi.org/10.1002/(SICI)1099-128X(199603)10:2<149::AID-CEM414>3.0.CO;2-F.

(299) Tennant, R. E.; Guesné, S. J.; Canipa, S.; Cayley, A.; Drewe, W. C.; Honma, M.; Masumura, K.; Morita, T.; Stalford, S. A.; Williams, R. V. Extrapolation of in Vitro Structural Alerts for Mutagenicity to the in Vivo Endpoint. *Mutagenesis* **2019**, *34* (1), 111–121. https://doi.org/10.1093/mutage/gey030.

(300) Vian, M.; Raitano, G.; Roncaglioni, A.; Benfenati, E. In Silico Model for Mutagenicity (Ames Test), Taking into Account Metabolism. *Mutagenesis* **2019**, *34* (1), 41–48. https://doi.org/10.1093/mutage/gey045.

(301) Eker, G.; Şengül, B.; Cindoruk, S. S. Performance Evaluation of Diethylamine to the Removal of Polycyclic Aromatic Hydrocarbons (PAHs) from Polluted Soils with Sunlight. *Polycycl. Aromat. Compd.* **2021**, *41* (2), 306–318. https://doi.org/10.1080/10406638.2019.1578809.

(302) Eker, G.; Sengul, B. Removal of Polycyclic Aromatic Hydrocarbons (PAHs) from Industrial Soil with Solar and UV Light. *Polycycl. Aromat. Compd.* **2020**, *40* (4), 1238–1250. https://doi.org/10.1080/10406638.2018.1539018.

(303) McIntosh, A. D.; Fryer, R. J.; Webster, L.; Cundy, A. B. Long-Term Fate of Polycyclic Aromatic Hydrocarbons (PAH) in Sediments from Loch Leven after Closure of an Aluminium Smelter. *J. Environ. Monit.* **2012**, *14* (5), 1335–1344. https://doi.org/10.1039/C2EM11006G.

(304) Zhao, G.; Wang, Z.; Huang, Y.; Ye, L.; Yang, K.; Huang, Q.; Chen, X.; Li, G.; Chen, Y.; Wang, J.; Zhou, Y. Effects of Benzoapyrene on Migration and Invasion of Lung Cancer Cells Functioning by TNF-α. *J. Cell. Biochem.* **2018**, *119* (8), 6492–6500. https://doi.org/10.1002/jcb.26683.

(305) Piegorsch, W. W.; Zeiger, E. Measuring Intra-Assay Agreement for the Ames Salmonella Assay. In *Statistical Methods in Toxicology*; Hothorn, L., Ed.; Lecture Notes in Medical Informatics; Springer: Berlin, Heidelberg, 1991; pp 35–41. https://doi.org/10.1007/978-3-642-48736-1_5.