Computational data-driven investigation of chemical exposome and its links to human and ecosystem health

By

Ajaya Kumar Sahoo

LIFE10201904002

The Institute of Mathematical Sciences
Chennai

A thesis submitted to the
Board of Studies in Life Sciences
In partial fulfillment of requirements
for the Degree of

DOCTOR OF PHILOSOPHY
of
HOMI BHABHA NATIONAL INSTITUTE

August 2024

Homi Bhabha National Institute

Recommendations of the Viva Voce Committee

As members of the Viva Voce Committee, we certify that we have read the dissertation prepared byAjaya Kumar Sahoo entitled: “Computational data-driven investigation of chemical exposome and its links to human and ecosystem health” and recommend that it may be accepted as fulfilling the dissertation requirement for the Degree of Doctor of Philosophy.

Sitabhra Sinha

______________________________ Date: 15/11/2024

Chair - Prof. Sitabhra Sinha

Areejit Samal

______________________________ Date: 15/11/2024

Supervisor/Convener - Prof. Areejit Samal

______________________________ Date: 15/11/2024

Member 1 - Prof. Rahul Siddharthan

______________________________ Date: 15/11/2024

Member 2 - Prof. Satyavani Vemparala

Dhiraj Kumar

______________________________ Date: 15/11/2024

Member 3 - Prof. Dhiraj Kumar

Anshu Bhardwaj

______________________________ Date: 15/11/2024

External Examiner - Prof. Anshu Bhardwaj

Final approval and acceptance of this dissertation is contingent upon the candidate’s submission of the final copies of the dissertation to HBNI.

I hereby certify that I have read this dissertation prepared under my direction and recommend that it may be accepted as fulfilling the dissertation requirement.

Date: 15/11/2024

Areejit Samal

Place: CHENNAI

Supervisor

Statement by Author

This dissertation has been submitted in partial fulfillment of requirements for an advanced degree at Homi Bhabha National Institute (HBNI) and is deposited in the Library to be made available to borrowers under rules of the HBNI.

Brief quotations from this dissertation are allowable without special permission, provided that accurate acknowledgement of source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the Competent Authority of HBNI when in his or her judgement the proposed use of the material is in the interests of scholarship. In all other instances, however, permission must be obtained from the author.

Ajaya Kumar Sahoo

Ajaya Kumar Sahoo

Declaration

I, hereby declare that the investigation presented in this thesis has been carried out by me. The work is original and has not been submitted earlier as a whole or in part for a degree or diploma at this or any other Institution or University.

Ajaya Kumar Sahoo

Ajaya Kumar Sahoo

List of Publications arising from the thesis

Journals

Published

1. *Identification of activity cliffs in structure-activity landscape of androgen receptor binding chemicals*, R.P. Vivek-Ananth[†], **A.K. Sahoo**[†], S.P. Baskaran[†], J. Ravichandran and A. Samal[⋆], Science of the Total Environment, 873: 162263 (2023). https://doi.org/10.1016/j.scitotenv.2023.162263
2. *Analysis of structure–activity and structure–mechanism relationships among thyroid stimulating hormone receptor binding chemicals by leveraging the ToxCast library*, **A.K. Sahoo**[†], S.P. Baskaran[†], N. Chivukula, K. Kumar and A. Samal[⋆], RSC Advances, 13: 23461-23471 (2023). https://doi.org/10.1039/D3RA04452A
3. *An integrative data-centric approach to derivation and characterization of an adverse outcome pathway network for cadmium-induced toxicity*, **A.K. Sahoo**[†], N. Chivukula[†], K. Ramesh, J. Singha, S.R. Marigoudar, K.V. Sharma and A. Samal[⋆], Science of the Total Environment, 920: 170968 (2024). https://doi.org/10.1016/j.scitotenv.2024.170968
4. *Leveraging integrative toxicogenomic approach towards development of stressor-centric adverse outcome pathway networks for plastic additives*, **A.K. Sahoo**[†], N. Chivukula[†], S.R. Madgaonkar, K. Ramesh, S.R. Marigoudar, K.V. Sharma and A. Samal[⋆], Archives of Toxicology, 98: 3299-3321 (2024). https://doi.org/10.1007/s00204-024-03825-z

Submitted

5. *Network-based investigation of petroleum hydrocarbons-induced ecotoxicological effects and their risk assessment*, **A.K. Sahoo**[†], S.R. Madgaonkar[†], N. Chivukula, P. Karthikeyan, K. Ramesh, S.R. Marigoudar, K.V. Sharma and A. Samal[⋆], bioRxiv 2024.07.18.604159. https://doi.org/10.1101/2024.07.18.604159

[[†] Joint-first authors; [⋆] Corresponding author(s)]

List of Publications not included in the thesis

Journals

Published

1. *MeFSAT: a curated natural product database specific to secondary metabolites of medicinal fungi*, R.P. Vivek-Ananth†, **A.K. Sahoo**†, K. Kumaravel†, K. Mohanraj and A. Samal⋆, RSC Advances, 11: 2596-2607 (2021). https://doi.org/10.1039/D0RA10322E

2. *Virtual screening of phytochemicals from Indian medicinal plants against the endonuclease domain of SFTS virus L polymerase*, R.P. Vivek-Ananth, **A.K. Sahoo**, A. Srivastava⋆ and A. Samal⋆, RSC Advances, 12: 6234-6247 (2022). https://doi.org/10.1039/D1RA06702H

3. *In silico identification of potential inhibitors of vital monkeypox virus proteins from FDA approved drugs*, **A.K. Sahoo**, P.D. Augusthian, I. Muralitharan, R.P. Vivek-Ananth, K. Kumar, G. Kumar, G. Ranganathan and A. Samal⋆, Molecular Diversity, 27: 2169–2184 (2023). https://doi.org/10.1007/s11030-022-10550-1

4. *Scaffold and Structural Diversity of the Secondary Metabolite Space of Medicinal Fungi*, R.P. Vivek-Ananth, **A.K. Sahoo**, S.P. Baskaran and A. Samal⋆, ACS Omega, 8: 3102–3113 (2023). https://doi.org/10.1021/acsomega.2c06428

5. *IMPPAT 2.0: An Enhanced and Expanded Phytochemical Atlas of Indian Medicinal Plants*, R.P. Vivek-Ananth, K. Mohanraj, **A.K. Sahoo** and A. Samal⋆, ACS Omega, 8: 8827–8845 (2023). https://doi.org/10.1021/acsomega.3c00156

6. *T9GPred: A Comprehensive Computational Tool for the Prediction of Type 9 Secretion System, Gliding Motility, and the Associated Secreted Proteins*, **A.K. Sahoo**, R.P. Vivek-Ananth, N. Chivukula, S.V. Rajaram, K. Mohanraj, D. Khare, C. Acharya and A. Samal⋆, ACS Omega, 8: 34091–34102 (2023). https://doi.org/10.1021/acsomega.3c05155

7. *Cheminformatics Analysis of the Multitarget Structure–Activity Landscape of Environmental Chemicals Binding to Human Endocrine Receptors*, S.P. Baskaran†,

A.K. Sahoo[†], N. Chivukula, K. Kumar and A. Samal[⋆], ACS Omega, 8: 49383–49395 (2023). https://doi.org/10.1021/acsomega.3c07920

8. *EPEK: Creation and analysis of an Ectopic Pregnancy Expression Knowledgebase*, A. Natarajan[†], N. Chivukula[†], G.B. Dhanakoti, **A.K. Sahoo**, J. Ravichandran[⋆] and A. Samal[⋆], Computational Biology and Chemistry, 104: 107866 (2023). https://doi.org/10.1016/j.compbiolchem.2023.107866
9. *Computational prediction of phytochemical inhibitors against the cap-binding domain of Rift Valley fever virus*, I. Muralitharan[†], **A.K. Sahoo**[†], P.D. Augusthian and A. Samal[⋆], Molecular Diversity, 28: 2637-2650 (2024). https://doi.org/10.1007/s11030-023-10702-x
10. *ViCEKb: Vitiligo-linked Chemical Exposome Knowledgebase*, N. Chivukula, K. Ramesh, A. Subbaroyan, **A.K. Sahoo**, G.B. Dhanakoti, J. Ravichandran and A. Samal[⋆], Science of the Total Environment, 913: 169711 (2024). https://doi.org/10.1016/j.scitotenv.2023.169711

[† Joint-first authors; ⋆ Corresponding author(s)]

Ajaya Kumar Sahoo

Ajaya Kumar Sahoo

Oral or Poster presentations

1. Poster presentation titled *An integrative data-centric approach to derivation and characterization of an adverse outcome pathway network for cadmium-induced toxicity* at the 17th Annual International Biocuration Conference (AIBC 2024) held at Indian Biological Data Centre, Regional Centre for Biotechnology, Faridabad, India from March 6-8, 2024.
2. Oral presentation titled *From Bench to Bytes: Prediction of Bacterial Type 9 Secretion System and Associated Functionalities* at the Contemporary Perspectives in Computational Biology held at The Institute of Mathematical Sciences, Chennai, India from February 19-20, 2024.
3. Oral presentation titled *Unravelling activity cliffs in structure-activity landscape of environmental chemicals* at the Modelling and Tackling Complex Biological Systems held at The Institute of Mathematical Sciences, Chennai, India from October 13-14, 2023.
4. Poster presentation titled *IMPPAT: An extensive resource on the phytochemical space of Indian medicinal plants and its potential applications in natural product based drug discovery* at the First National Symposium on Integrating Traditional Knowledge in Evidence Based Medicine held at Advanced Centre for Treatment, Research and Education in Cancer, Tata Memorial Centre, Navi Mumbai, India from September 21-22, 2023.
5. Poster presentation titled *T9GPred: A Predictor for Bacterial Type 9 Secretion System, Associated Gliding Motility and Secreted Proteins* at the Network Biology Day held at The Institute of Mathematical Sciences Chennai, India on July 20, 2023.
6. Poster presentation titled *Computational Prediction of Type 9 Secretion System in Bacteroidetes* at the IMSc60 Celebrations held at The Institute of Mathematical Sciences, Chennai, India from January 2-5, 2023.
7. Poster presentation titled *MeFSAT: a curated natural product database specific to secondary metabolites of medicinal fungi* at the Ayurvedic phytochemicals: recent

research and clinical implications held at Indian Institute of Technology Gandhinagar, India on October 8 2022.

8. Poster presentation titled *IMPPAT 2.0: an enhanced and expanded phytochemical atlas of Indian medicinal plants* at the 1st HBNI Theme Meeting on Life Sciences held at Raja Ramanna Centre for Advanced Technology, Indore, India from September 7-10 2022.

Ajaya Kumar Sahoo

Ajaya Kumar Sahoo

Acknowledgements

I would like to begin by expressing my sincere gratitude to my advisor, Prof. Areejit Samal, for his unwavering guidance, support, and encouragement throughout my PhD journey. His mentorship from the very beginning has been crucial in shaping my scientific progress, and his constructive criticism and insightful input have consistently enhanced the quality of my work. I am deeply thankful to him for suggesting a diverse range of research problems, including those reported in this thesis, which have significantly guided me in choosing the right and relevant scientific path for my career. Moreover, I sincerely appreciate his commitment to allocating time for my work and allowing open discussions on any subject that could influence my scientific output. Additionally, I am thankful for the opportunities he has provided, including participation in various conferences and discussions with leading researchers, which have greatly enriched my academic experience and broadened my perspective.

I am fortunate to have collaborated with several colleagues whose contributions have been instrumental to the published research reported in this thesis and other publications not included in this thesis. I am deeply grateful to N. Chivukula for his exceptional support in project design, execution and writing across many of my publications including those reported in Chapters 3-6. Similarly, I am thankful to S.P. Baskaran for her invaluable contributions to Chapters 2 and 3, and for many insightful scientific discussions we had. I would also like to extend my heartfelt thanks to Dr. R.P. Vivek-Ananth, for collaborative work on several publications during the early years of my PhD, including the research reported in Chapter 2, and for the valuable lessons in cheminformatics and computer programming on the whole. I acknowledge the efforts of S.R. Madgaonkar in finding an important paper that inspired the research presented in Chapter 5 and for leading the study reported in Chapter 6 during my absence due to a wrist fracture. Many thanks to K. Ramesh for his dedicated efforts in data curation for the studies reported across Chapters 4-6 and to K. Kumar for scientific illustrations presented in Chapter 3. I extend my deepest thanks to Mr. K.V. Sharma and Dr. S.R. Marigoudar, our collaborators in National Centre for Coastal Research, Chennai, for a fruitful collaboration which has led to publications reported in Chapters 4-6. Lastly, I thank Dr. J. Ravichandran, Dr. P. Karthikeyan and Dr. J. Singha for collaboration and scientific discussions while conducting some of the research reported in this thesis.

Furthermore, I am grateful to Prof. Ashutosh Srivastava for collaboration in one of my publications not included in this thesis, for teaching me molecular dynamics simulation, and for providing valuable tips and techniques in creating high-quality scientific illustrations. I would also like to thank Prof. Celin Acharya for valuable collaboration in another publication not included in this thesis. I would like to extend my appreciation to Dr. K. Mohanraj, G.B. Dhanakoti and Dr. J. Ravichandran for their assistance in creating databases and webserver associated with many of my publications. I also thank other co-authors namely, K. Kumaravel, Dr. K. Mohanraj, P.D. Augusthian, I. Muralitharan, G. Kumar, G. Ranganathan, A. Natarajan, G.B. Dhanakoti, S.V. Rajaram, D. Khare and Dr. A. Subbaroyan for their contributions in my publications.

Next, I would like to extend gratitude to all my doctoral committee members—Prof. Sitabhra Sinha, Prof. Rahul Siddharthan, Prof. Satyavani Vemparala and Prof. Dhiraj Kumar—for their constructive comments and suggestions. I also wish to thank all the faculty members, including Prof. A. Samal, Prof. S. Sinha, Prof. R. Siddharthan, Prof. S. Vemparala, and Prof. S. Krishnaswamy, who taught me during the PhD coursework.

My sincere thanks to Prof. Anshu Bhardwaj, Prof. Madhulika Dixit, Prof. Sanjay Jain, Prof. Mohit Kumar Jolly, Prof. Dhiraj Kumar, Prof. Vinay K Nandicoori, Prof. Basanta Kumar Nayak, Prof. Prasun Kumar Mukherjee, Prof. Mitali Mukerji, Prof. Sagar Pandit, Prof. Kaustuv Sanyal, Prof. Nagamani Sukumar and Prof. Smita Srivastava with whom I had the opportunity to engage in detailed scientific discussions on my research. Additionally, I was fortunate to present my research to several faculty members during their visits to IMSc and at conferences elsewhere.

I would like to thank the system administrators Mr. B. Raveendra Reddy, Mr. G. Srinivasan for their technical support. I am also grateful to the library staff for their prompt assistance in ordering books upon request. My sincere thanks also go to the IMSc administration for their support in various occasions. Additionally, I extend my appreciation to IMSc and DAE for the generous funding.

Thanks to the current and past lab members—B.S. Karthikeyan, Janani, Vivek, K. Mohanraj, Ajay, Yasharth, Priya, Nikhil, Priyotosh, Madhumita, Shreyes, Kishan, Gaurav, Geetha, Gokul, Priya Shree, Vinayak, Priya Dharshini, Ishwarya, Pavithra, Kundhanathan, Saumitra and Jayalakshmi—for their help and support. I will always be grateful to Janani, Ajay, Nikhil and Priya for their unwavering support and kindness during challenging moments. I would also like to express my gratitude to Dhananjaya for his help during my initial period at IMSc, and to Garima for taking time to engage in some meaningful discussions with me. Additionally, I extend my thanks to the canteen, housekeeping, and security staff for their diligent care of food and other essential needs during my stay at the IMSc hostel.

Lastly, I would like to extend my deepest gratitude to my parents, Minati and Ravi, my brother Anshuman, and all other family members for their love and encouragement throughout my journey. Their support and belief in my potential have been a constant source of inspiration. Special thanks to Supriya for her steadfast presence in every circumstance. Her consistent support and motivation has been instrumental in helping me stay committed to my goals. I am also grateful to all my friends for their support during various moments of my life. To conclude, I offer my heartfelt thanks to all who have supported me in different times, and I am grateful to the almighty for guiding and blessing me throughout this journey.

Ajaya Kumar Sahoo

Ajaya Kumar Sahoo

Contents

List of Figures

List of Tables

Abstract

Humans and ecosystems are frequently exposed to myriad of chemicals, including those found in consumer products, industrial pollutants, and pesticides, which collectively constitute the chemical exposome. These chemicals can persist in the environment and bioaccumulate, leading to detrimental effects on humans and other organisms, as well as long-term ecological impacts. Therefore, it is imperative to characterize the chemical exposome and assess its impact on human and ecosystem health. To this end, traditional toxicity testing often relies on animal models which can be low-throughput, expensive and time consuming, and therefore, computational approaches have emerged as effective alternatives to expedite the characterization of the ever-expanding chemical exposome. In this thesis, we employ various computational approaches to characterize the structure-activity landscape and structure-mechanism relationship among environmental chemicals within the chemical exposome. Further, we investigate chemical-induced health effects on humans and ecosystems through the adverse outcome pathway (AOP) framework.

For the characterization of structure-activity landscape of endocrine disruptors among environmental chemicals, we focus on two distinct chemical spaces, namely, androgen receptor (AR) binding chemicals and thyroid stimulating hormone receptor (TSHR) binding chemicals. In both cases, we employ several computational approaches to analyze heterogeneity in the structure-activity landscape of these chemical spaces and identify activity cliffs, i.e., structurally similar chemicals exhibiting large differences in their activities against a target receptor. Further we classify the identified activity cliffs based on their structural features. Additionally, we analyze the structure-mechanism relationships of the TSHR binding chemicals and identify structurally similar chemicals differing in their mechanism of actions. In sum, the inferences from these computational analyses will aid in development of improved toxicity predictors for characterization of the chemical exposome.

Next, we investigate the adverse health effects induced by environmental chemicals, by focusing on certain classes of chemicals namely, heavy metal - cadmium, plastic additives and petroleum hydrocarbons (PHs), through AOP framework. In each case, we curate a list of chemicals by relying on published reports and existing resources. We then integrate biological endpoint data from various toxicological resources to identify associations between the chemicals with the high quality and complete AOPs within AOP-Wiki. Thereafter, we utilize these chemical-AOP associations to construct chemical-specific AOP networks and analyze toxicity pathways to understand the mechanisms underlying chemical-induced adverse effects in both humans and ecological species. Further, we assess the toxicities of the PHs across diverse ecological species using network-based approaches and perform ecological risk assessment. In conclusion, this thesis presents a systematic computational approach that integrates heterogeneous toxicological data to investigate environmental chemicals and their adverse effects on humans and ecosystems, offering a holistic overview of the chemical exposome and its health implications from a One Health perspective.

Chapter 1

Introduction

Chemical corruption of the globe affects us from conception to death. Like the rest of nature, we are vulnerable to pesticides; we too are permeable. All forms of life are more alike than different.

- Linda Lear

1.1 Motivation

The concept of exposome encompasses a variety of environmental factors, such as chemicals, radiation and microbes, that interact with different species throughout their lifespan and have the potential to affect their health outcomes [1–4]. The exposome complements the genome by helping us understand how environmental and genetic factors interact to influence health and disease [1]. The chemical exposome, a key component of the broader exposome concept, constitute myriad of chemicals including chemicals in consumer products, industrial pollutants, and pesticides, among others [5]. These chemicals are released into the environment through various anthropogenic activities, where they can persist, bioaccumulate and potentially cause harmful effects on human health and diverse ecological species [6–9]. Therefore, a systematic investigation of these environmental chemicals and understanding their adverse biological effects is important for linking the chemical exposome to human and ecosystem health.

The space of chemical exposome is continuously expanding due to the advancements made by chemists and rapid industrialization. PubChem and Chemical Abstracts Service (CAS) registry are two of the largest repositories of chemical information, containing approximately 115 million chemical structures and 110 million substances, respectively to date [5]. However, among these known chemicals, less than 1% have been experimentally tested for their biological activity, including toxicity [5], largely due to the reliance on extensive animal testing strategies. To address this, alternative approaches that include computational and high-throughput *in vitro* strategies, are being sought that can quickly, efficiently, and cost-effectively screen vast numbers of chemicals [10]. In particular, computational approaches relying on existing chemical information can be employed to prioritize chemicals for further testing and providing mechanistic insights valuable for refining testing strategies [11, 12]. In short, computational approaches can be leveraged to expedite the characterization of the ever-expanding chemical exposome and elucidate their adverse impact on human and ecosystem health.

This thesis aims to utilize diverse computational approaches to achieve two broad objectives, where the first objective is to characterize the structure-activity landscape and structure-mechanism relationship among endocrine disrupting chemicals. Under this objective, we focus on the following research questions:

- Is there heterogeneity in the structure-activity landscape of environmental chemicals? If so, employ computational approaches to characterize such heterogeneity, in particular, identify ‘activity cliffs’ in the chemical space.
- What structural features are responsible for the formation of activity cliffs, and how can such features be identified?
- Is there heterogeneity in the structure-mechanism relationship of endocrine receptor binding chemicals? If so, employ computational approaches to explore such heterogeneity in the associated chemical space.

We address these questions in Chapters 2 and 3 of this thesis by leveraging chemical-activity data of environmental chemicals that bind to androgen receptor (AR) and thyroid stimulating hormone receptor (TSHR), respectively.

The second objective of this thesis is to investigate chemical-induced health effects through adverse outcome pathway (AOP) framework. Under this objective, we focus on following research questions:

- How can diverse toxicological data be integrated to enhance our understanding of chemical-induced toxicities?
- How can integration of network-based approaches with the AOP framework provide additional insights into the underlying mechanisms driving adverse effects associated with environmental chemicals?
- How can current regulations and product data be utilized to identify existing restrictions and pinpoint potential sources of exposure to these chemicals?
- Can we leverage ecotoxicologically relevant data to perform risk assessment and identify vulnerable species exposed to these chemicals?

We address these questions in Chapters 4, 5 and 6 of this thesis by focusing on a heavy metal - cadmium, plastic additives and petroleum hydrocarbons (PHs), respectively.

1.2 Environmental chemicals of concern

In this thesis, we have investigated several classes of environmental chemicals that pose significant risks to both human and ecosystem health. These include endocrine disrupting chemicals (EDCs), heavy metals, plastic additives and petroleum hydrocarbons (PHs). In the following, we provide a detailed description of these environmental chemicals of concern.

Endocrine disrupting chemicals

Endocrine disrupting chemicals (EDCs) are synthetic or naturally occurring compounds that can bind to hormone receptors, either stimulating or blocking hormonal activity, thereby disrupting body's normal physiological functions [8, 13, 14]. EDCs encompass

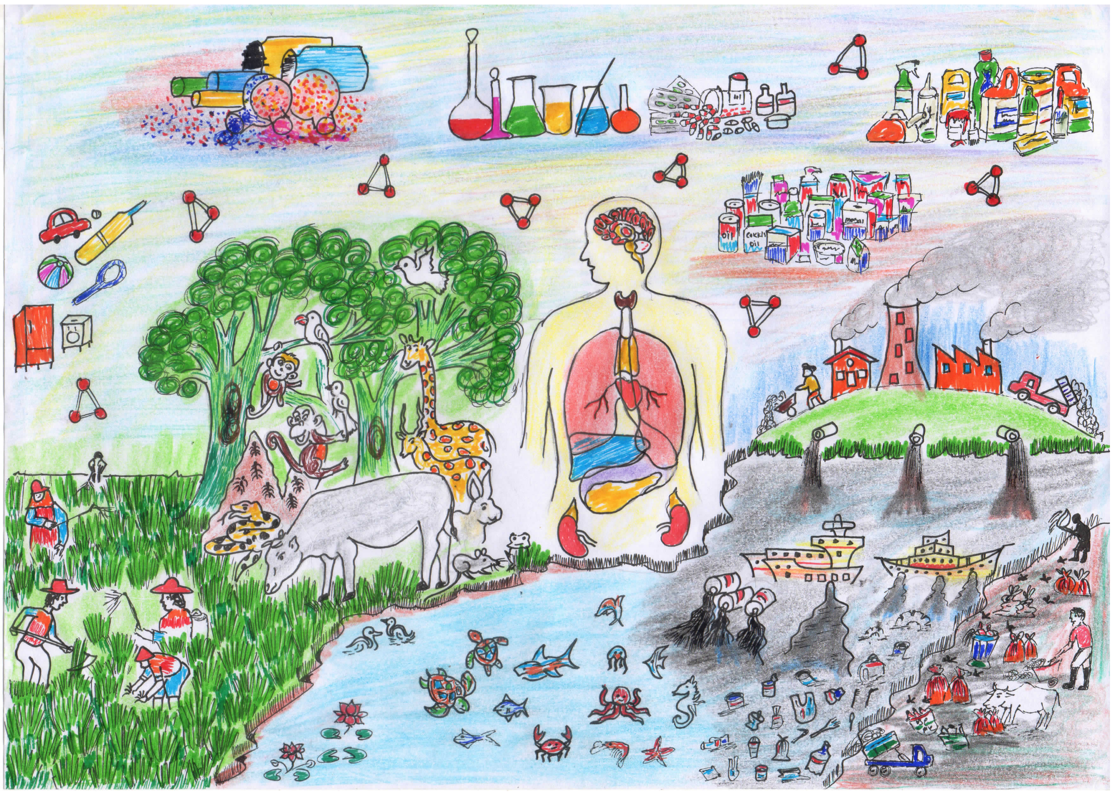


Figure 1.1: An overview of various environmental exposure sources contributing to the chemical exposome and their impact on human and ecosystem health.

a wide range of environmental chemicals including those in consumer products, synthesized in industry, pesticides, drugs and others (Figure 1.1) [8, 13, 14]. Importantly, these chemicals of concern are continuously subjected to various regulations worldwide due to their potential to cause different toxicities, including carcinogenic effects, developmental, reproductive, neurological and metabolic disorders [8, 13–15]. In this thesis, we have employed several computational approaches to investigate the chemical space of EDCs, contributing to a broader perspective on their potential health impacts. In Chapter 2, we have analyzed the structure-activity landscape of EDCs binding to the androgen receptor (AR), reported in published literature. In Chapter 3, we have extended our analysis to the structure-activity and structure-mechanism relationships of EDCs binding to the thyroid stimulating hormone receptor (TSHR), curated from the ToxCast [16] chemical library.

Heavy metals

Heavy metals including cadmium, arsenic, lead, chromium and copper are naturally occurring elements which are characterized by their high atomic weight and density [17]. Various anthropogenic activities have led to the release of these chemicals into the environment, where they can bioaccumulate and pose considerable risks to both human and ecosystem health (Figure 1.1) [17]. Notably, the Agency for Toxic Substances and Disease Registry (ATSDR) and the International Agency for Research on Cancer (IARC) have classified many of these heavy metals as carcinogens based on epidemiological studies in humans and experimental research in animals [17, 18]. Additionally, heavy metal exposure can cause non-carcinogenic effects including kidney damage, neuronal dysfunction, cardiovascular disease and developmental disorders in different species [17, 18]. In Chapter 4 of this thesis, we focused on cadmium and its inorganic compounds and analyzed their toxicities using computational approaches.

Plastic additives

Plastics are synthetic compounds which have become ubiquitous in the environment due to its low production cost, durability and adaptability [19]. Plastic additives are chemicals that are intentionally added to polymers during the plastic production process to achieve desired properties in the final product such as increase flexibility, reduce flammability and pigmentation [20,21]. Moreover, additives in plastics used in sectors such as food packaging, children's products including toys, agricultural products, and personal care products are more likely to come into contact with humans and other species in the ecosystem (Figure 1.1) [20]. As these additives are not covalently bonded to the plastic polymer, they can be easily leached into the surroundings under various environmental stresses [20, 21]. Environmental exposure to such additives has been linked to various adverse health effects, including cancer, endocrine disruptions, developmental and metabolic disorders in both humans and ecological species [6, 20, 21]. In Chapter 5 of this thesis, we curated a list of plastic additives from a detailed published report and analyzed their toxicities in humans and other species using computational approaches.

Petroleum hydrocarbons

Petroleum hydrocarbons (PHs) are organic compounds mainly composed of carbon and hydrogen, which are primarily derived from crude oil and its derivatives such as gasoline, kerosene, diesel and others [22]. PHs are complex mixtures of alkanes, alkenes and aromatic hydrocarbons, which can contaminate the environment via oil spills, transportation, offshore drilling and industrial discharge (Figure 1.1) [22]. In both ecological species and human, PHs can be absorbed through inhalation or dermal contact, leading to a range of toxic effects, including carcinogenic, developmental and endocrine disorders [22–26]. Moreover, the United States Environmental Protection Agency (US EPA) has identified 16 polycyclic aromatic hydrocarbons (PAHs) as priority pollutants due to their environmental prevalence and persistence. In Chapter 6 of this thesis, we curated a list of PHs from published literature and analyzed their toxicities in ecological species using network-based approaches.

Globally, several initiatives have been taken to develop large-scale knowledgebases for such environmental chemicals of concern which can aid in assessing their adverse health effects. To this end, the ToxCast program [16] developed by the US EPA has experimentally screened nearly 10000 environmental chemicals, including pesticides, food additives, chemicals in plastics, and industrially synthesized chemicals across various biological targets and profiled their bioactivity. As one of the largest and most comprehensive resources providing standardized experimental data on such diverse chemicals, ToxCast can significantly aid in the development of *in silico* predictive models for evaluating chemical toxicity. The Comparative Toxicogenomics Database (CTD) is one of the largest databases which compiles associations among chemicals, genes or proteins, phenotypes, and diseases upon chemical exposure, based on gathered information from published literature, to understand the effects of chemical exposome on human health [27]. ECOTOX is the largest resource on environmental toxicity, cataloging manually curated data on the impact of nearly 13000 chemicals on approximately 14000 aquatic or terrestrial species, to support ecological risk assessment [28]. Additionally, specialized resources have been developed focusing on particular diseases or classes of chemicals, further advancing research on chemical exposome [8,15,29]. In sum, the heterogeneous toxicological datasets from these available resources can be leveraged to assess toxicity of chemicals and predict their adverse biological effects.

1.3 Cheminformatics based approaches to investigate the structure-activity and structure-mechanism relationships in chemical datasets

Structure-activity relationship (SAR) based analyses are fundamental in contemporary toxicology as the obtained insights can facilitate the prediction and characterization of chemical toxicity through their application in both quantitative and qualitative models [30]. These analyses relate structure of the chemical with its biological activity, which enable identification of structural features associated with specific toxicological endpoints, thereby aiding in the development of toxicity predictors [30]. To this end, activity landscape analysis offers a powerful means to visualize and explore the SAR data [31]. In the structure-activity landscape of chemicals, smooth regions indicate continuous SAR, wherein small structural modification leads to minor change in the chemical activity, while the rugged regions indicate discontinuous SAR, wherein small structural modification can lead to a significant shift in the chemical activity [31, 32]. This discontinuous SAR region consists of activity cliffs, which are defined as structurally similar chemical pairs that exhibit large differences in their activity values [32].

The presence of discontinuity in the SAR data may indeed indicate the existence of activity cliffs and is not necessarily due to any experimental errors [33]. Quantitative structure-activity relationship (QSAR) models trained using SAR data containing activity cliffs can result in inaccurate predictions as these models rely on linear parameters and linear variables, making them inadequate for capturing the discontinuous SAR regions [33]. Importantly, activity cliffs can be thought of as artifacts arising from the choice of chemical fingerprints, which may not be suitable for a given dataset in capturing the minor structural variations that could lead to large differences in chemical activity [34]. In this context, we have utilized various computational approaches in this thesis to investigate the structure-activity landscape of environmental chemicals binding to several endocrine receptors. Specifically, we utilized computational approaches that rely on chemical fingerprints based similarity and substructure based similarity to uncover the heterogeneity in the structure-activity landscape (activity cliffs) of the chemicals. In the following, we provide a detailed explanation of such computational approaches employed in this thesis.

Structure-activity similarity (SAS) map

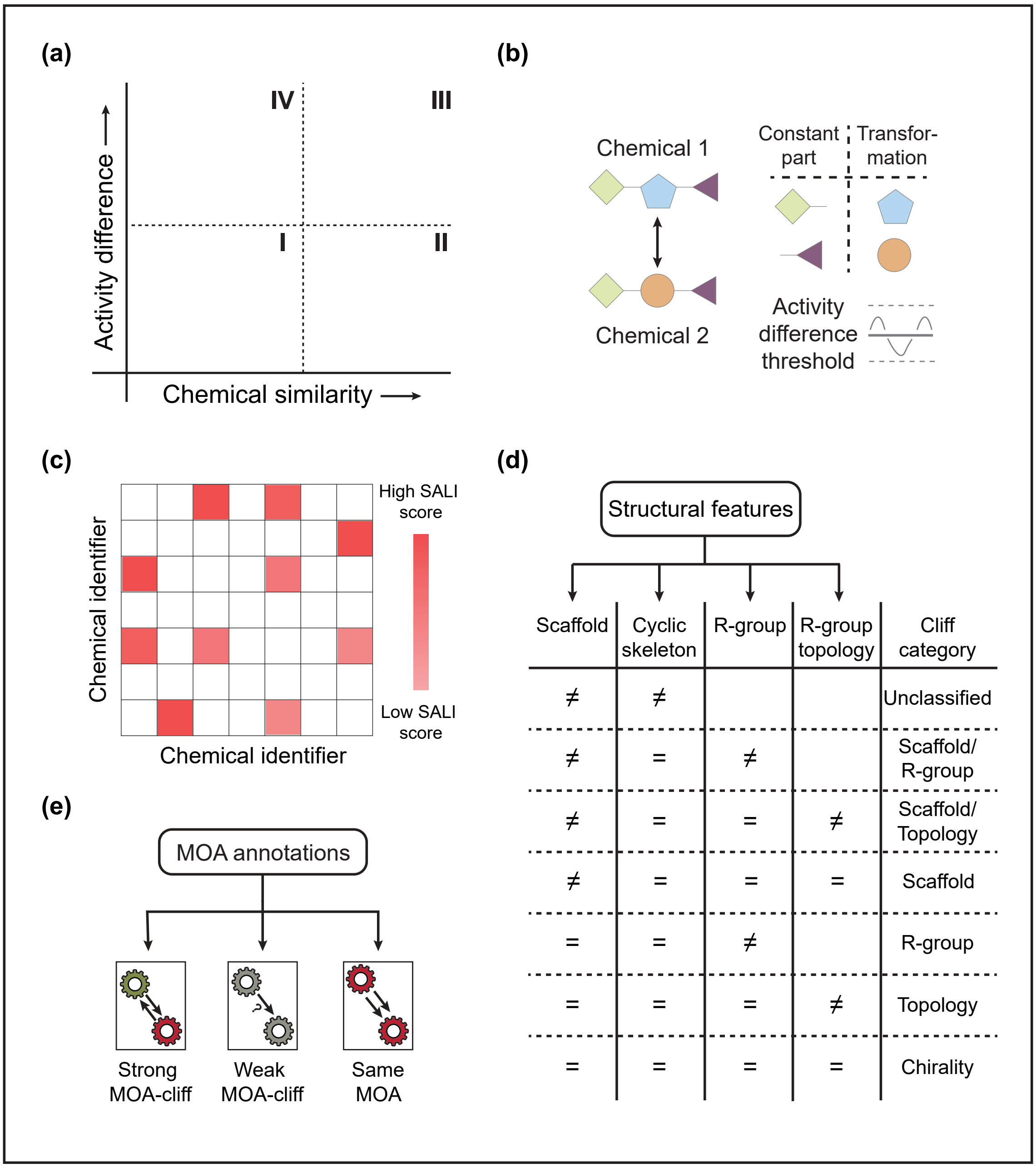
The structure-activity similarity (SAS) map is a two-dimensional (2D) representation of the structure-activity landscape of chemicals, initially proposed by Shanmugasundaram and Maggiora [35]. In a SAS map, the x-axis represents chemical similarity values computed using molecular fingerprints for pairs of chemicals, while the y-axis represents the absolute values of activity differences between these chemical pairs (Figure 1.2a). Each data point in the SAS map denotes one or more chemical pairs that have the corresponding chemical similarity and activity difference values. By using appropriately chosen thresholds to the chemical similarity and activity difference values, the SAS map can be divided into four quadrants (Figure 1.2a), which provide information on different regions of the structure-activity landscape of the chemicals. In the SAS map, quadrant I represents chemical pairs that are structurally very different but have similar activity values, indicating scaffold hops, quadrant II represents chemical pairs that are both structurally and activity-wise similar, indicating smooth region of the landscape, quadrant III represents chemical pairs that are structurally similar but have large differences in their activity values, indicating the activity cliff region, and quadrant IV represents chemical pairs that are structurally dissimilar and have large differences in their activity values (Figure 1.2a). Medina-Franco and colleagues have extensively employed the SAS map approach in distinct studies to identify activity cliffs among chemicals binding to peroxisome-proliferator-activated receptors [36], to DNA methyltransferases [37] and to estrogen receptor [38].

Structure-activity landscape index (SALI)

Guha and Van Drie developed the structure-activity landscape index (SALI) scoring method to numerically quantify activity cliffs in chemical datasets by directly comparing activity differences with chemical similarity [39]. The SALI score is calculated by dividing the absolute difference in activity values between two chemicals by one minus their chemical similarity computed using molecular fingerprints based approach [39]. Additionally, Guha and Van Drie introduced a SALI heatmap (Figure 1.2b) to analyze the SAR of chemical datasets. In this heatmap, the x- and y-axis represents the chemicals arranged according to their activity values, with each cell in the heatmap colored based on the computed SALI score for the corresponding chemical pair, where darker colors indicate higher SALI scores (Figure 1.2b). In the SALI based approach, chemical pairs with high SALI scores represent activity cliffs in the chemical dataset [39].

Matched molecular pair (MMP)

A matched molecular pair (MMP) is defined as a pair of chemicals that differ only at a single site and are characterized by a well defined transformation at that site [40]. Hussain and Rea previously proposed a chemical decomposition method based on the chemical fragments and fragment indexing to construct MMPs within chemical datasets [41]. Building on this approach, Dalke *et al.* developed an open-source platform namely mmpdb, which can more efficiently perform chemical fragmentation and fragment indexing in large chemical datasets [42]. By comparing the fragments of two different chemicals, the constant and transformed parts of the fragments can be identified (Figure 1.2c). Further, Hu *et al.* introduced specific criteria for the constant and transformed parts of the



Scaffold	Cyclic skeleton	R-group	R-group topology	Cliff category
≠	≠			Unclassified
≠	=	≠		Scaffold/ R-group
≠	=	=	≠	Scaffold/ Topology
≠	=	=	=	Scaffold
=	=	≠		R-group
=	=	=	≠	Topology
=	=	=	=	Chirality

Figure 1.2: Schematic summary of the different cheminformatics based approaches employed to analyze the structure-activity and structure-mechanism relationships among chemicals in this thesis. (a) SAS map approach to analyze the structure-activity landscape. (b) MMP approach to identify activity cliffs. (c) SALI heatmap to identify activity cliffs. (d) Classification of activity cliffs using different structural features. (e) MOA-based classification of chemical pairs.

fragments to generate size-restricted MMPs [43]. They also applied activity difference threshold on the chemical pairs forming size-restricted MMPs to identify MMP based activity cliffs, also termed as MMP-cliffs (Figure 1.2c) [43]. Notably, the MMP based approach does not rely on chemical fingerprints based similarity but instead focuses on structural transformations between chemical fragments to analyze the structure-activity landscape of chemicals (Figure 1.2c).

Chemical substructure based classification of activity cliffs

Previously, Hu and Bajorath proposed a method distinct from the MMP based approach, focusing on chemical substructures to identify the activity cliffs [44]. Specifically, they utilized various substructures of the chemical pairs namely, molecular scaffold (core structural framework), cyclic skeleton (scaffold topology), R-groups (side chains connected to the scaffold), and R-group topology (connectivity of the R-groups), to classify activity cliffs into following structural categories (Figure 1.2d):

(i) Chirality cliff: A chemical pair having same scaffold, R-groups and R-group topology.

(ii) Topology cliff: A chemical pair having same scaffold and R-groups but differ in the R-group topology.

(iii) R-group cliff: A chemical pair having same scaffold but differ in R-groups.

(iv) Scaffold cliff: A chemical pair having different scaffold but same cyclic skeleton and R-groups.

(v) Scaffold/Topology cliff: A chemical pair having different scaffold and R-group topology but same cyclic skeleton and R-groups.

In our work [45, 46], we have introduced another category Scaffold/R-group cliff, wherein a chemical pair have different scaffolds and R-groups but same cyclic skeleton. Notably, the structural classification of the activity cliffs does not rely on chemical fingerprints based similarity but instead focuses on chemical substructures, which can reveal the structural features behind the formation of activity cliffs in chemical datasets [44].

Mechanism of action (MOA)-based classification of chemicals

Similar to the presence of activity cliffs in SAR data, structurally similar chemicals can also differ in their mechanism of action (MOA). Thus, analyzing the structure-mechanism relationship among chemicals can reveal heterogeneity, namely MOA-cliffs, which highlight cases wherein minor structural changes can lead to differences in their MOA. Previously, Hao *et al.* have leveraged the MOA annotations of chemicals tested for binding to androgen receptor (AR) in both agonist and antagonist assays and classified the structurally similar chemical pairs into the following categories (Figure 1.2e) [47]:

(i) Strong MOA-cliff: A chemical pair in which both the chemicals have opposite MOA annotations.

(ii) Same MOA: A chemical pair in which both the chemicals have same MOA annotations.

(iii) Weak MOA: A chemical pair which could not be classified as either strong MOA or same MOA.

This MOA-based classification of chemicals can reveal subtle structural variations that can lead to opposite MOA of chemicals, and such insights can be crucial in the development toxicity predictors.

In Chapter 2 of this thesis, we have employed SAS map and SALI heatmap to identify activity cliffs among chemicals binding to the AR. We then analyzed the substructures of these activity cliffs to classify them into various structural categories. In Chapter 3 of this thesis, we utilized the SAS map and MMP approach to identify activity cliffs among chemicals binding to the thyroid stimulating hormone receptor (TSHR). Additionally, we used MOA annotations reported in both TSHR agonist and antagonist assays to identify MOA-cliffs. Overall, by leveraging these cheminformatics based approaches, we have examined in this thesis, the heterogeneity in the structure-activity and structure-mechanism relationships among environmental chemicals targeting several endocrine receptors.

1.4 Linking chemical exposome and health by leveraging adverse outcome pathway framework

Historically, chemical toxicity testing has largely depended on animal models, which are low throughput, time-consuming and expensive. To address this challenge, the United States National Research Council (US NRC) has published a landmark report titled 'Toxicity Testing in the 21st Century: A Vision and a Strategy', which emphasized the need for computational and high-throughput *in vitro* approaches to enable rapid, efficient and cost-effective screening of chemicals [10]. The report further recommended using the concept of toxicity pathways as a foundation for new approach to toxicity testing and chemical risk assessment. Toxicity pathways are described as cellular response pathways, which when perturbed by environmental agents, can potentially lead to adverse health effects [10]. Inspired by this report, Ankley *et al.* proposed a conceptual framework that involves aggregation and organization of existing mechanistic data on adverse outcomes induced by chemical exposure, and termed it as adverse outcome pathway (AOP) [48]. AOPs can serve as foundation for Integrated Approaches to Testing and Assessment (IATA) by integrating existing data and facilitating chemical prioritization based on their associated risks, thereby enhancing the efficiency and effectiveness of regulatory assessments [49].

To this end, several efforts are being made to develop AOPs tailored to chemical-induced toxicological events in humans and ecological species, with the aim of improving chemical risk assessment [50–52]. Subsequently, the Organisation for Economic Co-operation and Development (OECD) launched an international effort to develop AOPs that can guide chemical risk assessment and associated regulatory decisions. In particular, this led to the establishment of a user-friendly open-source repository namely, AOP-Wiki [53], that has enabled collaborative development and evaluation of AOPs. In brief, AOP-Wiki has served as a critical platform for the structured organization of toxicological knowledge and enabled the systematic construction of AOPs.

An AOP is a conceptual framework to organise toxicological knowledge that consists of sequentially ordered biological events underlying a stressor (e.g., chemical) induced adverse biological outcome [48, 54]. In an AOP, the biological events are organized into modular constructs termed as key events (KEs) and are connected to each other through directed links termed as key event relationships (KERs) (Figure 1.3a) [54–56]. The originating KE in an AOP that involves the molecular interaction between the chemical and the biological target is termed as molecular initiating event (MIE) (Figure 1.3a) [57]. The anchored biological events that are at organ or higher levels of biological organization and are of regulatory relevance, are termed as adverse outcomes (AOs) (Figure 1.3a) [57].

The linear set-up of individual AOPs limits its ability to capture the biological complexity and diversity of toxicity pathways induced by a chemical or exhaustively capture all perturbations leading to an adverse biological outcome [58]. To address this knowledge gap, the concept of AOP networks was proposed (Figure 1.3b) [58, 59]. An AOP network can be constructed by assembling two or more AOPs through their shared KEs (Figure 1.3b). [58, 59]. Notably, AOP network can highlight interactions among individual AOPs enabling the understanding of complex toxicity pathways [58–63].

Till date, AOP-Wiki has been leveraged to build over 32 different AOP networks for a variety of adverse outcomes such as reproductive disorders [64–67], neurologic disorders [68–71], endocrine disorders [58, 72–77], developmental disorders [66, 76, 78, 79], hepatic disorders [58, 80, 81], pulmonary disorders [82–84] and other effects [59, 63, 85–92]. These AOP networks also include those built to analyze adverse outcomes induced by specific chemicals [64, 67]. Since AOPs were envisaged to be agnostic to stressor or chemical associations, integrating datasets from various toxicological resources has facilitated the identification of novel chemical-AOP associations [58, 64, 77, 83, 93]. This approach has significantly expanded the coverage of chemical-induced adverse outcomes, enabling the construction of more comprehensive AOP networks. Notably, such constructions of AOP networks provide novel insights into chemical-induced toxicities that would not have been possible by relying solely on the data contained in AOP-Wiki.

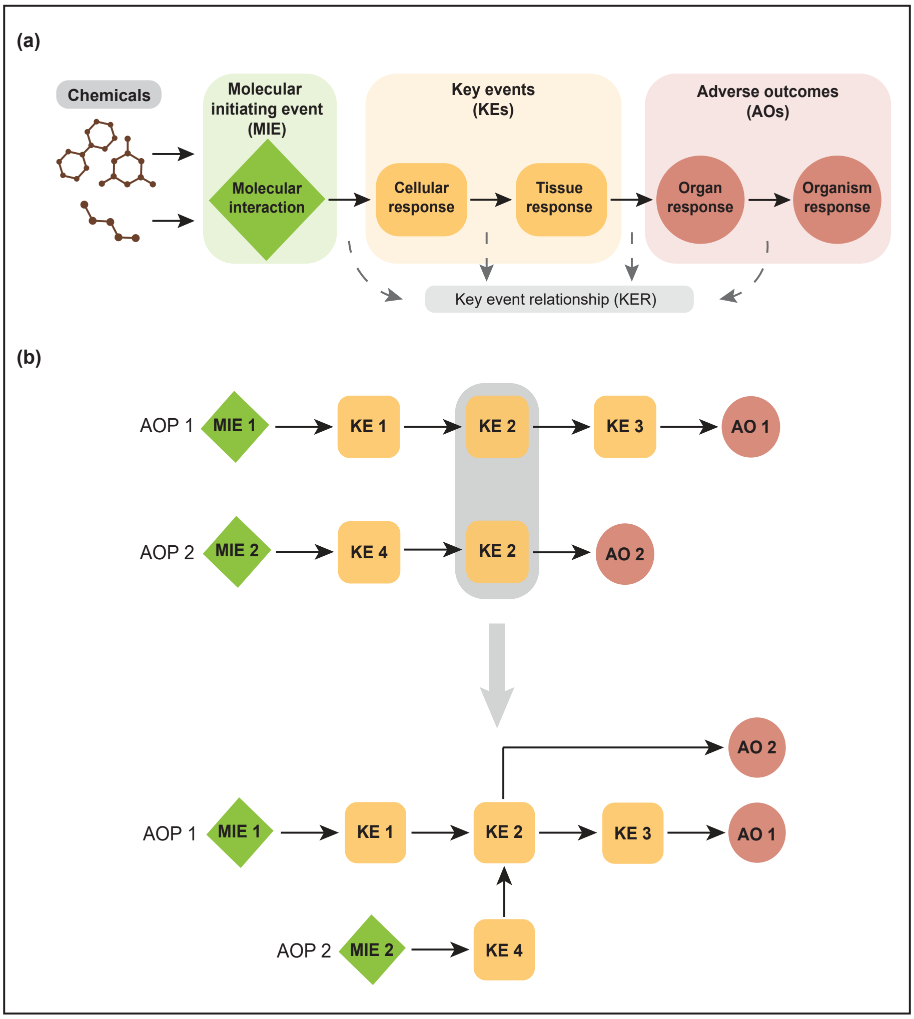


Figure 1.3: Schematic figure describing the adverse outcome pathway (AOP) framework and AOP network. (a) An AOP comprises of key events (KEs) which include molecular initiating events (MIEs) and adverse outcomes (AOs) spanning different biological levels of organization. In an AOP, two KEs are linked through a directed relationship referred to as key event relationship (KER). (b) An AOP network constructed by assembling two different AOPs via their shared KEs.

In this context, we integrated heterogeneous datasets from several toxicological resources with existing AOPs compiled in AOP-Wiki to construct a variety of AOP networks that link environmental chemical exposures to their adverse health effects in humans and ecological species. In the following, we provide a detailed explanation of different networks from stressor and AOP perspective constructed and analyzed in this thesis.

Stressor-AOP network

A stressor-AOP network visualizes the relationships between stressors (chemicals) and their associated AOPs (Figure 1.4a), aiding in the understanding of stressor-induced adverse biological effects. Previously, Aguayo-Orozco *et al.* constructed a stressor-AOP network by linking chemicals screened in ToxCast to AOPs within AOP-Wiki by leveraging assay endpoint data from ToxCast [93]. This network construction facilitated exploration of stressor-induced adverse effects from a mechanistic perspective. In this thesis, we have utilized biological endpoint data for environmental chemicals from various toxicological resources to construct stressor-AOP networks. Additionally, we have introduced two criteria to analyze the stressor-AOP network: (a) *coverage score*, which quantitatively evaluates the significance of the stressor-AOP association based on the fraction of KEs in the AOP linked to the stressor through our data integrative approach, and (b) *level of relevance*, which qualitatively assesses the significance of the stressor-AOP association based on the relationship between the stressor and the types of KEs in the AOP. Notably, these criteria aided in the identification of highly relevant AOPs associated with the stressor of interest.

Undirected AOP network

An undirected AOP network represents the relationships between various AOPs that are linked through shared KEs (Figure 1.4b) [58, 59]. Construction of such a network can highlight connected components (where two or more AOPs are connected) (Figure 1.4b), which can further reveal interactions between different toxicity pathways associated with the stressor [77]. In this thesis, we constructed such undirected AOP networks for environmental chemicals by identifying highly relevant AOPs from the stressor-AOP network. Thereafter, we have analyzed these networks to understand chemical-induced toxicities in different species.

Directed AOP network

A directed AOP network is a structured representation that illustrates interactions among AOPs through their KEs and KERs (Figure 1.4c). In this network, KEs including MIEs and AOs can be arranged according to their biological levels of organization (Figure 1.4c). Unlike stressor-AOP and undirected AOP networks, the directed AOP network provides a

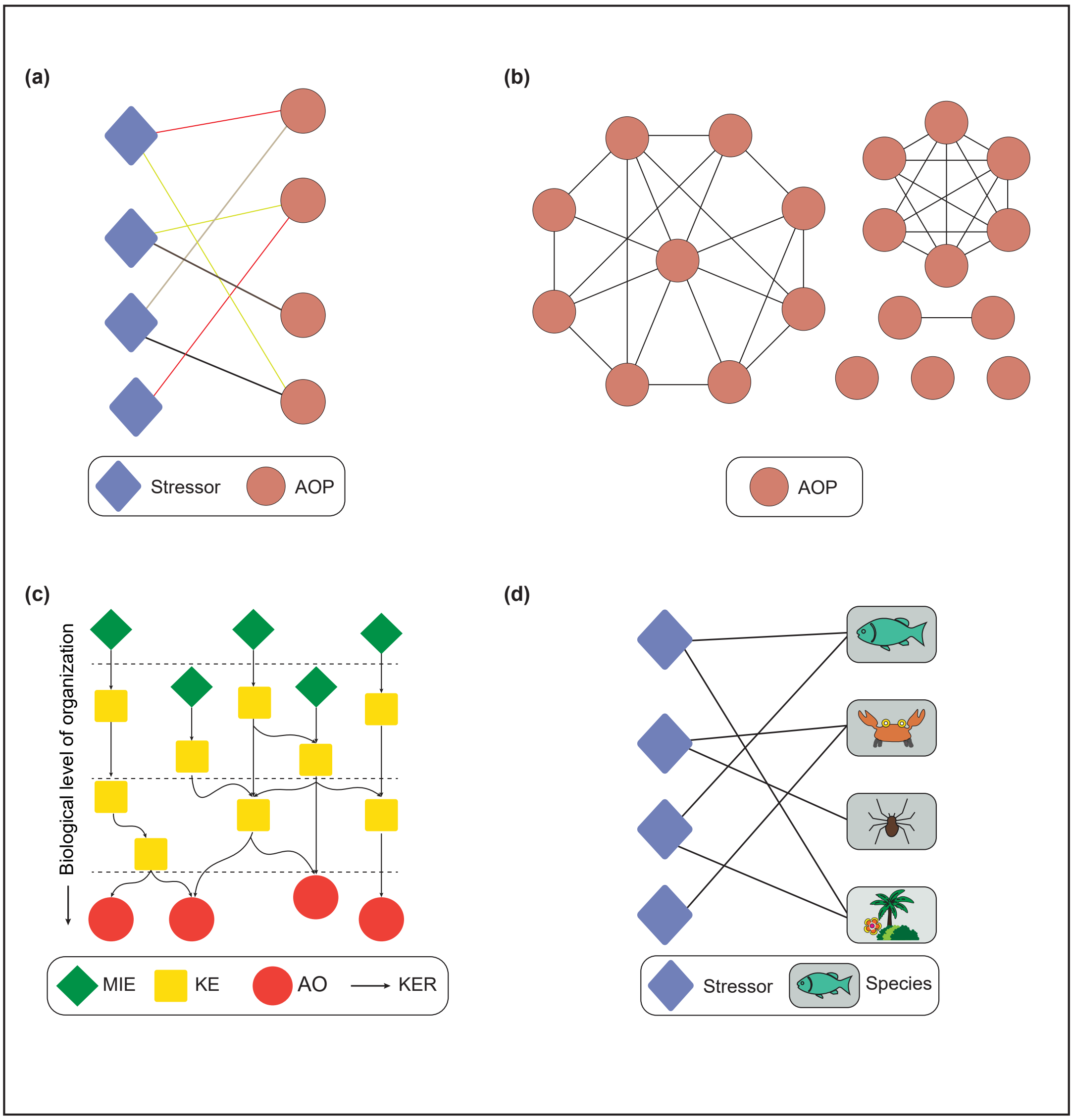


Figure 1.4: Schematic figure describing different networks constructed from stressor and AOP perspective, and analyzed in this thesis. (a) A stressor-AOP network linking stressors with their associated AOPs, where the stressor-AOP link is annotated based on coverage score and level of relevance. (b) An undirected AOP network of the AOPs associated with the stressor, where two AOPs are connected if they share at least one KE. (c) A directed AOP network comprising MIEs, KEs, and AOs, where the KEs including MIEs and AOs are linked through the KERs and are arranged according to their biological levels of organization. (d) A stressor-species network linking stressors and species.

directional flow of toxicological information from MIEs to AOs through the KEs (Figure 1.4c). Additionally, topological analysis of the directed AOP network can be performed using graph-theoretic frameworks to gain insights such as identifying points of convergence and critical biological events [58,59,77]. These insights can aid in the development of *in vitro* assays which can be tailored to capture upstream biological events including adverse outcomes, thereby reducing the need for animal based experiments [58]. Moreover, a directed AOP network can highlight emergent toxicity pathways within the context of systems biology that are not captured when considering individual AOPs alone [59]. Furthermore, incorporating existing chemical toxicity data into the directed AOP network can aid in the understanding of specific toxicity pathways through which chemicals induce adverse health effects across various species [64]. Notably, due to its comprehensive nature in capturing stressor associated complex toxicity pathways, the directed AOP network can be considered as a functional unit of toxicity prediction [54]. In this thesis, we have constructed directed AOP networks for several known chemical pollutants by considering the AOPs in the connected components of the corresponding undirected AOP network. Thereafter, we have analyzed these networks using network-based measures and investigated the associated toxicity pathways in both humans and ecological species using published experimental data.

Stressor-species network

A stressor-species network is a visualization that depicts the relationships between stressors (chemicals) and different species (Figure 1.4d). In general, a stressor-species link can be established using various sources of information, including toxicity data from stressor-specific experiments performed across different species or through AOP taxonomic applicability information. Previously, Wang *et al.* constructed a stressor-species network for polycyclic aromatic hydrocarbons (PAHs) using toxicity concentration data, highlighting the species and species groups most affected by PAH exposure [94]. In this thesis, we have constructed stressor-species networks using the toxicity concentration data and bioconcentration factor data for environmental chemicals documented in a published resource.

In Chapter 4 of this thesis, we have constructed both directed and undirected AOP networks to analyze cadmium-induced toxicity. In Chapter 5 of this thesis, we have expanded this analysis by constructing a stressor-AOP network for plastic additives which provided a holistic understanding of the adverse effects induced by these environmental chemicals. In Chapter 6, we have constructed a stressor-species network for petroleum hydrocarbons (PHs) and assessed their toxicities across various ecological species. Overall, by leveraging a variety of network representations, we have investigated the toxicities induced by different environmental chemicals and linked chemical exposome to its health effects in humans and ecosystems.

1.5 Thesis organization

The remaining chapters of this thesis are organized as follows:

Chapter 2 presents a systematic investigation of the chemical diversity along with the structure-activity landscape of AR binding chemicals. We cluster and visualize the AR binding chemical space, and assess its global diversity. Subsequently, we investigate the structure-activity landscape of AR binders using the SAS map [38] and identify chemicals forming activity cliffs. Additionally, we compute the SALI score [39] for all pairs of AR binding chemicals and use SALI heatmap to evaluate the activity cliffs identified using SAS map. Finally, we provide a classification of the activity cliffs into six categories using structural information of AR binding chemicals at different levels. **The work reported in this chapter is contained in the published manuscript [45].**

Chapter 3 presents a systematic investigation of the heterogeneity in structure-activity as well as structure-mechanism relationships among the TSHR binding chemicals from ToxCast [16]. By employing SAS map [38], we identify activity cliffs among chemicals in both TSHR agonist dataset and TSHR antagonist dataset. Further, by using the MMP approach, we find that the resultant activity cliffs (MMP-cliffs) [43] are a subset of activity cliffs identified via the SAS map approach. Subsequently, by leveraging the ToxCast MOA annotations for chemicals common to both TSHR agonist and TSHR antagonist datasets, we identify Strong MOA-cliffs and Weak MOA-cliffs. **The work reported in this chapter is contained in the published manuscript [46].**

Chapter 4 presents derivation and characterization of an AOP network for inorganic cadmium-induced toxicity through integration of heterogeneous data from different exposome-relevant resources. From AOP-Wiki [53], we filter high confidence AOPs and identify KEs associated with inorganic cadmium from five exposome-relevant databases using a data-centric approach. We then curate cadmium relevant AOPs (cadmium-AOPs) and construct an undirected AOP network, to identify the large connected component of cadmium-AOPs. Further, we analyze the directed network of KEs and KERs within the largest component using graph-theoretic approaches. Subsequently, we mine published literature using artificial intelligence-based tools to provide auxiliary evidence of cadmium association for all KEs in the largest component. Finally, we perform case studies to verify the rationality of cadmium-induced toxicity in humans and aquatic species. **The work reported in this chapter is contained in the published manuscript [95].**

Chapter 5 presents the development of a stressor-centric AOP network by leveraging integrative toxicogenomic approach to understand plastic additives-induced toxicites. We first identify a list of plastic additives from chemicals documented in plastics [20]. Next, we leverage heterogeneous toxicogenomics and biological endpoints data from five exposome-relevant databases and identify associations between plastic additives, and high quality and complete AOPs within AOP-Wiki. Based on these plastic additive-AOP associations, we construct a stressor-centric AOP network, wherein the stressors are categorized into priority use sectors and AOPs are linked to disease categories. We visualize the plastic additives-AOP network for each of the plastic additives and make them available in a dedicated website: https://cb.imsc.res.in/saopadditives/. Finally, we show the utility of the constructed plastic additives-AOP network by identifying highly relevant

AOPs associated with three known pollutants, and explore the associated toxicity pathways in humans and aquatic species. **The work reported in this chapter is contained in the published manuscript [96].**

Chapter 6 presents our network-based investigation of PHs-induced ecotoxicological effects and their risk assessment. First, we systematically curate a list of PHs from published reports [97, 98]. Next, we integrate biological endpoints data from different toxicological databases and construct a stressor-centric AOP network linking PHs with ecotoxicologically-relevant high confidence AOPs within AOP-Wiki. Further, we construct stressor-species networks based on reported toxicity concentrations and bioconcentration factors data, and analyze the effect of PHs across different ecological species. Finally, we utilize the aquatic toxicity data within ECOTOX [28] to construct Species Sensitivity Distributions (SSDs) [99] for polycyclic aromatic hydrocarbons (PAHs) prioritized by the United States Environmental Protection Agency (US EPA), and derive their corresponding hazard concentrations (HC05) that is not harmful to 95% of species in the aquatic ecosystem. **The work reported in this chapter is contained in the manuscript [100].**

In concluding **Chapter 7**, we present a brief summary and limitations of the research reported across different chapters of this thesis. The chapter also discusses the future prospects and the scope of our research in investigating environmental chemicals in the exposome, and linking them to human and ecosystem health.

Chapter 2

Identification of activity cliffs in structure-activity landscape of androgen receptor binding chemicals

Androgen receptor (AR) is a ligand-dependent nuclear transcription factor, mediated by male sex hormones dihydrotestosterone (DHT) and testosterone which are also known as androgens [101, 102]. The binding of these androgens with AR plays a key role in the development of male reproductive system and secondary sexual characteristics [101, 102]. Apart from the androgens, several endocrine disrupting chemicals (EDCs) have been reported to bind with AR and interfere with the normal functioning of the hormones [103–105]. The modes of action of the EDCs on the human endocrine system are manifold, and these include blocking the binding of the hormones to their native receptors [8, 15, 106]. These EDCs have a deleterious effect on human reproductive health which includes developmental abnormalities in the reproductive tract, poor semen quality and testicular cancer [104, 105].

Among the myriad environmental chemicals in the human exposome, it is therefore imperative to identify the EDCs which can bind with AR and interfere with the normal functioning of the male hormones. To this end, activity landscape analysis and identification of the activity cliffs in the space of AR binding chemicals will enable creation of better predictive models for EDCs.

In this chapter, we perform chemical space exploration, clustering and diversity analysis of the AR binding chemicals curated from published literature. Notably, we utilize the reported relative binding affinity (RBA) of these AR binders and systematically investigate their local and global structure-activity landscapes using various computational approaches to identify and characterize activity cliffs. **The work reported in this chapter is contained in the published manuscript [45].**

2.1 Methods

2.1.1 Chemical dataset curation and annotation

Previously, Fang *et al.* [107] had experimentally determined the AR binding affinity of 202 natural, synthetic and environmental chemicals against recombinant rat AR protein using competitive receptor binding assay. In particular, the Fang *et al.* dataset provides the Chemical Abstracts Service (CAS) identifiers, experimentally determined half maximal inhibitory concentration (IC_{50}), and the RBA for the 202 chemicals. Note that, the RBA value for a particular chemical was obtained by dividing the IC_{50} of the reference chemical R1881 by the IC_{50} of the particular chemical and expressing it as a percent. Moreover, Fang *et al.* classified the chemicals into 14 broad classes namely, steroids, diethylstilbestrols (DESs), phytoestrogens, phenols, flutamides, diphenylmethanes, polychlorinated biphenyls (PCBs), organochlorines, phthalates, aromatic hydrocarbons, noncyclic chemicals, aromatic acids, phenol like chemicals, and others.

Here, we leveraged the Fang *et al.* dataset to study the structure-activity landscape of AR binding chemicals. First, we removed chemicals which were labeled as ‘nonbinders’ and ‘slight binders’ in Fang *et al.* [107] since the binding affinity value for such chemicals was not reported in the publication. Afterwards, we collected the two-dimensional (2D) chemical structures for the remaining 146 AR binders from ChemIDplus [108]. The compiled chemical structures were processed using a cleaning protocol which included removing the invalid structures, duplicate structures and salts using MayaChemTools [109]. Further, we removed two acyclic chemicals from the dataset since the scaffold definition considered in this study is specific to chemicals with cyclic structure. In total, we curated a dataset of 144 natural, synthetic and environmental chemicals (Supplementary Table S2.1), along with their AR binding affinities and chemical structures starting from the information in Fang *et al.* [107]. Lastly, using the ClassyFire webserver [110], we structurally classified the 144 chemicals (Supplementary Table S2.1) in our dataset. Although the chemical class information was provided by Fang *et al.*, we decided to use here the comparatively more detailed chemical class information predicted from ClassyFire.

2.1.2 Chemical structure characterization

We characterized the structures of the 144 chemicals in our dataset using structural fingerprints, physicochemical properties and molecular scaffolds. First, we used Extended-Connectivity Fingerprints with diameter 4 (ECFP4) fingerprints implemented in RDKit [111, 112] to capture the structural features of the 144 chemicals. Thereafter, we used ECFP4 fingerprints to compute pairwise chemical structure similarity for the 144 chemicals. Second, we computed six physicochemical properties (PCP) namely, hydrogen bond donors (HBD), hydrogen bond acceptors (HBA), octanol/water partition coefficient (LogP), molecular weight (MW), topological polar surface area (TPSA) and number of rotatable bonds (RTB) which are known to be important for the bioavailability of the chemicals [113]. Third, we computed the molecular scaffolds for the 144 chemicals using the Bemis-Murcko definition (Supplementary Table S2.1) [114], since this definition is widely used and provides maximum information for the chemicals.

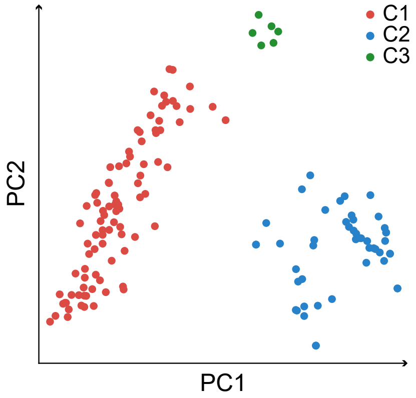


Figure 2.1: Principal Component Analysis (PCA) of the 144 AR binding chemicals. PCA was performed on the similarity matrix which encapsulates the pairwise structural similarities between all pairs of chemicals. The data points in the two-dimensional PCA plot (PC1; PC2) corresponds to the 144 chemicals in the whole library, and the data points are colored based on the three chemical clusters identified from the chemical similarity network (CSN).

2.1.3 Structure based clustering

We analyzed the global and local structure-activity relationship (SAR) for the 144 chemicals by clustering the chemicals based on structural similarity. To cluster the chemicals based on structural similarity, we first computed the pairwise chemical structure similarity for all pairs of chemicals in the library. The chemical structure similarity between any two chemicals was computed via Tanimoto coefficient (Tc) [115] using ECFP4 fingerprints. Thereafter, we constructed a similarity matrix for the whole library using the Tc values for all pairs of chemicals in our dataset. In order to visualize the high-dimensional dataset, we used principal component analysis (PCA) [116] to project the data to two dimensions. Notably, the first two principal components (PC1; PC2) capture 53.15% variance in the whole library. From the PCA plot (Figure 2.1), we find that the 144 chemicals in the whole library can be grouped into three clusters.

To further assist the clustering of chemicals in our dataset, we constructed a chemical similarity network (CSN) of the 144 chemicals. The nodes in the CSN represent the chemicals and the edges are drawn between two nodes if the Tc between the respective chemical pair is ≥ 0.2 (Figure 2.2). The similarity threshold of 0.2 was used since above this value, we observed isolated nodes in the CSN. Since our aim was to perform clus-

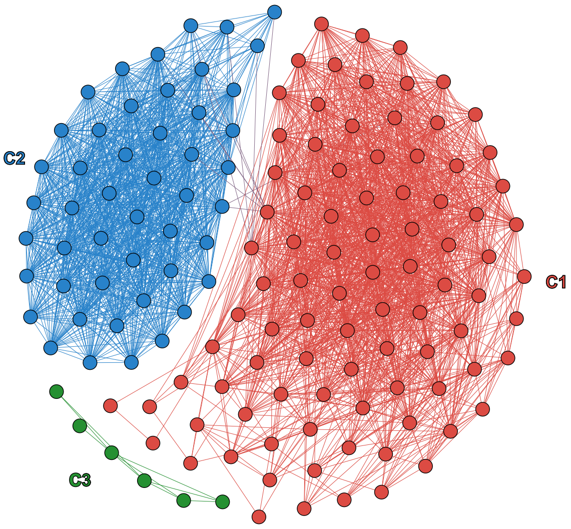


Figure 2.2: Chemical Similarity Network (CSN) of the 144 AR binding chemicals. The CSN was constructed based on the pairwise structural similarity computed via Tc using ECFP4 fingerprints. Edges were assigned between two nodes if Tc ≥ 0.2. The nodes and edges belonging to the chemical clusters C1, C2 and C3 are colored in red, blue and green, respectively. The edges between clusters C1 and C2 are colored in shades of magenta.

tering of the 144 AR binders, to avoid isolated nodes as separate clusters, we chose the similarity threshold to be 0.2. The CSN was visualized using Gephi software package version 0.9.7 [117]. Thereafter, we used the Louvain community detection (with resolution parameter set to 5.0) in Gephi package to identify clusters of chemicals within the CSN [118]. Lastly, we colored the data points in the PCA plot (Figure 2.1) based on the clusters identified within the CSN (Figure 2.2; Supplementary Table S2.1). We remark that the two approaches, PCA and CSN, were used to check the agreement between clustering of chemicals obtained from different approaches.

2.1.4 Global diversity

The Consensus Diversity Plot (CDP) [38, 113] helps in visualizing and comparing the global diversity of different chemical libraries. We used CDP to compare the diversity of the three chemical clusters and the whole library of 144 chemicals analyzed here (Figure 2.3). Following González-Medina *et al.* [113] we considered Molecular ACCess System (MACCS) keys fingerprints to analyze the structural diversity of the chemical clusters and the whole library. The x-axis of the CDP represents median Tc obtained using MACCS

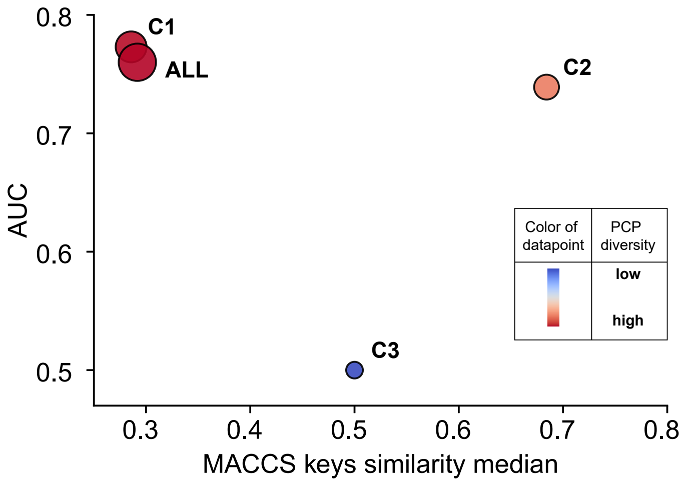


Figure 2.3: Consensus Diversity Plot (CDP) depicting the global diversity of the three chemical clusters (C1, C2 and C3) and the whole library (ALL). The x-axis of the CDP represents the median Tc obtained using MACCS keys fingerprints and the y-axis represents AUC. The PCP diversity computed from the mean Euclidean distance using the six physicochemical properties are represented by the color of the data points: red represents high PCP diversity whereas blue represents low PCP diversity. The relative size of the dataset is represented by the size of the data points.

keys fingerprints, capturing the structural diversity. The y-axis of the CDP represents the area under the curve (AUC) obtained from the cyclic system retrieval curve (CSR) [119, 120], capturing the scaffold diversity. The color of the data points in the CDP represents the PCP diversity of the clusters or the whole library. Note, red color indicates high PCP diversity and blue color indicates low PCP diversity [120, 121]. PCP diversity is the mean Euclidean distance between all pairs of chemicals in a cluster or the whole library computed using the six physicochemical properties. Finally, the size of a data point in the CDP represents the relative size of a cluster or the whole library.

2.1.5 Activity difference

For any pairs of chemicals in a cluster or the whole library, the absolute value of pairwise activity difference was calculated using the formula:

$$\Delta act_{i,j} = \left| log(RBA_i) - log(RBA_j) \right|$$

where, RBA_i and RBA_j are the relative binding affinities of the i^{th} and j^{th} chemicals, respectively as determined by Fang *et al.* [107].

2.1.6 Activity cliff identification

Activity landscape analysis helps in studying the nature of the SAR of a chemical space. In particular, 'activity cliffs' or discontinuous SAR identified by the analysis can help in capturing key pharmacophore regions necessary for biological activity. Activity landscape analysis can be done using various approaches. Here, we used structure-activity similarity (SAS) map [35, 38] and structure-activity landscape index (SALI) [31, 39] for the identification of activity cliffs.

Structure-activity similarity (SAS) map

Here, we generated SAS maps for the three clusters and the whole library of 144 chemicals (Figure 2.4). Briefly, the SAS map has 4 quadrants (I to IV). Quadrant I denotes scaffold hopping region (with chemicals having low structural similarity and low activity difference), quadrant II corresponds to smooth region of the SAR space (with chemicals having high structural similarity and low activity difference), quadrant III identifies activity cliffs (with chemicals having high structural similarity and high activity difference), and quadrant IV represents an uncertain region (with chemicals having low structural similarity and high activity difference). Importantly, quadrant III of the SAS map is the region of interest in this work. The x-axis of the SAS map represents the Tc obtained using ECFP4 fingerprints for different pairs of chemicals. Note that, we used ECFP4 fingerprints since it has been extensively used to analyze the activity landscape in earlier literature [37, 38]. The y-axis of the SAS map represents the absolute activity difference between different pairs of chemicals ($\Delta act_{i,j}$). The median plus two standard deviations for the distribution of Tc obtained using ECFP4 fingerprints was computed for all pairs of chemicals in the whole library, and thereafter, it was used as x-axis threshold to demarcate the quadrants in the SAS map. The x-axis threshold for the whole library of 144

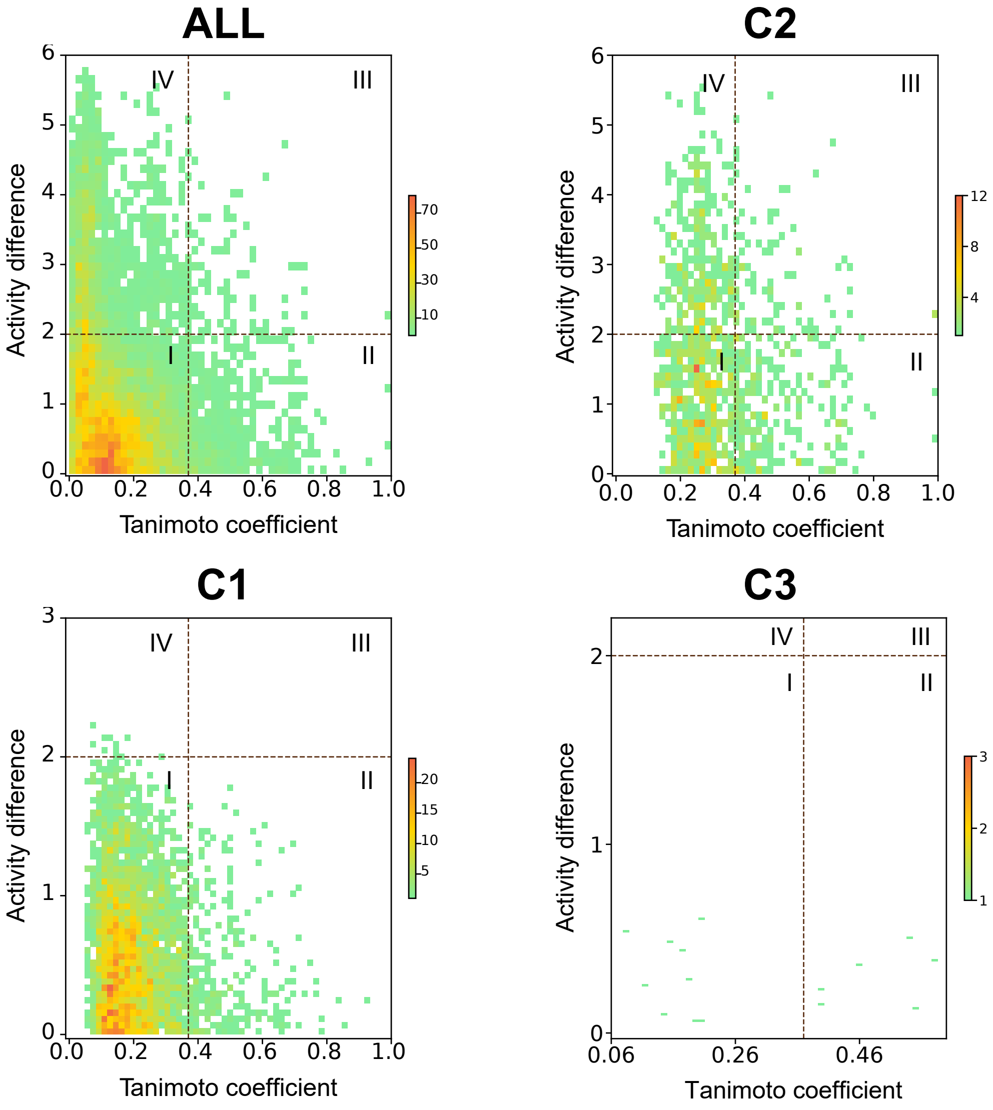


Figure 2.4: Structure-activity similarity (SAS) map for the whole library (ALL) and three chemical clusters (C1, C2 and C3). In each case, the SAS map is divided into 4 quadrants (denoted by I, II, III and IV) by considering x-axis threshold as 0.37 and y-axis threshold as 2 logarithmic units. Regions in each plot are colored based on the number of data points (green for the low dense region and red for the high dense region).

chemicals was determined to be 0.37. Note that the same x-axis threshold was used for both global and local SAS maps. Following Naveja *et al.*, the y-axis threshold was set at 2 logarithmic units in the activity difference or 100-fold change in the activity [38]. We remark that SAS maps are two-dimensional heatmaps wherein a continuous color gradient is used to represent the number of data points in a region.

Structure-activity landscape index (SALI)

SALI can be used to quantitatively characterize activity landscapes [39]. We computed the SALI scores for all pairs of chemicals in the whole library. SALI score is based on the activity difference and pairwise structural similarity, and is calculated as follows:

$$SALI_{i,j} = \frac{\left|\log(RBA_i) - \log(RBA_j)\right|}{1 - sim_{i,j}}$$

where RBA_i and RBA_j are the relative binding affinities of the i^{th} and j^{th} chemicals and $sim_{i,j}$ is the pairwise structural similarity between i^{th} and j^{th} chemicals. Note, when the $sim_{i,j}$ value for a pair of chemicals becomes 1, the SALI score becomes undefined. For such a pair of chemicals, the SALI score is assigned equal to that of the chemical pair with the maximum SALI score in the dataset. Tc using ECFP4 fingerprints was used for computing pairwise structural similarity of chemicals. Pairs of chemicals with high values of SALI correspond to activity cliffs and the computed scores were visualized using the SALI heatmap [39]. Importantly, we highlight the activity cliffs identified from the SAS map by marking them as black boxes in the SALI heatmap (Figure 2.5). Notably, this combined approach helps in visually identifying the overlap between the activity cliffs from SAS map and chemical pairs with high SALI scores. The plots, including heatmaps were generated via in-house python scripts using Matplotlib package [122].

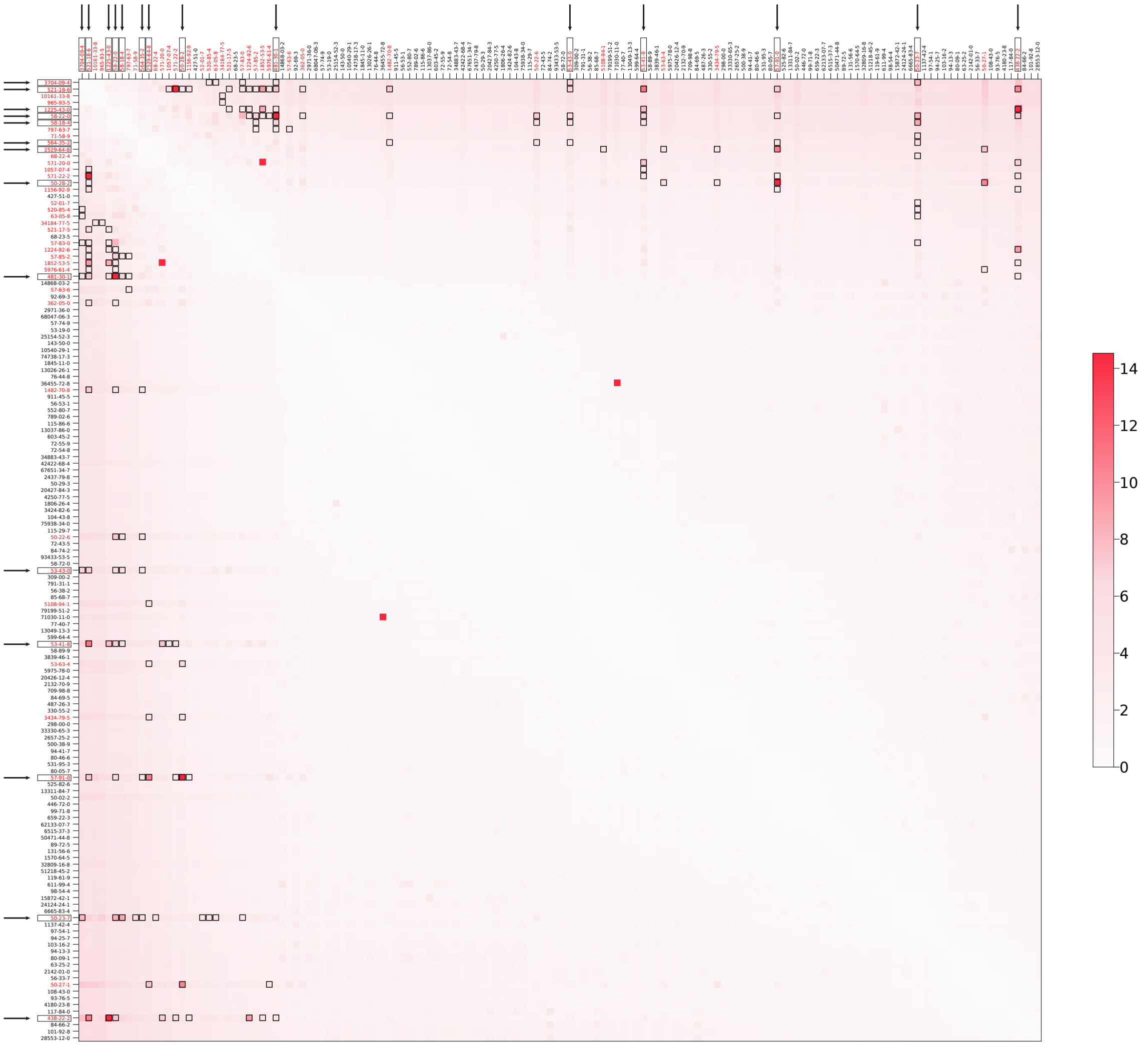


Figure 2.5: SALI score based heatmap for all chemical pairs of the 144 AR binding chemicals. The x-axis and y-axis of the heatmap are labeled with CAS identifiers of the 144 chemicals arranged from left to right and top to bottom in the descending order of their respective AR binding affinities. Each cell in the heatmap represents the chemical pair and is colored based on the computed SALI score: darker the cell, higher the SALI score. The chemical pairs (cells in heatmap) identified as activity cliffs in the SAS map are highlighted with black boxes. The CAS identifiers of the 41 chemicals which makeup the 86 activity cliffs in the SAS map are shown in red color. Further, the labels of the 14 ACGs identified in the SAS map are boxed and marked with arrows.

2.1.7 Computational tools

We used several computational tools to analyze the AR binding chemicals. MayaChemTools is a collection of several scripts which is useful in analyzing chemical data, computing physicochemical properties and generating fingerprints [109]. Using MayaChemTools, we followed a cleaning protocol to remove salts and duplicates. ClassyFire [110], a web-based tool, was used to predict the chemical classification considering the 2D structure of the chemicals. We used RDKit [123], an open source cheminformatics library to compute ECFP4 fingerprints, physicochemical properties and Bemis-Murcko scaffold of chemicals. Further, we used RDKit to perform the R-group decomposition of the chemicals forming activity cliffs. Gephi is an open source software for visualization and analysis of networks and graphs [117]. Using Gephi, we visualized the CSN and clustered the chemicals. All the plots were made using the Matplotlib package [122] in Python3.

2.2 Results

2.2.1 Visualization of the chemical space of AR binders

Figure 2.1 displays the 2D PCA plot generated from the similarity matrix for the whole library of 144 AR binding chemicals. In the PCA plot, we observed three chemical clusters, spatially separated in the two-dimensional space obtained using the first two principal components (PC1; PC2) (Figure 2.1). Further, we built the CSN of the 144 AR binding chemicals. Applying Louvain community detection method on the CSN, we again identified three chemical clusters (Supplementary Table S2.1) which are more likely to contain structurally similar chemicals (Figure 2.2) [110]. We annotated the data points in the PCA plot using three different colors corresponding to the three chemical clusters identified in the CSN (Figures 2.1 and 2.2). From Figure 2.1, we can see that the cluster information obtained from the CSN concurs with the clustering obtained from the PCA plot. The chemical cluster C1 contains 90 chemicals (62.5% of the whole library), C2 contains 48 chemicals (33.33% of the whole library) and C3 contains 6 chemicals (4.17% of the whole library).

2.2.2 Exploration of the chemical space of AR binders

We annotated the chemicals in the three clusters using the chemical class information predicted from ClassyFire [110] (Supplementary Table S2.1), and their occurrence in the list of EDCs compiled in DEDuCT database [8, 15] and publicly available lists on chemical regulation. Firstly, we find that the 144 chemicals in the whole library belong to 21 chemical classes. The chemicals in cluster C1 belong to a diverse set of chemical classes. 'Benzene and substituted derivatives' (47 of 90 chemicals) is the most prevalent chemical class in C1. The chemicals in cluster C2 are dominated by the chemical class 'Steroids and steroid derivatives' (44 of 48 chemicals). The chemicals in cluster C3 belong to diverse chemical classes. Importantly, we observed that the chemical classes of each cluster are unique, i.e., there is no overlap among the clusters in terms of the occurrence of chemical classes (Supplementary Table S2.1).

Secondly, we find 62 chemicals in the whole library are EDCs with documented adverse health effects based on a comparative analysis with the list of 792 potential EDCs compiled in DEDuCT [15]. The distribution of these 62 EDCs among the chemical clusters C1, C2 and C3 was found to be 46, 11 and 5 chemicals, respectively. Thirdly, to comprehend the regulatory status of the AR binding chemicals in the whole library, in view of their potential adverse health effects on humans, we considered six publicly available lists on chemical regulations. We find that, among the 144 AR binding chemicals, 31 are present in 'California Proposition 65 (CP65)' [124] list, 7 are present in 'Restricted substances under REACH' [125] list, 8 are present in 'SVHC under REACH' [126] list, 2 are present in 'Toxic chemicals restricted to be imported or exported in China' [127] list, 18 are present in 'EU list of substances prohibited in cosmetic products' [128] list, and 10 are present in 'Schedule 1 hazardous chemicals list in India' [129] list. In total, 42 of the 144 AR binding chemicals are present in at least one of the six chemical regulations considered here. Further, of the 62 chemicals in the whole library identified as potential EDCs, 29 are present in at least one of the six chemical regulations considered here. We remark that this dataset of 144 AR binding chemicals containing potential EDCs and regulated chemicals of concern can serve as a benchmark dataset to study the local and global SAR of the AR binding chemicals.

2.2.3 Scaffold content of the AR binding chemicals

In order to explore the scaffold content, we computed the Bemis-Murcko scaffolds for each of the 144 AR binding chemicals. Figure 2.6 shows the top 3 scaffolds in terms of the frequency of its occurrence in each chemical cluster. The number of unique molecular scaffolds in the chemical clusters C1, C2 and C3 are 29, 21 and 6, respectively. The scaffold content of cluster C1 is dominated by the benzene scaffold (i.e., present in 29 chemicals). The scaffold content of cluster C2 is dominated by scaffolds belonging to the steroid class. In cluster C3, all the 6 chemicals have a unique scaffold. Of note, the scaffold content of each cluster is unique, i.e., there is no overlap among the clusters in terms of the scaffolds of the chemicals (Supplementary Table S2.1). The presence of non-overlapping molecular scaffolds and chemical classes among the three clusters further justifies the clustering of the chemical space of the 144 AR binders into three clusters. Therefore, we used these chemical clusters and the whole library to study the local and global SAR of the AR binding chemicals.

2.2.4 Consensus Diversity Plot (CDP) and global diversity of AR binding chemicals

Figure 2.3 shows the CDP which helps in analyzing the global diversity of the chemical clusters and the whole library of AR binders. The data points in CDP namely, C1, C2, C3 and ALL correspond to the three chemical clusters and the whole library, respectively. The data points in the left of the CDP have high structural diversity, those in the bottom

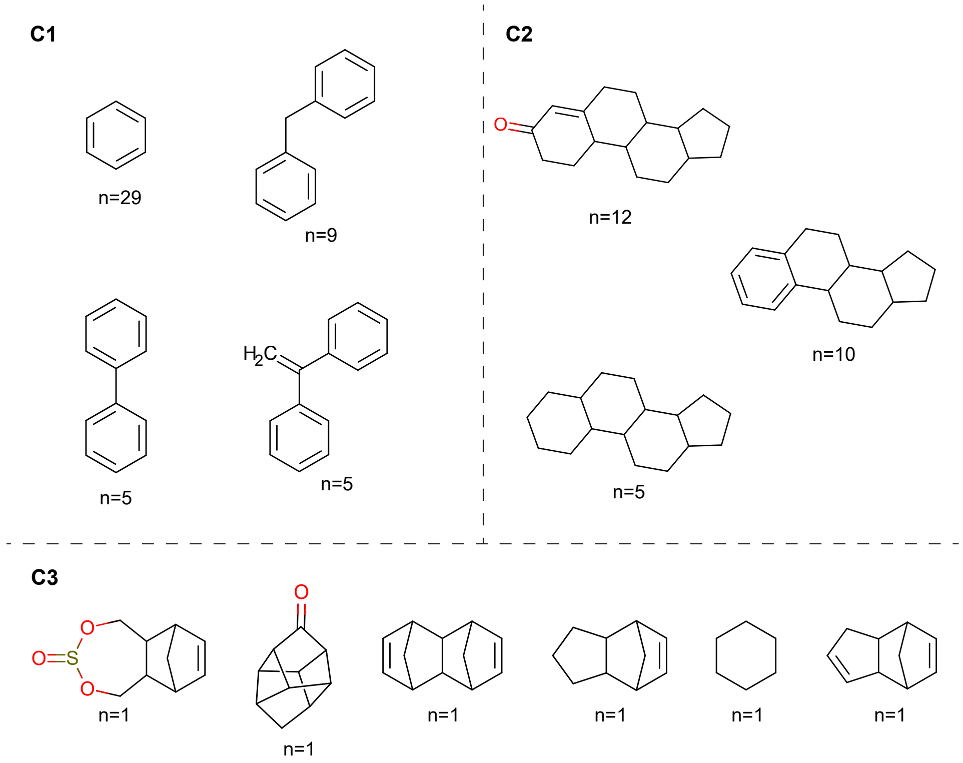


Figure 2.6: Top three Bemis-Murcko scaffolds for the three chemical clusters of AR binding chemicals in terms of the frequency (n) of their occurrence in each chemical cluster.

of the CDP have high scaffold diversity, and data points colored in red have high PCP diversity. From the CDP, we observed that the chemicals in cluster C1 have high structural diversity based on the median of Tc obtained using the MACCS keys fingerprints and low scaffold diversity based on AUC (Figure 2.3). We can see from Figure 2.3 that the whole library (ALL) has high structural diversity and low scaffold diversity. The chemical cluster C1 also has similar global diversity to the ALL, possibly since C1 accounts for 62.5% of the chemicals in the whole library (Table 2.1). The low structural diversity of C2 suggests that the chemicals in C2 are structurally very similar. The chemical cluster C2 has AUC similar to C1 and ALL, suggesting that C1, C2 and ALL have similar scaffold diversity (Figure 2.3; Table 2.1). The chemical cluster C3 has the highest scaffold diversity with AUC value of 0.5. This is because all the six chemicals in C3 have unique molecular scaffolds. While assessing PCP diversity using CDP, values close to 1 reflect high diversity in a chemical library, whereas values close to 0 reflect low diversity. We observed

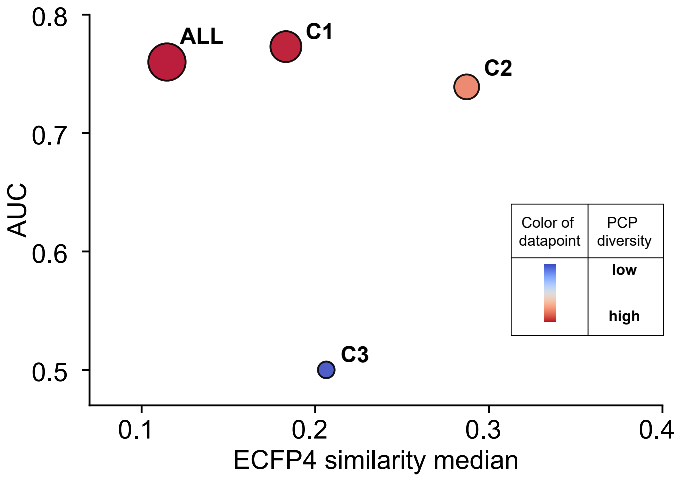


Figure 2.7: Consensus Diversity Plot (CDP) depicting the global diversity of the three chemical clusters (C1, C2 and C3) and the whole library (ALL). The x-axis of the CDP represents the median Tc obtained using ECFP4 fingerprints and the y-axis represents AUC. The PCP diversity computed from the mean Euclidean distance using the six physicochemical properties are represented by the color of the data points: red represents high PCP diversity whereas blue represents low PCP diversity. The relative size of the dataset is represented by the size of the data points.

that among the clusters considered here, C1 has highest PCP diversity which means that physicochemical properties of the chemicals in C1 are very different, and C3 has the lowest PCP diversity which means that physicochemical properties of the chemicals in C3 are quite similar (Figure 2.3; Table 2.1). In sum, C1 has a diverse set of chemicals and C3 has similar chemicals in terms of physicochemical properties. A similar CDP made using ECFP4 fingerprints is shown in Figure 2.7. We observed that CDP using MACCS keys fingerprints is better suited for analyzing the structural diversity of datasets (Figures 2.3 and 2.7).

2.2.5 Activity landscape analysis to explore the SAR of the AR binding chemicals

Activity landscape analysis is widely used to analyze the SAR of diverse chemical spaces [40]. SAS map is the first method to be developed for 2D visual representation of activity landscape [35]. SAS map takes into account the structural similarity and activity difference of all possible pairs of chemicals in a given dataset. A SAS map is a 2D heatmap with continuous color gradient reflecting the number of data points in a region. In particular, the color gradient ranges from green for the low density regions to red for high density regions (Figure 2.4). Figure 2.4 shows the SAS maps for the whole library (ALL), and the three chemical clusters (C1, C2 and C3). For the whole library (ALL), the majority of the chemical pairs are in region I (Figure 2.4; Table 2.2). This suggests that the corresponding chemical pairs are structurally diverse with low activity difference. Notably, we find only 86 pairs of chemicals in region III. Region III contains pairs of chemicals with high structural similarity and high activity difference, i.e., the activity cliffs. In particular, the 86 activity cliff pairs (Supplementary Table S2.2) in region III are formed by 41 chemicals.

Upon analyzing the local SAS maps for the three clusters, we found that all of these 86 activity cliff pairs belong to cluster C2. Table 2.2 provides a quantitative summary of the percentage of chemical pairs in the different regions of the SAS map for the three chemical clusters and the whole library. As noted above, the cluster C2 is the only cluster with activity cliffs wherein 7.62% of the chemical pairs are in activity cliff region. In other words, the two other clusters C1 and C3 do not have activity cliff pairs. The local activity landscape for cluster C2 is rough and heterogeneous since C2 has chemical pairs in both smooth and activity cliff regions. We find that the 41 AR binding chemicals that form activity cliff pairs in cluster C2 belong to the chemical class ‘Steroids and steroid derivatives’.

Further, we identified the activity cliff generators (ACGs) among the 86 activity cliff pairs using the criterion proposed by Naveja *et al.* [38]. We considered a chemical to be an ACG if it is part of at least 5 activity cliff pairs. Following this criterion, we identified 14 chemicals as ACGs (Supplementary Table S2.3). Notably, we find that the two principal androgens, DHT and testosterone form the maximum number of activity cliff pairs (i.e., 17 and 13, respectively). In a later subsection, we provide a detailed classification of the identified activity cliff pairs.

2.2.6 SALI based exploration of the AR binding chemicals

To date, multiple numerical approaches have been proposed to quantify SAR discontinuity or activity cliffs [32]. SALI introduced by Guha *et al.* is another widely used method for identifying and quantifying the activity cliffs in a SAR landscape [39]. Herein, we followed Guha *et al.* to compute the SALI scores for all pairs between the 144 AR binding chemicals. The computed SALI scores for all chemical pairs can be visualized using the SALI heatmap [39]. Figure 2.5 shows the SALI heatmap for all pairs between the 144 AR binding chemicals. Briefly, both x- and y-axis of this heatmap are annotated with the CAS identifiers of the AR binders, and the chemicals are arranged from left to right and top to bottom according to descending order of their AR binding affinities (Figure 2.5). Further, each cell in the SALI heatmap represents a chemical pair and the cells are colored based on the SALI score of the corresponding chemical pair (darker the cell color, higher the SALI score). Certain cells of the SALI heatmap are marked with black boxes, if the corresponding chemical pairs were identified as activity cliffs from the analysis of the SAS map (Figure 2.4; Supplementary Table S2.2). The CAS identifiers (axes labels) corresponding to the 41 chemicals which makeup the 86 activity cliff pairs (Supplementary Table S2.2) identified from the SAS map are colored in red. Also, the CAS identifiers of the 14 ACGs (Supplementary Table S2.3) identified from the SAS map are boxed and marked with arrows (Figure 2.5).

From Figure 2.5, we can see that most of the chemical pairs identified as activity cliffs in the SAS map, also have high SALI scores. Further, we noted that some chemical pairs having high SALI score (dark colored cell) were not identified as activity cliffs in the SAS map (and thus, not highlighted with black boxes in the SALI heatmap). We find such pairs to be stereoisomers with low activity difference. We note that the chemical fingerprints used to compute structural similarity does not capture the stereoisomer information, and this renders the structural dissimilarity (i.e., denominator in the SALI score definition) to be zero. Due to this, in spite of the low activity difference, the stereoisomer pairs were assigned the highest SALI score. However, in the case of SAS map, such stereoisomer pairs are located in region II (as they have high structural similarity and low activity difference), and hence are not identified as activity cliffs. Lastly, we observed that the SALI heatmap cells with intermediate SALI score have not been identified as activity cliffs in the SAS map approach, possibly due to the x- and y-axis thresholds used to demarcate the four regions in the SAS maps (Figure 2.5). In sum, there is a large overlap between the activity cliff pairs identified from the SAS map and the chemical pairs with high SALI score. These observations further support the 86 activity cliff pairs identified in this study.

2.2.7 Structural classification of activity cliffs

Previously, Hu *et al.* have proposed a conceptually different methodology from SAS map or SALI score to identify the activity cliffs [44]. The proposed methodology includes the structural classification of chemical pairs with high activity difference, using the molecular scaffolds, R-groups, and topology of the R-groups in the chemical structures. Here, we followed Hu *et al.* to systematically classify the 86 activity cliff pairs identified using the SAS map.

Figure 2.8 shows the computational workflow used to structurally classify the activity cliff pairs by considering the structural information of chemicals at different levels. Here we classified the activity cliff pairs into 6 structural categories namely:

(i) Chirality cliff: An activity cliff pair having same molecular scaffold, R-groups and topology of R-groups.

(ii) Topology cliff: An activity cliff pair having same molecular scaffold and R-groups but different topology of R-groups.

(iii) R-group cliff: An activity cliff pair having same molecular scaffold but different R-groups.

(iv) Scaffold cliff: An activity cliff pair having different molecular scaffolds, but same

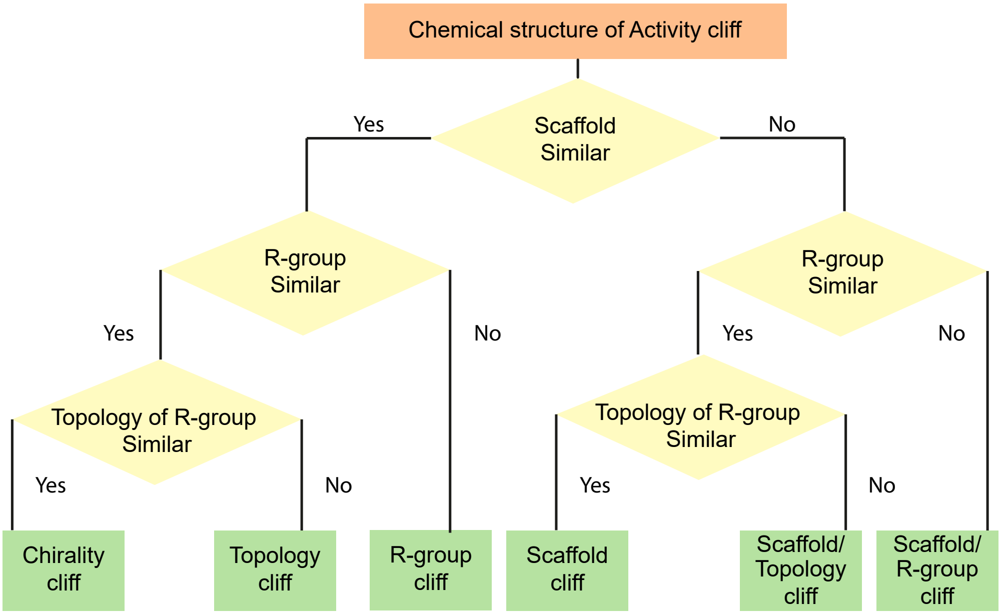


Figure 2.8: Computational workflow to structurally classify the activity cliff pairs using the structural information of the chemicals at different levels.

R-groups and topology of the R-groups.

(v) Scaffold/Topology cliff: An activity cliff pair having different molecular scaffolds and topology of the R-groups, but same R-groups.

(vi) Scaffold/R-group cliff: An activity cliff pair having different molecular scaffolds and R-groups.

Note that the 'Scaffold/R-group' cliff category listed above was not considered by Hu *et al.* while defining the types of activity cliffs [44].

To classify the activity cliffs into six different structural categories, we used the R-group decomposition function in RDKit [123], to decompose the chemical structure into its core structure (scaffold) and R-groups (Supplementary Table S2.4). Then using the workflow as depicted in Figure 2.8, we manually classified the 86 activity cliffs into 6 different structural categories (Supplementary Table S2.5). We find that among the 86 activity cliffs, 3 are Chirality cliffs, 3 are Topology cliffs, 29 are R-group cliffs, 2 are Scaffold cliffs, 9 are Scaffold/Topology cliffs, and 40 are Scaffold/R-group cliffs. In the next subsection, we illustrate with examples the six categories of activity cliffs by considering two ACGs and their activity cliff pairs.

2.2.8 Activity cliff classification of ACGs: DHT and 5α-Androstan-17β-ol

Here, we show the structural classification of the activity cliff pairs for two ACGs namely, DHT (CAS:521-18-6) and 5α-Androstan-17β-ol (CAS:1225-43-0). Figure 2.9 shows the activity cliff classification of the activity cliff pairs of the two ACGs. Further, Figure 2.9 also provides the CAS identifier and the logarithm of RBA value, i.e., $log(RBA)$ of the chemicals which makeup the above activity cliff pairs. DHT is one of the principal androgens with high binding affinity to AR. DHT forms activity cliffs with 17 other chemicals, which is the maximum number of activity cliff pairs formed by any chemical in our dataset. All the 17 chemicals which form activity cliffs with DHT have low binding affinity compared to DHT. Using the computational workflow as shown in Figure 2.8, we were able to classify the 17 activity cliff pairs of DHT into 4 categories namely, Chirality cliff, Scaffold cliff, Scaffold/Topology cliff and Scaffold/R-group cliff (Figure 2.9a). Figure 2.9b shows the cliff classification of another ACG namely, 5α-Androstan-17β-ol forming activity cliffs with 7 other chemicals in our dataset. The 7 activity cliffs can be structurally classified into Topology cliff, R-group cliff, Scaffold cliff, Scaffold/Topology cliff and Scaffold/R-group cliff. Below, we explain the structural classification of a few activity cliff pairs for the two ACGs.

DHT and Etiocholan-17β-ol-3-one (CAS:571-22-2) are stereoisomers and form Chirality cliff, with activity difference (i.e., difference in $log(RBA)$) of 2.15 (Figure 2.9a). 5α-Androstan-17β-ol and 5α-Androstan-3β-ol (CAS:1224-92-6) form Topology cliff since both the chemicals have the same scaffold and R-groups (-CH3 and -OH), but only differ in the topology of the R-group (-OH), with activity difference of 2.19 (Figure 2.9b). 5α-Androstan-17β-ol forms R-group cliffs with two chemicals as these two chemicals when compared to 5α-Androstan-17β-ol have different R-groups but the same scaffolds (Figure 2.9b). 5α-Androstan-17β-ol forms Scaffold cliff with epitestosterone since both the chemicals differ in molecular scaffold but have the same R groups (-CH3, -OH) and

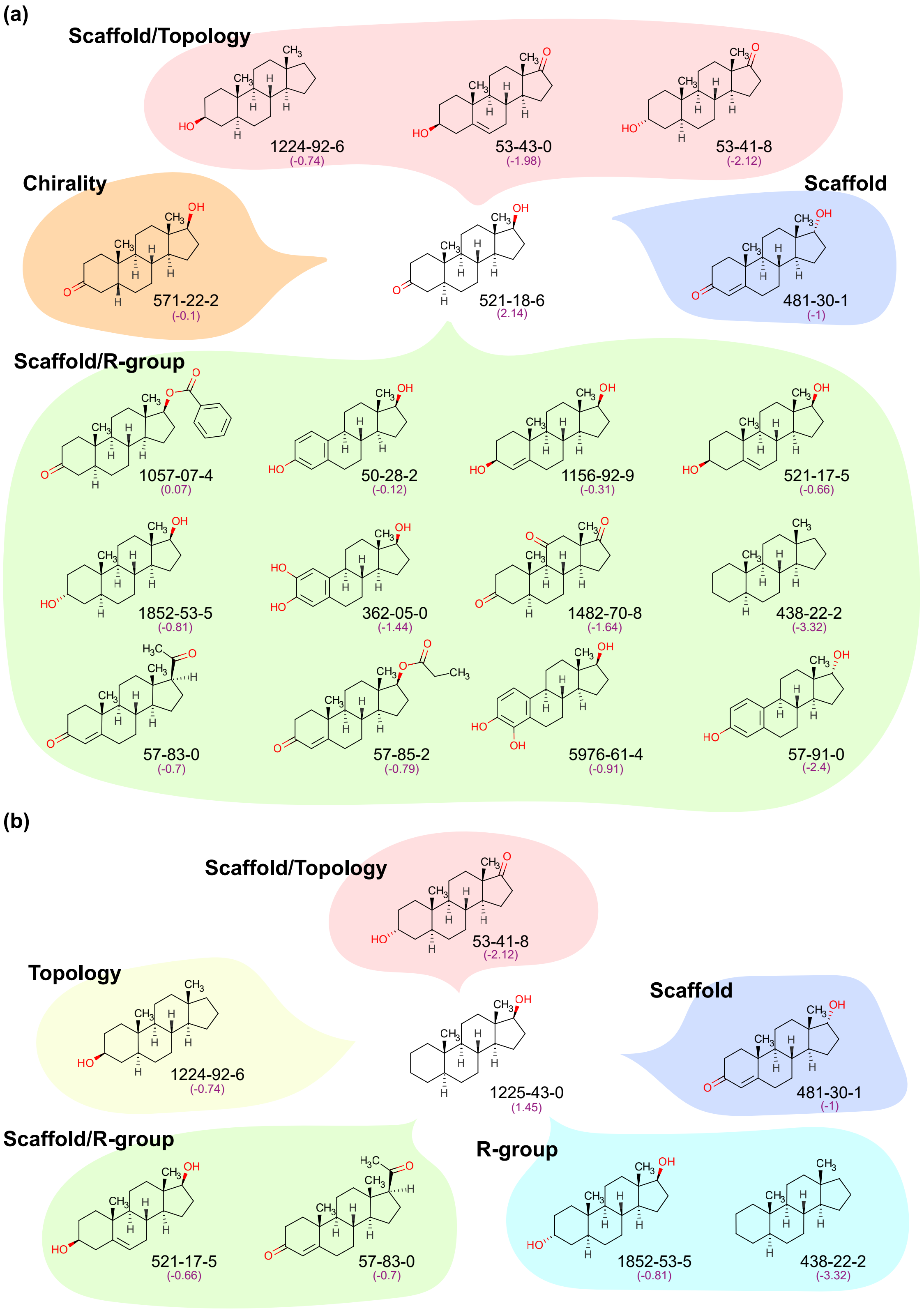


Figure 2.9: Structural classification of the activity cliffs. (a) Structural classification of the 17 activity cliff pairs of the ACG DHT. (b) Structural classification of the 7 activity cliffs of the ACG 5α-Androstan-17β-ol. For each chemical, CAS identifier and $log(RBA)$ value have been provided. Here, each activity cliff category is highlighted in a different color.

topology of R-groups, with activity difference of 2.45 (Figure 2.9b). DHT forms Scaffold/Topology cliffs with three other chemicals which share the same set of R-groups (-CH3, -OH) with DHT, but have different molecular scaffolds and topology of R-groups (Figure 2.9a). DHT forms 12 Scaffold/R-group cliffs since these chemicals, when compared to DHT, differ in both molecular scaffolds and R-groups (Figure 2.9a). Note that we are unable to provide mechanistic interpretation of the structural features of activity cliffs identified from the classification since the co-crystallized structure of AR with the chemicals forming activity cliffs is not available.

2.3 Discussion

Structure-activity relationship (SAR) based analysis can help in predicting potential AR binding chemicals. Activity landscape analysis is a powerful tool for systematically analyzing the relationships between the chemical structures and their biological activities, thereby providing crucial insights for the development of highly predictive quantitative SAR (QSAR) models. In this chapter, we used several computational techniques to perform activity landscape analysis of the AR binders (Figure 2.10). Our efforts led to the identification of activity cliffs and the key structural features behind the formation of such activity cliffs. To the best of our knowledge, this study is the first attempt to computationally analyze the structure-activity landscape of the AR binding chemicals.

To date, activity landscape analysis has been extensively performed on chemical datasets relevant to drug discovery research. However, such attempts on environmental chemical spaces are limited in the literature. A notable exception is the work of Naveja *et al.* [38] wherein the SAS map approach was used to computationally detect activity cliffs and ACGs among environmental chemicals reported to bind the estrogen receptor. However, there were no similar attempt to analyze the activity landscape of AR binding chemicals prior to this work. Here, we employed not only SAS map but also SALI scoring based numerical method and observed that both methods are complementary in identify-

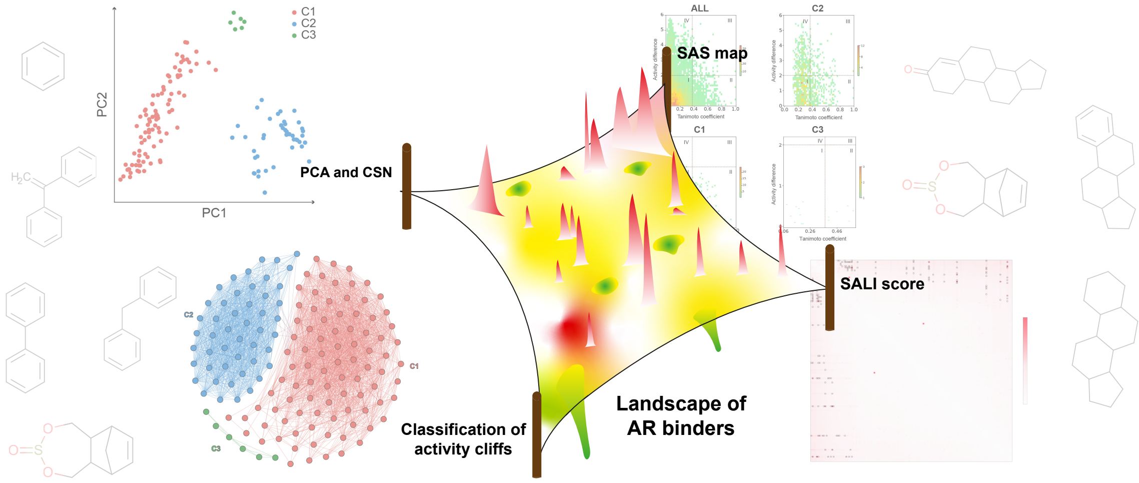


Figure 2.10: Schematic summary of the different computational approaches used to analyze the structure-activity landscape of AR binding chemicals.

ing the heterogeneity in the structure-activity landscape of AR binders. Importantly, the present study stands apart from the previously published studies by Fang *et al.* [107] and Naveja *et al.* [38] in analyzing the structure-activity landscape using different approaches in tandem and the structural interpretation of activity cliffs.

However, we were unable to provide a mechanistic interpretation of the formation of activity cliffs as no experimentally determined co-crystallized protein structures for any pair of chemicals forming activity cliffs were available in the Protein Data Bank (PDB) [130]. Instead, we provided a detailed classification of the activity cliffs using their structural features and this enabled better interpretation of the identified activity cliffs from a chemistry perspective [44]. We expect that the insights from this study will facilitate ongoing efforts in computational toxicology to build predictive models for identifying endocrine or hormone disruptors in the ever-expanding chemical exposome.

Supplementary Information

Supplementary Tables S2.1-S2.5 associated with this chapter are available for download from the GitHub repository: https://github.com/asamallab/PhDThesis-Ajaya_Kumar_Sahoo/blob/main/SI/ST_Chapter2.xlsx.

Code Availability

The computer programs used to perform the computations reported in this chapter are available in the following GitHub repository:

https://github.com/asamallab/PhDThesis-Ajaya_Kumar_Sahoo/tree/main/Codes.

Chemical dataset	Number of chemicals	Median Tc using MACCS keys fingerprints	Number of scaffolds	AUC	PCP diversity
ALL	144	0.29	56	0.76	3.14
C1	90	0.29	29	0.77	3.13
C2	48	0.68	21	0.74	2.95
C3	6	0.5	6	0.5	2.13

Table 2.1: Global diversity of the three chemical clusters (C1, C2 and C3) and the whole library (ALL). For each set, the table lists the number of chemicals present, median Tc using MACCS keys fingerprints, number of scaffolds, AUC and the PCP diversity.

Chemical cluster	Number of chemical pairs	Median Tc using ECFP4 fingerprints	Hops (Region I)	Smooth (Region II)	Cliffs (Region III)	Uncertain (Region IV)	Cliffs/Smooth
ALL	10296	0.11	74.38%	5.12%	0.84%	19.67%	0.16
C1	4005	0.18	92.41%	7.37%	0	0.22%	0
C2	1128	0.29	45.57%	20.04%	7.62%	26.77%	0.38
C3	15	0.21	60%	40%	0	0	0

Table 2.2: The table summarises the percentage of chemical pairs in the four regions or quadrants of the SAS maps for the whole library (ALL) and the three chemical clusters (C1, C2 and C3). The table also provides the ratio of the chemical pairs present in the activity cliff region (quadrant III) to the chemical pairs present in the smooth region (quadrant II). Further, the table provides the number of chemical pairs and median Tc computed using the ECFP4 fingerprints for the whole library and the three chemical clusters.

Chapter 3

Analysis of structure-activity and structure-mechanism relationships among thyroid stimulating hormone receptor binding chemicals by leveraging the ToxCast library

Thyroid stimulating hormone receptor (TSHR) plays an important role in the hypothalamic-pituitary-thyroid axis where it mediates the production of thyroid hormone upon activation by the physiologic agonist, thyroid stimulating hormone (TSH) [131–133]. The hypothalamic-pituitary-thyroid axis is crucial for development and metabolism, and is prone to disruptions by endocrine disrupting chemicals (EDCs) [134–136] in the human exposome. EDCs can bind to an endocrine receptor and dysregulate the hormonal activity in the human body, thus affecting the metabolism, immune system and reproductive system [137]. In particular, animal studies have shown that EDCs binding to TSHR disrupt the thyroid system, ultimately leading to developmental toxicity [138–140]. In human, the overproduction of thyroid hormone caused by the binding of M22 autoantibody with

TSHR can lead to Grave's disease [141], and underproduction of thyroid hormone caused by the binding of K1-70 autoantibody can lead to hypothyroidism and Hashimoto's disease [142].

Therefore, screening of environmental chemicals in the human exposome that can bind to TSHR is important for their proper management. Towards this, several machine learning based quantitative structure-activity relationship (QSAR) models have been developed [143, 144]. However, the predictions of these models can be inaccurate due the presence of heterogeneity in the structure-activity and structure-mechanism relationships of the chemicals binding to TSHR.

In this chapter, we describe our investigation of the structure-activity landscape and structure-mechanism relationships of chemicals binding to TSHR compiled from the ToxCast [16] chemical library. Previously in Chapter 2, we described our work analyzing the heterogeneity in the structure-activity landscape of androgen receptor (AR) binding chemicals using multiple computational approaches. Here, in addition to the activity landscape analysis, we leveraged the mechanism of action (MOA) annotations of the chemicals in TSHR agonist and TSHR antagonist datasets to analyze their structure-mechanism relationships. **The work reported in this chapter is contained in the published manuscript [46].**

3.1 Methods

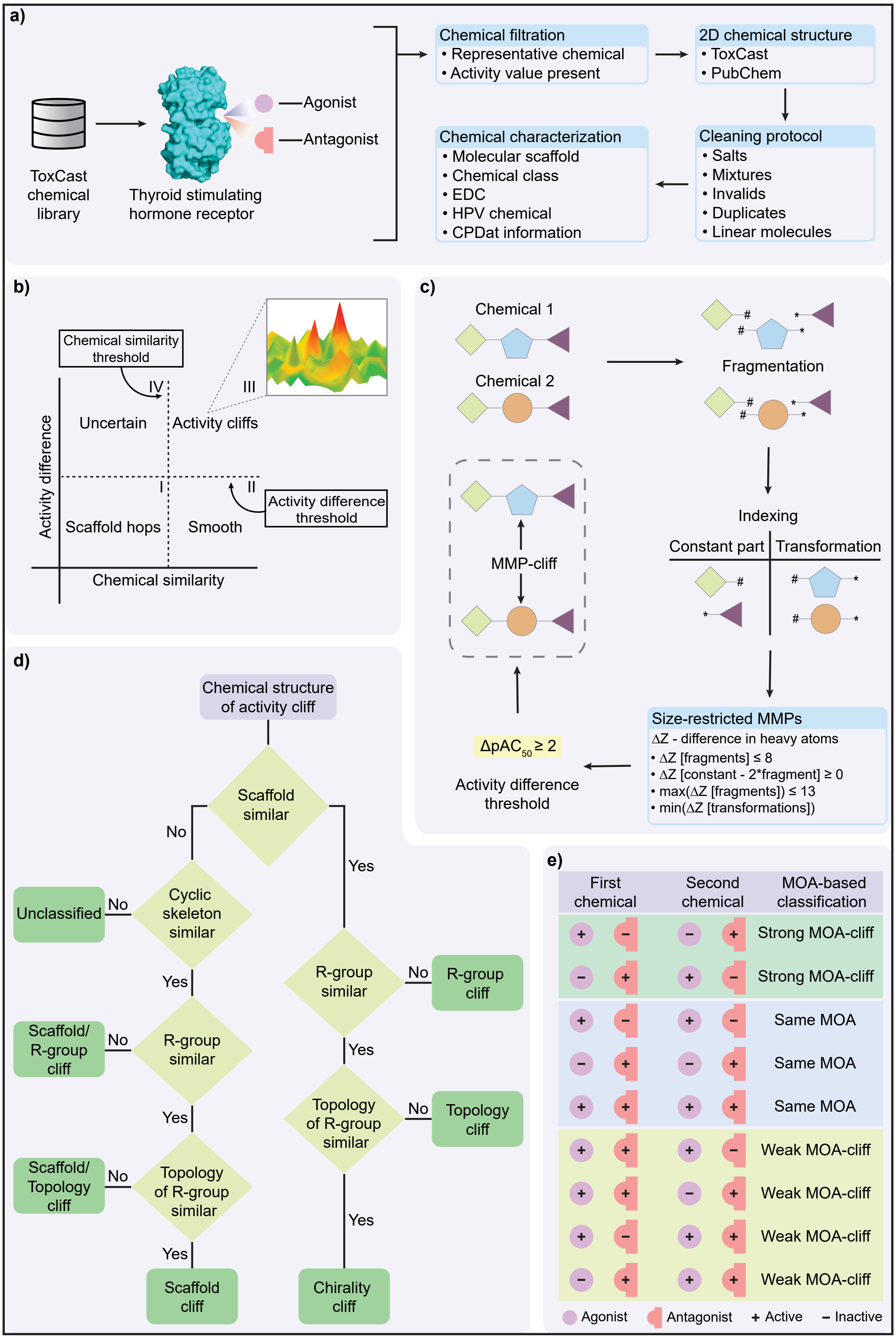
3.1.1 Chemical dataset comprising agonists and antagonists of the TSHR

The objective of this investigation is the analysis of the structure-activity landscape of the agonists and antagonists of the TSHR (Figure 3.1). For this investigation, we retrieved the chemicals, their corresponding activity values, and endpoints from Tox21 assays (assay source identifier 7) within ToxCast version 3.5 [16] using level 5 and 6 processing.

First, we used an in-house R script to filter the Tox21 multi-concentration summary file in order to identify chemicals based on their endpoint being either TSHR agonist (assay endpoint identifier 2040) or TSHR antagonist (assay endpoint identifier 2043) screened in HEK293T cell line. TSHR agonist is a chemical that binds to TSHR and fully activates it, whereas TSHR antagonist is a chemical that binds to TSHR but does not activate it and can additionally block the activation by any other agonist. Next, we filtered chemicals annotated as representative samples (i.e., gsid_rep is 1) and with reported activity value (i.e., modl_ga value is present) (Figure 3.1a). Subsequently, for these shortlisted chemicals, we accessed the two-dimensional (2D) structures provided by ToxCast version 3.5, or PubChem [145] if the 2D structures were not provided by ToxCast. Thereafter, we used MayaChemTools [109] to remove salts, mixtures, invalid structures and duplicated chemicals (Figure 3.1a). We also removed linear chemicals using the scaffold definition employed in our previous work [45]. Finally, we curated a TSHR agonist dataset of 509 chemicals (Supplementary Table S3.1) and a TSHR antagonist dataset of 650 chemicals (Supplementary Table S3.2). For each chemical in the two datasets, we additionally compiled the Chemical Abstracts Service (CAS) registry number or PubChem compound identifiers, reported biological activity (i.e., either active: hit_c is 1; or inactive: hit_c is 0), and the chemical concentration that generates the half maximal response (modl_ga, i.e., logarithm of AC_{50} value in micromolar concentration).

3.1.2 Molecular representation and annotation

We annotated chemicals in both TSHR agonist and TSHR antagonist datasets using molecular scaffolds and chemical classifications and their presence in different databases (Figure 3.1a). Following our previous work [45], we used the Bemis-Murcko definition [114] to compute the molecular scaffolds from chemical structures. Next, we relied on ClassyFire [110] to provide the corresponding chemical classifications. Further, we used DEDuCT [8, 15] database which compiles information on 792 EDCs curated from published literature with supporting evidence for endocrine disruption from experiments in



First chemical ●	First chemical ◖	Second chemical ●	Second chemical ◖	MOA-based classification
+	–	–	+	Strong MOA-cliff
–	+	+	–	Strong MOA-cliff
+	–	+	–	Same MOA
–	+	–	+	Same MOA
+	+	+	+	Same MOA
+	+	+	–	Weak MOA-cliff
+	+	–	+	Weak MOA-cliff
+	–	+	+	Weak MOA-cliff
–	+	+	+	Weak MOA-cliff

Figure 3.1 ***(previous page)*****:** Summary of structure-activity landscape analysis and activity cliff identification in a chemical dataset curated from ToxCast library. (a) Curation and annotation of TSHR agonist and antagonist datasets. (b) Structure-activity similarity (SAS) map based approach to identify the activity cliffs in a chemical dataset. (c) Steps involved in generation of a matched molecular pair (MMP) and associated MMP-cliff. (d) Classification of activity cliff pairs based on respective structural information. (e) MOA based classification of the chemical pairs (common to both TSHR agonist and antagonist datasets and having Tanimoto coefficient based similarity of > 0.35) into three different categories.

humans and rodents, to identify the known EDCs among chemicals in the TSHR agonist or TSHR antagonist dataset. We also used Organisation for Economic Co-operation and Development High Production Volume (OECD HPV) [146] or United States High Production Volume (USHPV) [147] databases to identify high production volume chemicals in our datasets. Additionally, we leveraged the CAS identifiers of the chemicals in TSHR agonist and TSHR antagonist datasets, which are also compiled in Distributed Structure-Searchable Toxicity (DSSTox) database, to retrieve annotations such as functional uses, occupational health hazard reports and product use composition from Chemical and Products Database (CPDat) [148] (Figure 3.1a).

3.1.3 Computation of activity difference

The activity difference for a pair of chemical is considered as the difference between their corresponding pAC$_{50}$ values, where pAC$_{50}$ is the negative logarithm of AC$_{50}$ value in molar concentration [36, 47, 149]. The activity values of the chemicals in the compiled TSHR agonist and TSHR antagonist datasets are reported as the logarithm of AC$_{50}$ values in micromolar concentrations (modl_ga). Therefore, we converted the modl_ga value to pAC$_{50}$ value using the following formulae:

$$AC_{50}(M) = 10^{modl\_ga} \times 10^{-6}$$

$$pAC_{50} = -log_{10}(AC_{50}(M)) = 6 - modl\_ga$$

Thereafter, we calculated the activity difference between two chemicals i and j using the following formula:

$$\text{Activity difference} = \left|(pAC_{50})_i - (pAC_{50})_j\right|$$

wherein the $(\mathrm{pAC}_{50})_i$ and $(\mathrm{pAC}_{50})_j$ are the pAC_{50} values of chemicals i and j, respectively.

3.1.4 Identification of activity cliffs using structure-activity similarity (SAS) map

We independently analyzed the activity landscape of the chemicals in TSHR agonist and TSHR antagonist datasets using SAS map (Figure 3.1b) [36–38, 45]. SAS map is a 2D representation where the structural similarity between the chemicals is plotted along the x-axis and the activity difference between the chemicals is plotted along the y-axis (Figure 3.1b). We computed structural similarity between chemical pairs based on Tanimoto coefficient (Tc) between the corresponding Extended-Connectivity Fingerprints with diameter 4 (ECFP4) of the chemicals. As there is no strict rule to choose a threshold for high structural similarity [150], we considered a similarity threshold of 0.35 which was close to three standard deviations from median of the computed Tc for chemical pairs in both TSHR agonist and TSHR antagonist datasets. We considered an activity difference threshold of 100 fold change which is equivalent to 2 logarithmic units. We designated the highly similar chemical pairs ($Tc > 0.35$) with high activity difference (≥ 2) as the activity cliffs in both TSHR agonist and TSHR antagonist datasets (Region III in Figure 3.1b). Additionally, we considered chemicals which form at least 5 activity cliff pairs as activity cliff generators (ACGs) [38, 45].

3.1.5 Identification of activity cliffs based on matched molecular pairs (MMPs)

In addition to SAS map based activity landscape analysis, we employed the MMP based approach to identify the activity cliffs (MMP-cliffs) [43] independently in TSHR agonist and TSHR antagonist datasets (Figure 3.1c). We used mmpdb platform [42] to generate MMPs for chemicals in both datasets. First, the mmpdb fragment module was used to fragment the chemical structure with 'none' value for both maximum number of non-hydrogen atoms and maximum number of rotatable bonds arguments. Next, the mmpdb index module was used to generate an exhaustive list of MMPs with 'none' value for maximum number of non-hydrogen atoms in the variable fragment argument. This gave us an exhaustive list of MMPs with detailed information on the constant part and transformations containing the exchanged fragments between chemical pairs. Further, to generate size-restricted MMPs, we implemented the following four criteria (Figure 3.1c) [43]:

(i) The difference in number of heavy atoms of the exchanged fragments in transformation should not exceed 8.

(ii) The constant part should be at least twice the size of each fragment in the transformation.

(iii) The number of heavy atoms of each fragment in the transformation should not exceed 13.

(iv) For a chemical pair with multiple MMPs, the transformation with the least difference in the number of heavy atoms between the exchanged fragments is considered.

Finally, we identified MMP-cliffs among the size-restricted MMPs by selecting those pairs with an activity difference ≥ 2 in logarithmic units (i.e., 100 fold change) (Figure 3.1c).

3.1.6 Activity cliff classification

In this study, we followed the activity cliff classification described in Vivek-Ananth *et al.* [45], to classify the activity cliffs by considering their molecular scaffolds, R-groups, R-group topology and chirality of chemical structures. Further, we modified the workflow in Vivek-Ananth *et al.* [45] to also check for topologically equivalent scaffolds (cyclic skeleton) when a pair of chemicals do not share the same scaffolds (Figure 3.1d) [44]. We used the R-group decomposition module available in RDKit [123] to decompose the chemical structure into its core structure (scaffold) and R-groups. Further, we used the modified workflow (Figure 3.1d) to manually classify the activity cliffs into the following 7 types:

(i) Chirality cliff: These are chemical pairs having the same scaffold, R-groups and R-group topology.

(ii) Topology cliff: These are chemical pairs having different R-group topologies while their scaffolds and R-groups remain unchanged.

(iii) R-group cliff: These are chemical pairs having different R-groups while their scaffolds remain unchanged.

(iv) Scaffold cliff: These are chemical pairs having different scaffolds while their cyclic skeletons, R-groups and R-group topologies remain unchanged.

(v) Scaffold/Topology cliff: These are chemical pairs having different scaffolds and R-group topologies while their cyclic skeletons and R-groups remain unchanged.

(vi) Scaffold/R-group cliff: These are chemical pairs having different scaffolds and R-groups while their cyclic skeletons remain unchanged.

(vii) Unclassified: These are chemical pairs having different scaffolds and cyclic skeletons.

3.1.7 Mechanism of action (MOA) based classification of chemical pairs

In addition to the activity cliffs in TSHR agonist and TSHR antagonist datasets, we were interested in identifying chemical pairs in which the chemicals have similar structures but differ in their mechanism of action (MOA). Such chemical pairs are designated as mechanism of action cliffs (MOA-cliffs) [47]. We considered chemicals which were common to both the TSHR agonist and TSHR antagonist datasets, and removed those chemicals which were reported as inactive MOA in both assays. We then computed the structural similarity of chemical pairs by using the Tc between the ECFP4 fingerprints of the shortlisted chemicals. We chose 0.35 as the similarity threshold (which is the structural similarity threshold used in SAS map analysis) to filter similar chemical pairs. Based on their MOA annotations in TSHR agonist and TSHR antagonist datasets, we classified these chemical pairs into 3 types (Figure 3.1e):

(i) Strong MOA-cliff: These are chemical pairs in which the chemicals have opposite MOA annotations.

(ii) Same MOA: These are chemical pairs in which both the chemicals have same MOA annotations.

(iii) Weak MOA-cliff: These are chemical pairs which could not be classified as either Strong MOA-cliff or Same MOA.

3.2 Results

3.2.1 Exploration of the chemical space of TSHR agonist and antagonist datasets

From ToxCast library, we curated 509 chemicals in TSHR agonist (Supplementary Table S3.1) and 650 chemicals in TSHR antagonist (Supplementary Table S3.2) datasets, and thereafter, annotated the chemicals in the two datasets with information on their molecular scaffolds, chemical classifications, and their presence in public documentation or databases (Figure 3.1a). Notably, there were 89 chemicals common between TSHR agonist and TSHR antagonist datasets. Additionally, we observed that chemicals in both TSHR agonist and TSHR antagonist datasets are structurally diverse (median Tc based similarity using ECFP4 fingerprints of ~0.11), which could be attributed to the diverse composition of environmental chemicals in the ToxCast chemical library, which are assessed for their adverse biological effects [16, 151].

For the 509 chemicals in the TSHR agonist dataset, after computing the molecular scaffolds we observed that the benzene scaffold is highly represented (as it is found in 122 chemicals). Many of the chemicals in TSHR agonist dataset are also categorized under the chemical class of ‘Benzene and substituted derivatives’ (195 chemicals) (Supplementary Table S3.1). Importantly, 79 chemicals in the TSHR agonist dataset are documented in DEDuCT [8, 15] as EDCs with experimental evidence, of which 29 EDCs have Category II evidence (Supporting evidence from *in vivo* rodent and *in vitro* human experiments but not from *in vivo* human experiments), 28 EDCs have Category III evidence (Supporting evidence from only *in vivo* rodent experiments), 21 EDCs have Category IV evidence (Supporting evidence from only *in vitro* human experiments) and 1 EDC has Category I evidence (Supporting evidence from *in vivo* human experiments). Among the 79 identified EDCs, 21 chemicals are also documented as high production volume chemicals as per OECD HPV or USHPV databases (Supplementary Table S3.1). Chemical and Products Database (CPDat) provided various functional use annotations for 102 chemicals, of which biocides, fragrance and antioxidants are the major reported functional categories (Supplementary Table S3.1). CPDat also provided the product use composition data for 70 chemicals, of which personal care, and cleaning products and household care are the major categories (Supplementary Table S3.1). Additionally, 4 chemicals namely, 3-Carene, Butylated hydroxytoluene, Hydroquinone and Triphenyl phosphate have been documented in various occupational health hazard reports (Supplementary Table S3.1).

Similarly, for the 650 chemicals in the TSHR antagonist dataset, we observed that benzene scaffold is the most represented molecular scaffold (as it is found in 127 chemicals), while ‘Benzene and substituted derivatives’ is the most represented chemical class (254 chemicals) (Supplementary Table S3.2). Notably, 65 chemicals in the TSHR antagonist dataset are documented as EDCs in DEDuCT, of which 26 EDCs have Category III evidence (Supporting evidence from only *in vivo* rodent experiments), 22 EDCs have Category II evidence (Supporting evidence from *in vivo* rodent and *in vitro* human experiments but not from *in vivo* human experiments) and 17 EDCs have Category IV evidence (Supporting evidence from only *in vitro* human experiments). Among the 65 identified EDCs, 13 are also documented as high production volume chemicals in OECD HPV or USHPV databases (Supplementary Table S3.2). CPDat provided functional uses for 156 chemicals, of which biocides, fragrance and antioxidants are reported as the major functional categories (Supplementary Table S3.2). CPDat also provided the product use composition data for 107 chemicals, of which personal care, pesticides, and cleaning products and household care are the major categories (Supplementary Table S3.2). Additionally, 4 antagonists namely, 2,2’,4,4’,5-Pentabromodiphenyl ether, 2,2’,4,4’-Tetrabromodiphenyl ether, Bibenzyl and Styrene are documented in various occupational health hazard reports (Supplementary Table S3.2).

3.2.2 Activity landscape analysis of TSHR agonist dataset

SAS map has been employed in the literature to identify activity cliffs by investigating the structure-activity relationship [36–38, 45]. Accordingly, we analyzed the activity landscape of the TSHR agonist dataset using the SAS map approach (Figure 3.2a). We observed that the majority of chemical pairs show similar activity while they are structurally diverse (SAS map Region 1 in Figure 3.2a). Importantly, we identified 79 chemical pairs showing high activity difference while being structurally similar (SAS map Region III in Figure 3.2a). We designated these 79 chemical pairs (formed by 60 unique chemicals) as activity cliffs (Supplementary Table S3.3), of which 9 chemicals are additionally identified as activity cliff generators (ACGs) (Supplementary Table S3.4). The chemicals forming activity cliffs are represented by 34 unique scaffolds with benzene and triphenyltin scaffolds being the highly represented scaffolds, and are categorized under 13 chemical classes with ‘Benzene and substituted derivatives’ class being the largest category. Moreover, triphenyltin scaffold is highly represented in chemicals forming ACGs. The chemicals forming pairs in the Region I (Scaffold hops) and Region IV (Unknown) are dominated by ‘Benzene and substituted derivatives’ chemical class followed by ‘Prenol lipids’ chemical class. Similarly, the chemicals forming pairs in the Region II (Smooth) are dominated by ‘Benzene and substituted derivatives’ chemical class followed by ‘Steroids and steroid derivatives’ chemical class.

MMP based activity landscape analysis has been alternatively employed in the literature to identify the activity cliffs [44, 47]. We also used the MMP approach to analyze the activity landscape of the TSHR agonist dataset. We identified 523 MMPs formed by 170 chemicals in the TSHR agonist dataset (Supplementary Table S3.5), of which 38 MMPs (formed by 19 unique chemical pairs) are identified as MMP-cliffs based on an activity difference threshold consideration similar to SAS map. Notably, the MMP-cliffs identified by the MMP approach are a subset of the activity cliffs identified by the SAS map approach, which could be attributed to the highly restrictive fragment transformation conditions imposed in the generation of MMPs [44]. Interestingly, the constant part containing three benzene rings identified in 14 of the 38 MMP-cliffs is similar to the highly represented triphenyltin scaffold among the chemicals forming activity cliffs identified through SAS map. Figure 3.2b shows chemical pairs of N,N’-Diphenyl-p-phenylenediamine (CAS:74-31-7) and N-Phenyl-1,4-benzenediamine (CAS:101-54-2), Triphenyl phosphate (CAS:115-86-6) and Triphenyltin acetate (CAS:900-95-8) that are identified as MMP-cliffs. N,N’-Diphenyl-p-phenylenediamine is an ACG which is documented as an EDC in DEDuCT and present in the OECD HPV or USHPV databases. Notably, Triphenyl phosphate and Triphenyltin acetate are documented as EDCs in DEDuCT and Triphenyl phosphate is also present in the OECD HPV or USHPV databases.

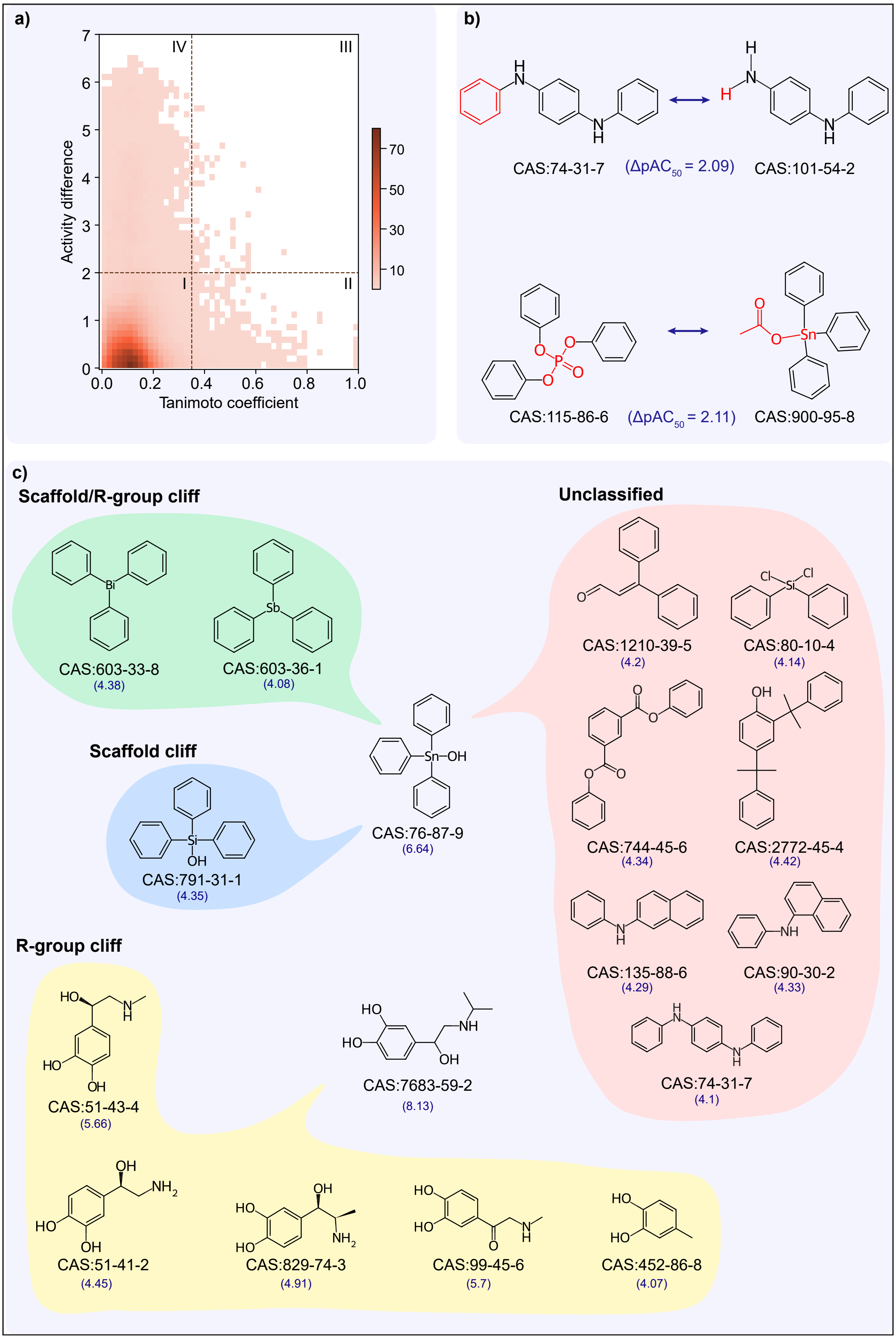

a)
Activity difference
Tanimoto coefficient
0.0
0.2
0.4
0.6
0.8
1.0
0
1
2
3
4
5
6
7
I
II
III
IV
70
50
30
10
b)
CAS:74-31-7
(ΔpAC_{50} = 2.09)
CAS:101-54-2
CAS:115-86-6
(ΔpAC_{50} = 2.11)
CAS:900-95-8
c)
Scaffold/R-group cliff
CAS:603-33-8
(4.38)
CAS:603-36-1
(4.08)
Scaffold cliff
CAS:791-31-1
(4.35)
CAS:76-87-9
(6.64)
Unclassified
CAS:1210-39-5
(4.2)
CAS:80-10-4
(4.14)
CAS:744-45-6
(4.34)
CAS:2772-45-4
(4.42)
CAS:135-88-6
(4.29)
CAS:90-30-2
(4.33)
CAS:74-31-7
(4.1)
R-group cliff
CAS:51-43-4
(5.66)
CAS:7683-59-2
(8.13)
CAS:51-41-2
(4.45)
CAS:829-74-3
(4.91)
CAS:99-45-6
(5.7)
CAS:452-86-8
(4.07)

Figure 3.2 ***(previous page)*****:** Activity landscape analysis of TSHR agonist dataset. (a) SAS map for TSHR agonist dataset. SAS map is divided into 4 quadrants by considering a similarity threshold of 0.35 and activity difference threshold of 2. Further, the density of data points in different regions of the SAS map is shown using a color gradient. (b) MMP-cliffs formed by N,N'-Diphenyl-p-phenylenediamine (CAS:74-31-7) with N-Phenyl-1,4-benzenediamine (CAS:101-54-2) [ΔpAC_{50} = 2.09] and Triphenyl phosphate (CAS:115-86-6) with Triphenyltin acetate (CAS:900-95-8) [ΔpAC_{50} = 2.11]. The transformed fragments resulting in MMP-cliff are highlighted in red color. (c) Activity cliff classifications for the ACGs, Triphenyltin hydroxide (CAS:76-87-9; 10 activity cliff pairs) and Isoproterenol (CAS:7683-59-2; 5 activity cliff pairs). The activity value (pAC_{50}) is mentioned below for each chemical.

Subsequently, we classified the 79 activity cliffs and identified 11 as R-group cliffs, 1 as scaffold cliff, 11 as Scaffold/R-group cliffs and 56 as unclassified (Supplementary Table S3.3). Figure 3.2c shows the different classifications of the activity cliffs formed by Triphenyltin hydroxide (CAS:76-87-9) and Isoproterenol (CAS:7683-59-2). Triphenyltin hydroxide forms 10 activity cliff pairs where 2 are Scaffold/R-group cliffs (same cyclic skeleton but differ in the scaffold as well as R-group), 1 is Scaffold cliff (same R-group, R-group topology and cyclic skeleton but differ only in scaffold) and remaining are Unclassified (differ in scaffold as well as the cyclic skeleton). Similarly, Isoproterenol forms 5 activity cliff pairs where all are R-group cliffs (same scaffold and cyclic skeleton but differ in R-groups). Further, we noted that majority of the identified activity cliffs (56 of 79) are classified under the Unclassified category as the chemicals forming these cliffs differ in their scaffolds as well as their scaffold topology (cyclic skeleton).

3.2.3 Activity landscape analysis of TSHR antagonist dataset

Similar to the activity landscape analysis of the TSHR agonist dataset, we analyzed the TSHR antagonist dataset through both SAS map and MMP approaches. From the SAS map approach, while most chemical pairs show similar activity despite having diverse structures (SAS map Region I in Figure 3.3a), 69 chemical pairs showed high activity difference while they are structurally similar (SAS map Region III in Figure 3.3a). We designated these 69 chemical pairs as activity cliffs, and observed that they are formed by 75 unique chemicals (Supplementary Table S3.6), of which 4 chemicals are ACGs

(Supplementary Table S3.7). The chemicals forming activity cliffs are represented by 39 unique scaffolds with benzene and biphenyl scaffolds being the highly represented scaffolds, and are categorized under 17 chemical classes with 'Benzene and substituted derivatives' class being the largest category. Similar to the activity cliff region, chemicals forming pairs in other three regions (Region I, II and IV) are also dominated by 'Benzene and substituted derivatives' chemical class followed by 'Steroids and steroid derivatives' chemical class.

From the MMP approach, we identified 590 MMPs (formed by 195 chemicals), of which 3 are MMP-cliffs (Supplementary Table S3.8). Notably all the MMP-cliffs are also activity cliffs identified through SAS map approach. Figure 3.3b shows chemical pairs of Styrene (CAS:100-42-5) and Phenylmercuric chloride (CAS:100-56-1), and Styrene and beta-Nitrostyrene (CAS:102-96-5). Styrene is an ACG which is documented as an EDC in DEDuCT and present in the OECD HPV or USHPV databases.

Further, we classified the 69 activity cliffs and identified 18 as R-group cliffs (same scaffold but differ in R-groups), 1 as Scaffold/R-group cliff (same cyclic skeleton but differ in both scaffold and R-group) and 50 as Unclassified (differ in both scaffold and cyclic skeleton) (Supplementary Table S3.6). Figure 3.3c shows 6 activity cliffs formed by Styrene, 5 R-group cliffs, and 1 Unclassified (differ in both scaffold and cyclic skeleton) and 1 Scaffold/R-group cliff formed by Norgestimate (CAS:35189-28-7) and Testosterone propionate (CAS:57-85-2). Finally, similar to the activity cliff classification in the TSHR agonist dataset, we noted that majority of the activity cliffs in the TSHR antagonist dataset (50 of 69) are classified under the Unclassified category.

3.2.4 Mechanism of action cliffs

Apart from the differences in activity, structurally similar chemicals also show a difference in their identified MOA. Hao *et al.* [47] have earlier explored the MMPs with different MOAs from androgen receptor agonist and antagonist datasets, and designated them as

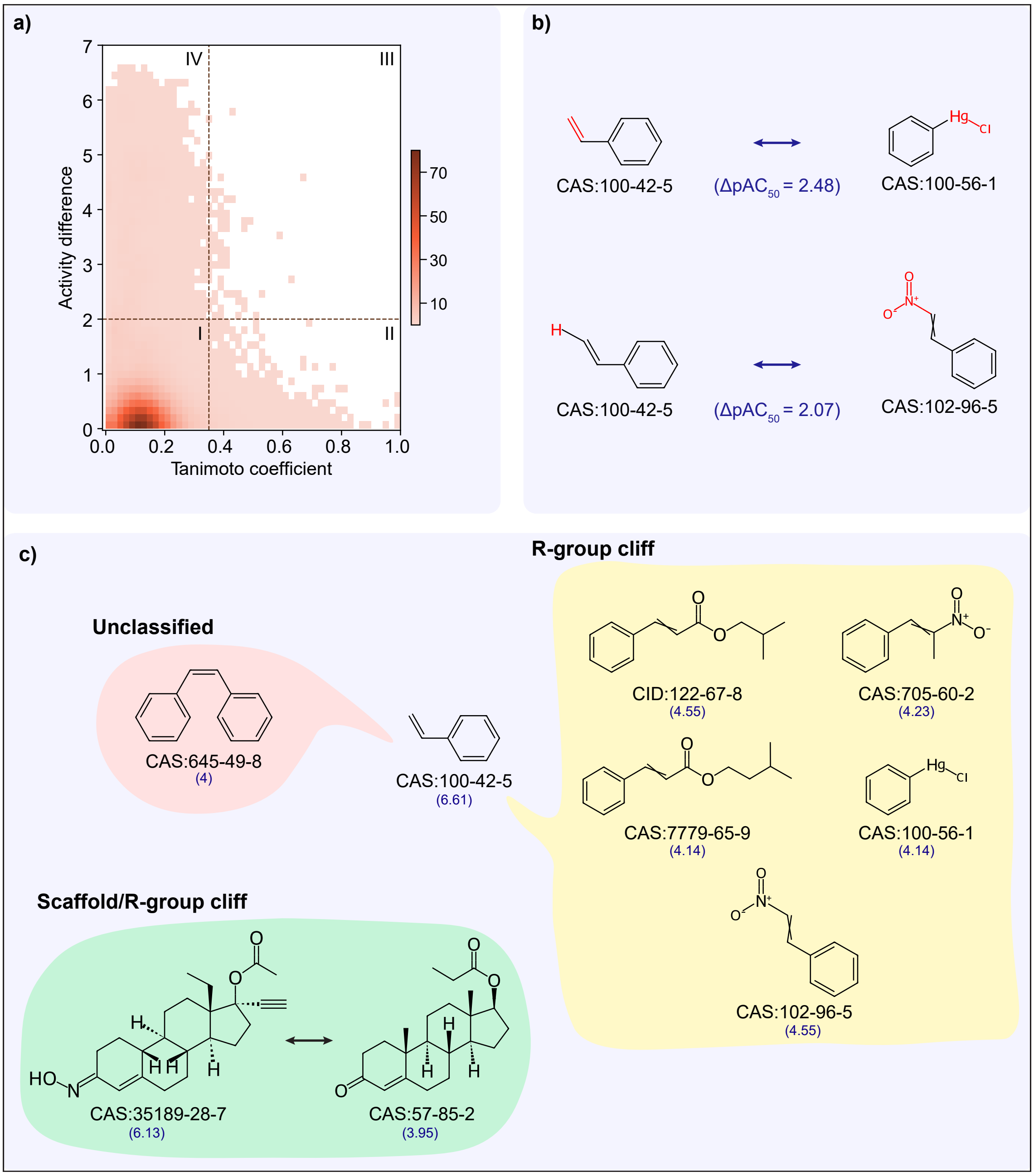


Figure 3.3: Activity landscape analysis of TSHR antagonist dataset. (a) SAS map for TSHR antagonist dataset. SAS map is divided into 4 quadrants by considering a similarity threshold of 0.35 and activity difference threshold of 2. Further, the density of data points in different regions of the SAS map is shown using a color gradient. (b) MMP-cliffs formed by Styrene (CAS:100-42-5) with Phenylmercuric chloride (CAS:100-56-1) [ΔpAC_{50} = 2.48] and with beta-Nitrostyrene (CAS:102-96-5) [ΔpAC_{50} = 2.07]. The transformed fragments resulting in MMP-cliff are highlighted in red color. (c) Activity cliff classifications for the activity cliff generator, Styrene (6 activity cliff pairs) and an activity cliff pair of Norgestimate (CAS:35189-28-7) with Testosterone propionate (CAS:57-85-2). The activity value (pAC_{50}) is mentioned below for each chemical.

MOA-cliffs. We shortlisted 75 chemicals which have endpoints in both TSHR agonist and TSHR antagonist datasets and identified 38 chemical pairs which have high structural similarity (Supplementary Table S3.9). We classified these 38 chemical pairs based on their MOA annotations and identified 3 as Strong MOA-cliffs, 16 as Same MOA and 19 as Weak MOA-cliffs (Figure 3.1e; Supplementary Table S3.9). Notably, 1 Strong MOA-cliff and 8 Weak MOA-cliffs are also classified as activity cliffs identified through the SAS map approach. Figure 3.4 shows examples of different MOA based classifications of highly similar chemical pairs (Tc > 0.35). 3,3'-Diaminobenzidine (CAS:91-95-2; inactive agonist and active antagonist) and 3,3'-Dimethylbenzidine (CAS:119-93-7; active agonist and inactive antagonist) form Strong MOA-cliff, Triphenyltin chloride (CAS:639-58-7; active agonist and active antagonist) and Triphenyltin hydroxide (CAS:76-87-9; active agonist and active antagonist) form Same MOA, and Endosulfan sulfate (CAS:1031-07-8; active agonist and active antagonist) and Endosulfan I (CAS:959-98-8; active agonist and inactive antagonist) form Weak MOA-cliff.

3.3 Discussion

The ToxCast program has screened nearly 10000 environmental chemicals for their adverse effects on various biological targets including TSHR, and characterized them based on their bioactivity and mechanisms of action [151, 152]. To date, ToxCast stands as the largest repository, providing experimentally determined activity data for thousands of chemicals through a standardized pipeline. Thus, the ToxCast dataset has greatly enabled the development of several QSAR models in predicting toxicity of chemicals and in prioritizing chemicals for further testing [16, 153]. In particular, the ToxCast library has been used to develop machine learning based QSAR models to predict chemicals that bind to TSHR [143,144]. However, there were no previous research focusing on the activity landscape analysis of chemicals from ToxCast library, in particular for the chemicals that can bind to TSHR, prior to this study.

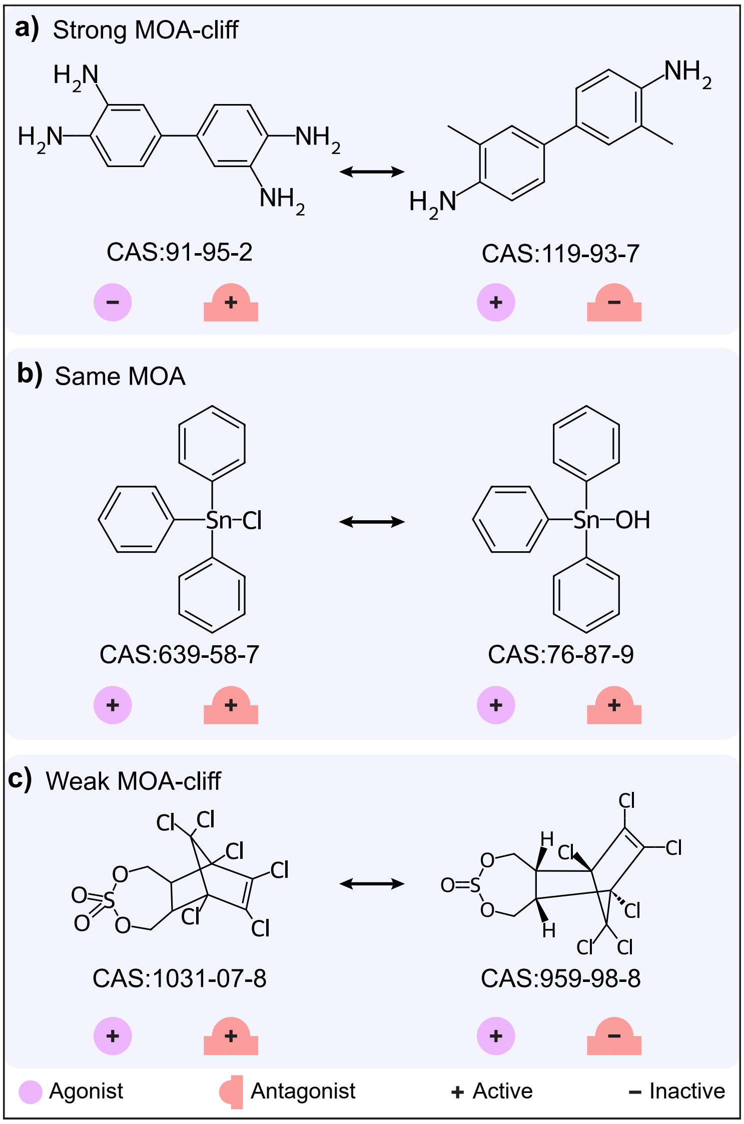


Figure 3.4: Examples for three different MOA based classifications of chemical pairs. (a) Strong MOA-cliff formed by 3,3'-Diaminobenzidine (CAS:91-95-2) with 3,3'-Dimethylbenzidine (CAS:119-93-7). (b) Same MOA formed by Triphenyltin chloride (CAS:639-58-7) with Triphenyltin hydroxide (CAS:76-87-9). (c) Weak MOA-cliff formed by Endosulfan sulfate (CAS:1031-07-8) with Endosulfan I (CAS:959-98-8).

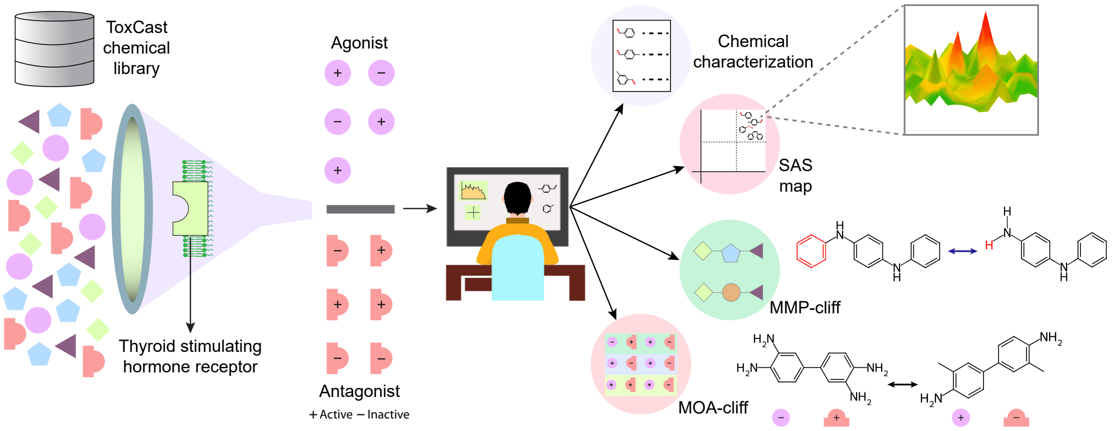


Figure 3.5: Schematic summary of the different computational approaches used to analyze the structure-activity and structure-mechanism relationships of TSHR binding chemicals.

Notably, this is the first study to report the heterogeneity of the structure-activity landscape as well as the structure-mechanism relationships of the TSHR binding chemicals compiled from ToxCast library (Figure 3.5). Here, along with the chemical fingerprint-based SAS map approach, we have also utilized a substructure-based MMP approach (independent of chemical fingerprints) and observed that the MMP is a more stringent approach than SAS map in identifying the activity cliffs. Further, from the analysis of the structure-mechanism relationships of TSHR binding chemicals, we identified structurally similar chemicals differing in their mechanisms of action i.e., agonist and antagonist (MOA-cliffs) (Figure 3.5). However, we were unable to investigate the molecular mechanisms behind the formation of activity cliffs and MOA-cliffs due to the lack of experimentally determined co-crystallized protein-ligand structures in the Protein Data Bank (PDB) [130]. We believe that the insights from this study will aid in development of better toxicity prediction models, and thereby, contribute towards characterization of the human exposome.

Supplementary Information

Supplementary Tables S3.1-S3.9 associated with this chapter are available for download from the GitHub repository: https://github.com/asamallab/PhDThesis-Ajaya_Kumar_Sahoo/blob/main/SI/ST_Chapter3.xlsx.

Code Availability

The computer programs used to perform the computations reported in this chapter are available in the following GitHub repository:

https://github.com/asamallab/PhDThesis-Ajaya_Kumar_Sahoo/tree/main/Codes.

Chapter 4

An integrative data-centric approach to derivation and characterization of an adverse outcome pathway network for cadmium-induced toxicity

Heavy metals are naturally occurring dense elements that are usually toxic in nature showing varying levels of toxicity depending on the dosage and time of exposure [17, 154]. The rising levels of heavy metals in the environment, owing to various industrial and anthropogenic activities, is of grave concern as they can negatively affect human health and environment [17, 155]. Cadmium is one such heavy metal which contaminates both terrestrial and aquatic environments, and is a major contributor of toxicity in various exposomes [156–158]. The prolonged biological half-life of cadmium coupled with its low excretion rates promotes accumulation of cadmium in humans and causes a wide range of disorders such as Itai-Itai disease, rheumatoid arthritis, cardiac diseases and reproductive disorders to name a few [156, 159–163]. Additionally, the toxic effects of cadmium are known to be dosage dependent, with varying impacts observed at different exposure levels [156, 160]. Similarly, cadmium accumulation in marine organisms, especially fish, has been reported to disrupt the endocrine systems and cause various reproductive and developmental disorders [164–168]. Due to its wide range of toxicities, cadmium has been identified as a priority pollutant by the United States Environmental Protection Agency (US EPA) [169]. Moreover, cadmium is classified as a carcinogen by the International Agency for Research on Cancer (IARC) [170].

The concept of adverse outcome pathway (AOP) network was proposed to elucidate the complex toxicity pathways underlying stressor-induced adverse effects [58]. Previously, Chai *et al.* [64] had integrated data within the Comparative Toxicogenomics Database (CTD) [27] and AOP-Wiki [53] to construct an AOP network specific to arsenic-induced reproductive toxicity. Jeong *et al.* [83] had additionally leveraged the chemical, gene, phenotype and disease associations within CTD to construct an AOP network specific to pulmonary fibrosis. Ravichandran *et al.* [77] had leveraged endocrine-mediated endpoints from DEDuCT [8, 15] to construct an AOP network specific to endocrine disruption. Knapen *et al.* [58] had leveraged ToxCast [16] assay endpoints to construct an AOP network specific to chemical mixtures in wastewater. In each of the above-mentioned studies, the integration of data from an external source had aided in identification of novel associations with existing AOPs, which resulted in expanded coverage of possible toxicity pathways. Therefore, it is relevant to integrate heterogeneous datasets from various exposome-relevant resources to construct an AOP network relevant for cadmium-induced toxicity.

In the previous two chapters, Chapters 2 and 3, we investigated the structure-activity and structure-mechanism relationships of environmental chemicals binding to endocrine receptors using various computational approaches. These analyses revealed the presence of heterogeneity namely, activity cliffs and MOA-cliffs, thereby characterizing the chemical space within the exposome. In this chapter, we aim to analyze the adverse health effects induced by a heavy metal namely, cadmium and its inorganic compounds, by leveraging the AOP framework. Notably, we construct and analyze the first AOP network specific to inorganic cadmium-induced toxicity by integrating heterogeneous data from several exposome-relevant resources. **The work reported in this chapter is contained in the published manuscript [95].**

4.1 Methods

4.1.1 Compilation of AOPs from AOP-Wiki

AOP-Wiki [53] is a large public repository hosted by the Society for the Advancement of Adverse Outcome Pathways (SAAOP), and the online resource compiles detailed qualitative information on AOPs that are being developed globally. To access the information compiled in AOP-Wiki, we downloaded the XML file (released on 1 April 2023) from the AOP-Wiki 'Projects Downloads' page which was last accessed on 31 October 2023. Thereafter, we used an in-house python script to parse and retrieve information from the downloaded XML file. For each AOP in the downloaded XML file, we retrieved information on AOP identifier, AOP title, associated key events (KEs) (including molecular initiating events - MIEs and adverse outcomes - AOs), key event relationships (KERs), linked stressors, and the status according to Organisation for Economic Co-operation and Development (OECD) and SAAOP. Additionally, we retrieved the biological applicability information for AOPs such as taxonomy, sex and life-stage of the organism, and their weight of evidence. For each KE associated with an AOP, we retrieved the corresponding KE identifier, KE title, level of biological organization, action name, object name and identifiers, and process name, source and identifiers. For each KER associated with an AOP, we retrieved corresponding information on upstream and downstream KEs, adjacency, evidence for biological plausibility, and extent of quantitative understanding. Note that, an adjacency value of 'Adjacent' suggests the existence of direct link between the upstream and downstream KEs, whereas a value of 'Non-adjacent' suggests the existence of intermediate KE(s) [171].

4.1.2 Filtration of ‘high confidence AOPs’ within AOP-Wiki

AOP-Wiki is a living document as several AOPs are under development resulting in continuous update and improvement of the resource [171]. Therefore, it is crucial to assess the quality and completeness of AOPs including the associated information before considering them to build specific AOP networks [59, 77]. Building upon the earlier work by Ravichandran *et al.* [77], we developed a detailed workflow (Figure 4.1) that employs both computational methods and manual curation efforts in tandem to filter AOPs that are non-empty, connected, complete and are of sufficient quality based on the information provided in AOP-Wiki. Note, the AOPs that satisfy the above criteria are referred to as ‘high confidence AOPs’ in this work.

Initially, using an in-house python script, we retrieved information on 437 AOPs from the downloaded AOP-Wiki XML file. First, we checked the SAAOP status and removed 6 ‘archived’ AOPs (Figure 4.1), as they are not under active development and are marked as unsuitable for further adoption [172]. Then, we checked and removed 3 AOPs that contained at least one KE with title as ‘unknown’, due to the uncertainty in the associated biological event (Figure 4.1).

Subsequently, we checked the remaining 428 AOPs to remove empty AOPs which lack KEs. To ensure that the downloaded XML file from AOP-Wiki was up-to-date with information on the AOP page in the online repository, we manually updated any empty AOP determined based on information in the downloaded XML file, if KEs were listed in the ‘Events’ table in the ‘Summary of the AOP’ section of the corresponding AOP page in AOP-Wiki (last accessed on 19 November 2023). Furthermore, for an empty AOP determined based on information in the downloaded XML file, we also checked the ‘Graphical Representation’ section when the ‘Events’ table was empty on the corresponding AOP page in AOP-Wiki (last accessed on 19 November 2023). This combined computational and manual effort led to the removal of 20 empty AOPs which lack KEs (Figure 4.1).

Next, we checked the remaining 408 AOPs for complete absence of KERs. We man-

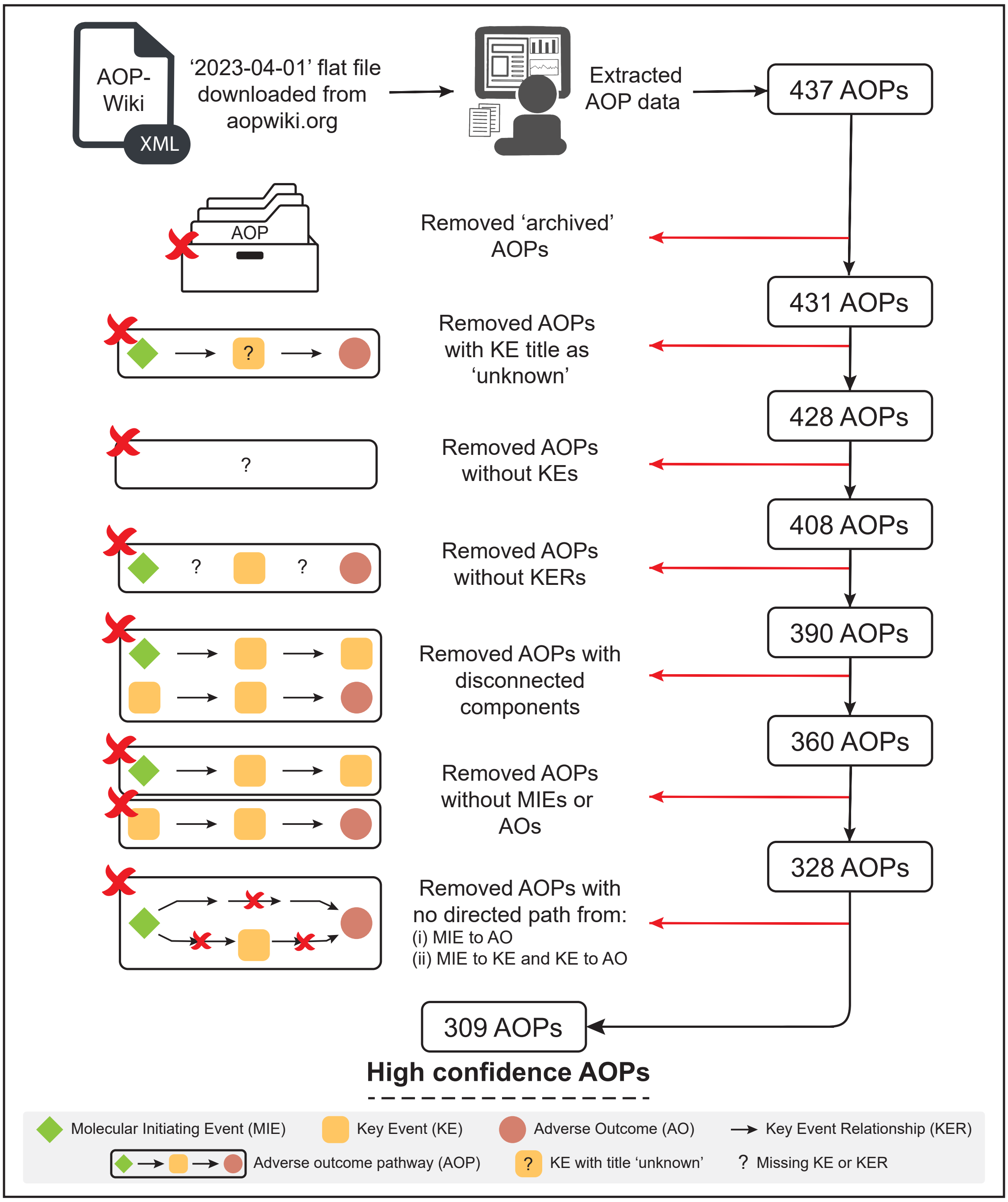

AOP-Wiki
XML
'2023-04-01' flat file downloaded from aopwiki.org
Extracted AOP data
437 AOPs
AOP
Removed 'archived' AOPs
431 AOPs
?
Removed AOPs with KE title as 'unknown'
428 AOPs
?
Removed AOPs without KEs
408 AOPs
? ?
Removed AOPs without KERs
390 AOPs
Removed AOPs with disconnected components
360 AOPs
Removed AOPs without MIEs or AOs
328 AOPs
Removed AOPs with no directed path from:
(i) MIE to AO
(ii) MIE to KE and KE to AO
309 AOPs
High confidence AOPs
Molecular Initiating Event (MIE)
Key Event (KE)
Adverse Outcome (AO)
Key Event Relationship (KER)
Adverse outcome pathway (AOP)
? KE with title 'unknown'
? Missing KE or KER

Figure 4.1: Workflow to filter high confidence AOPs from AOP-Wiki by employing computation and manual curation in conjunction.

ually updated the AOPs lacking KERs based on information in the downloaded XML file, if KERs were listed in the 'Relationships Between Two Key Events' table in the 'Summary of the AOP' section of the corresponding AOP page in AOP-Wiki (last accessed on 19 November 2023). Furthermore, for an AOP lacking KERs based on information in the downloaded XML file, we also checked the 'Graphical Representation' section when the 'Relationships Between Two Key Events' table was empty. This combined computational and manual effort led to the removal of 18 AOPs with complete absence of KERs (Figure 4.1).

Next, we checked for the presence of disconnected components in the remaining 390 AOPs. This computation of the number of connected components in an AOP was performed using the NetworkX [173] python package. We manually updated the AOPs containing disconnected components (i.e., containing more than one connected component) based on information in the downloaded XML file, if additional KERs were available in the 'Relationships Between Two Key Events' table in the 'Summary of the AOP' section of the corresponding AOP page in AOP-Wiki (last accessed on 19 November 2023). This combined computational and manual effort led to removal of 30 AOPs containing disconnected components (Figure 4.1).

Subsequently, we checked the remaining 360 AOPs for the presence of at least one MIE and at least one AO (Figure 4.1). We manually updated the AOPs containing no MIE and/or no AO based on information in the downloaded XML file, if additional information was available on the 'Events' table in the 'Summary of the AOP' section of the corresponding AOP page in AOP-Wiki (last accessed on 19 November 2023). This led to removal of 32 AOPs lacking MIE and/or AO (Figure 4.1).

Finally, we checked the remaining 328 AOPs for the existence of: (i) a directed path that originates from at least one MIE and terminates in at least one AO; (ii) a directed path to every KE that originates from at least one MIE; (iii) a directed path from every KE that terminates in at least one AO. We manually updated AOPs that were not complying with all the three path criteria based on information in the downloaded XML file, if additional information was available from the ‘Relationships Between Two Key Events’ table in the ‘Summary of the AOP’ section of the corresponding AOP page in AOP-Wiki (last accessed on 19 November 2023). This led to removal of 19 AOPs based on the three path criteria, and the remaining 309 AOPs were designated as high confidence AOPs (Figure 4.1).

Supplementary Table S4.1 contains the list of 309 high confidence AOPs filtered using the above-mentioned criteria in this study. These 309 high confidence AOPs comprise 1054 unique KEs (Supplementary Table S4.2) and 1599 unique KERs (Supplementary Table S4.3). Note that, while filtering for the high confidence AOPs (Figure 4.1), we had to assign identifiers to certain KEs and KERs that were manually compiled from AOP-Wiki, as the corresponding identifiers were not available in AOP-Wiki.

4.1.3 Identification of KEs associated with inorganic cadmium

The aim of this study is to build and investigate the network of AOPs within AOP-Wiki that are relevant for inorganic cadmium-induced toxicity. To this end, we first identified KEs associated with inorganic cadmium using five different sources namely, AOP-Wiki [53], Comparative Toxicogenomics Database (CTD) [27], ToxCast [16], DEDuCT [8, 15] and NeurotoxKb [29] (Figure 4.2).

KEs associated with inorganic cadmium from AOP-Wiki

For each AOP, we have retrieved information on the stressors that can trigger its progression from the downloaded XML file. We find that 2 of the 309 high confidence AOPs namely, AOP:257 and AOP:296, are documented to be associated with inorganic cadmium, specifically, cadmium and cadmium chloride, in AOP-Wiki. Thus, we considered the 10 KEs comprising the 2 AOPs to be associated with inorganic cadmium (Figure 4.2). Additionally, we perused through the AOP-Wiki pages of these 2 AOPs, and compiled any information on study type and dosages for the corresponding KEs.

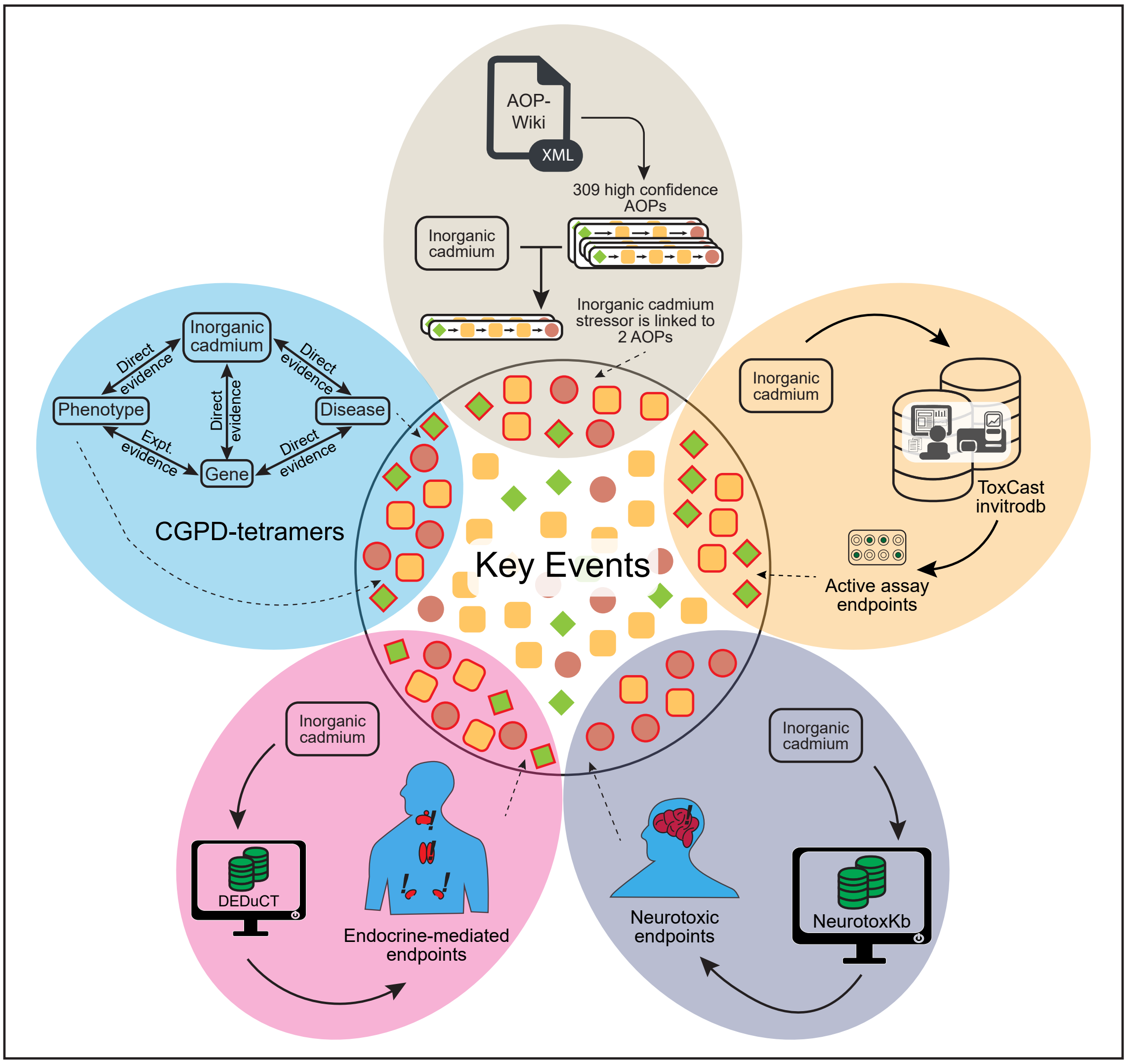


Figure 4.2: Workflow to identify the KEs in AOP-Wiki that are associated with inorganic cadmium based on information in five resources namely, AOP-Wiki, CTD, ToxCast, DEDuCT and NeurotoxKb.

Identification of KEs associated with inorganic cadmium using CGPD-tetramers from CTD

CTD [27] is among the largest toxicogenomics resources that compiles published information on health effects due to chemical exposures. CTD provides information on associations among chemicals, genes or proteins, pathways, phenotypes, and diseases through systematic curation from published literature and other resources. Recently, Jeong *et al.* [83] leveraged the chemical (C), gene (G), phenotype (P) and disease (D) tetramers, i.e., CGPD-tetramers, constructed from the data compiled in CTD, to identify the KEs associated with pulmonary fibrosis. Here, we followed the workflow proposed by Davis *et al.* [174] to construct the CGPD-tetramers specific to inorganic cadmium, and thereafter, leveraged them to identify the associated KEs in AOP-Wiki.

Specifically, we leveraged data from CTD's September 2023 release (last accessed on 31 October 2023) to construct the CGPD-tetramers specific to inorganic cadmium. First, we identified the list of chemicals within CTD which correspond to inorganic cadmium, namely, cadmium, cadmium chloride, cadmium nitrate, cadmium oxide, cadmium sulfate, cadmium sulfide, cadmium telluride and cadmium selenide. For each of these chemicals, we compiled the corresponding chemical-gene, chemical-phenotype, chemical-disease and gene-disease pairs from CTD. Further, we leveraged Gene Ontology (GO) annotations of genes from NCBI Gene resource [175] (last accessed on 31 October 2023) to identify gene-phenotype pairs. Thereafter, to construct the inorganic cadmium specific CGPD-tetramers, we considered: (i) chemical-gene and chemical-phenotype associations with literature evidence; (ii) chemical-disease and gene-disease associations with 'marker/mechanism' or 'marker/mechanism|therapeutic' evidence; (iii) gene-phenotype associations with GO annotations based on only the experimental results [176]. This construction procedure resulted in a non-redundant list of 9873 CGPD-tetramers which comprise 3 chemicals (cadmium, cadmium chloride and cadmium sulfate), 849 genes, 309 phenotypes (GO terms), and 163 disease terms (Supplementary Table S4.4). Subsequently, we manually mapped the phenotype and disease terms in 9873 CGPD-tetramers specific to inorganic cadmium to the KEs in AOP-Wiki, in order to identify the KEs associated with inorganic cadmium.

For the CGPD-tetramer phenotype linked GO terms, we generated the immediate neighbor terms (both parent and children GO terms) using the GOSim [177] package available in R programming language. Next, we overlapped the GO terms (along with their neighbor terms) with the process identifiers of KEs in AOP-Wiki, and manually inspected the KE title and phenotype name before accepting any mapping between CTD phenotypes and KEs in AOP-Wiki. Through this exercise, we were able to map 88 phenotypes in CGPD-tetramers to 181 KEs in AOP-Wiki (Figure 4.2). Additionally, for each of these mapped phenotypes, we compiled the corresponding literature reference from CTD, and thereafter manually curated the chemical name, study type and dosage information from these literature references.

For the CGPD-tetramer disease terms, we used disease identifier and disease name, and manually mapped them to KEs in AOP-Wiki using the process identifiers and the title of KEs. Through this exercise, we were able to map 70 disease terms in CGPD-tetramers to 60 KEs in AOP-Wiki (Figure 4.2). Additionally, for each of these mapped diseases, we compiled the corresponding literature reference from CTD, and thereafter manually curated the chemical name, study type and dosage information from these literature references.

Identification of KEs associated with inorganic cadmium using ToxCast assay endpoints

ToxCast is a US EPA project, which has experimentally screened nearly 10000 environmental chemicals to assess their adverse effects. Here, we also leveraged the ToxCast assay endpoints reported for inorganic cadmium to identify the associated KEs in AOP-Wiki. To this end, we downloaded the latest ToxCast invitrodb version 4.1 dataset from the US EPA repository [178], and thereafter, retrieved the active assay endpoints (‘hitc’

$\geq$ 0.9) for two inorganic cadmium compounds namely, cadmium chloride and cadmium nitrate, from the ‘mc5-6_winning_model_fits-flags_invitrodb_v4_1_SEPT2023.csv’ file. In ToxCast, the ‘activatory’ or ‘inhibitory’ effect of a chemical is obtained from the sign of the ‘top’ value of the winning model mentioned in the ‘mc4_all_model_fits_invitrodb_v4_1_SEPT2023.csv’ file [179]. Further, we accessed the ‘assay_gene_mappings_invitrodb_v4_1_SEPT2023.xlsx’ file to obtain the biological metadata for the shortlisted assay endpoints.

Subsequently, we manually mapped the shortlisted assay endpoints in ToxCast for the 2 inorganic cadmium compounds to KEs in AOP-Wiki. To elaborate, we first overlapped the target biological process, gene identifier, gene names and gene aliases corresponding to the assay endpoints in ToxCast with the KEs in AOP-Wiki based on their titles, object identifier and object name. Next, we manually inspected the assay endpoint description, ‘activatory’ or ‘inhibitory’ effect of the chemical in the assay endpoint, and the action of the overlapped KE, prior to accepting a mapping between an assay endpoint in ToxCast and KE in AOP-Wiki. This procedure resulted in the mapping of 30 KEs in AOP-Wiki to 28 ToxCast assay endpoints specific to 2 inorganic cadmium compounds (Figure 4.2). Additionally, for each of these mapped endpoints, we compiled the corresponding chemical name, cell type and AC_{50} values (concentration of half maximal activity).

Identification of KEs associated with inorganic cadmium using DEDuCT and NeurotoxKb

DEDuCT [8,15] is among the largest resources on endocrine disrupting chemicals (EDCs) that has compiled manually curated information on such chemicals along with supporting evidence from published literature. Here, we also leveraged DEDuCT to identify KEs in AOP-Wiki associated with inorganic cadmium. Specifically, we compiled the endocrine-mediated endpoints for three inorganic cadmium compounds namely, cadmium, cadmium chloride, and cadmium nitrate, from DEDuCT, and thereafter, manually mapped the endpoints to KEs in AOP-Wiki based on their titles. This procedure resulted in the mapping of 58 KEs in AOP-Wiki to 33 DEDuCT endpoints specific to 3 inorganic cadmium compounds (Figure 4.2). Additionally, for each of these mapped endpoints, we compiled the chemical name, study type, dosage information and corresponding literature evidence from DEDuCT.

NeurotoxKb [29] is a dedicated resource that has compiled manually curated information on neurotoxicants along with supporting evidence in mammals from published literature. Here, we also leveraged NeurotoxKb to identify KEs in AOP-Wiki associated with inorganic cadmium. Specifically, we compiled the neurotoxic endpoints for two inorganic cadmium compounds namely, cadmium and cadmium chloride, from NeurotoxKb, and thereafter, manually mapped the endpoints to KEs in AOP-Wiki based on their titles. This procedure resulted in the mapping of 7 KEs in AOP-Wiki to 5 NeurotoxKb endpoints specific to 2 inorganic cadmium compounds (Figure 4.2). Additionally, for each of these mapped endpoints, we compiled the literature reference mentioned in NeurotoxKb and thereafter manually curated the chemical name, study type and dosage information from the corresponding literature evidence.

Overall, by integrating information contained in AOP-Wiki, CTD, ToxCast, DEDuCT and NeurotoxKb, we compiled a list of 312 KEs (Supplementary Table S4.5) in AOP-Wiki with published evidence of being associated with inorganic cadmium, specifically, cadmium, cadmium chloride, cadmium sulfate and cadmium nitrate. Additionally, from each of these sources we curated the type of study (*in vivo* / *in vitro* / *in silico* / cohort study) and dosage information (if available) (Supplementary Table S4.6).

4.1.4 Curated subset of AOPs relevant for cadmium-induced toxicity

After compiling the list of 312 KEs associated with inorganic cadmium, we find that 241 of the 309 high confidence AOPs contain at least one KE associated with inorganic cadmium (Figure 4.3). Of these, we find that 34 high confidence AOPs have at least one MIE and at least one AO associated with inorganic cadmium (Figure 4.3). Moreover, we ascertained that the 34 high confidence AOPs have at least one directed path that originates from a MIE associated with inorganic cadmium and terminates in an AO associated with inorganic cadmium (Figure 4.3). Lastly, for each of these 34 high confidence AOPs, we computed the coverage score [64] which is the ratio of the number of KEs associated with inorganic cadmium to the total number of KEs in an AOP. By imposing a coverage score threshold of ≥ 0.4 [64], we identified a subset of 30 high confidence AOPs as relevant for cadmium-induced toxicity and designated them as 'cadmium-AOPs' (Figure 4.3). Subsequently, we considered the subset of 30 cadmium-AOPs to construct an AOP network relevant for cadmium-induced toxicity.

The 30 cadmium-AOPs comprise 98 unique KEs and 130 unique KERs. For each KER in an AOP, AOP-Wiki provides corresponding information on the weight of evidence (WoE) for its biological plausibility (Supplementary Table S4.7). Following Ravichandran *et al.* [77], we leveraged this WoE information for KERs in terms of 'High', 'Moderate', 'Low' or 'Not Specified', to compute the fraction of KERs in an AOP with 'High' WoE [i.e., F(High)], the fraction of KERs in an AOP with 'Moderate' WoE [i.e., F(Moderate)], the fraction of KERs in an AOP with 'Low' WoE [i.e., F(Low)], and the fraction of KERs in an AOP with 'Not Specified' WoE [i.e., F(Not Specified)]. Thereafter, we assigned the cumulative WoE using the following criteria [77]:

(i) If $F(\text{High}) \geq 0.5$, the cumulative WoE of the AOP is assigned as 'High'.

(ii) If $F(\text{High}) < 0.5$, but $(F(\text{High}) + F(\text{Moderate})) \geq 0.5$, the cumulative WoE of the AOP is assigned as 'Moderate'.

(iii) If $(F(\text{High}) + F(\text{Moderate})) < 0.5$, but $(F(\text{High}) + F(\text{Moderate}) + F(\text{Low})) \geq 0.5$, the cumulative WoE of the AOP is assigned as 'Low'.

(iv) If none of the above-mentioned three conditions are satisfied, the cumulative WoE of the AOP is assigned as 'Not Specified'.

Supplementary Table S4.8 provides the cumulative WoE for each of the 30 cadmium-AOPs. For each of the 30 cadmium-AOPs, we have also compiled information on taxonomic, sex, and life-stage applicability along with the corresponding levels of evidence

(Supplementary Table S4.8).

4.1.5 AOP network construction and visualization

To better understand the shared relationships among the 30 cadmium-AOPs, we constructed an undirected AOP network based on shared KEs among the AOPs. In the undirected AOP network, nodes correspond to 30 cadmium-AOPs, and there exists an edge between any pair of cadmium-AOPs if they share at least one KE (Figure 4.4). After constructing the undirected AOP network comprising 30 cadmium-AOPs, we visualized and identified the connected components in the network using Cytoscape [180].

Moreover, we constructed a directed AOP network comprising KEs and KERs in the 30 cadmium-AOPs. Since the undirected AOP network consists of multiple disconnected components (Figure 4.4), the directed AOP network also comprises multiple components. In particular, we constructed the directed network corresponding to the largest connected component of the AOP network (Figure 4.5). In the directed network, the nodes represent KEs and a directed edge represents a KER linking its upstream KE to its downstream KE. For the directed AOP network, we computed different network measures namely, in-degree, out-degree, eccentricity, and betweenness centrality using NetworkX [173] python package.

4.2 Results

4.2.1 Integrative data-centric construction and analysis of cadmium-AOP network

In this study, we aimed to construct and analyze an AOP network relevant for inorganic cadmium-induced toxicity. To achieve this, we first retrieved the AOP data from AOP-Wiki, assessed their quality and completeness, and curated 309 high confidence AOPs (Figure 4.1; Supplementary Table S4.1). Thereafter, by leveraging various exposome-

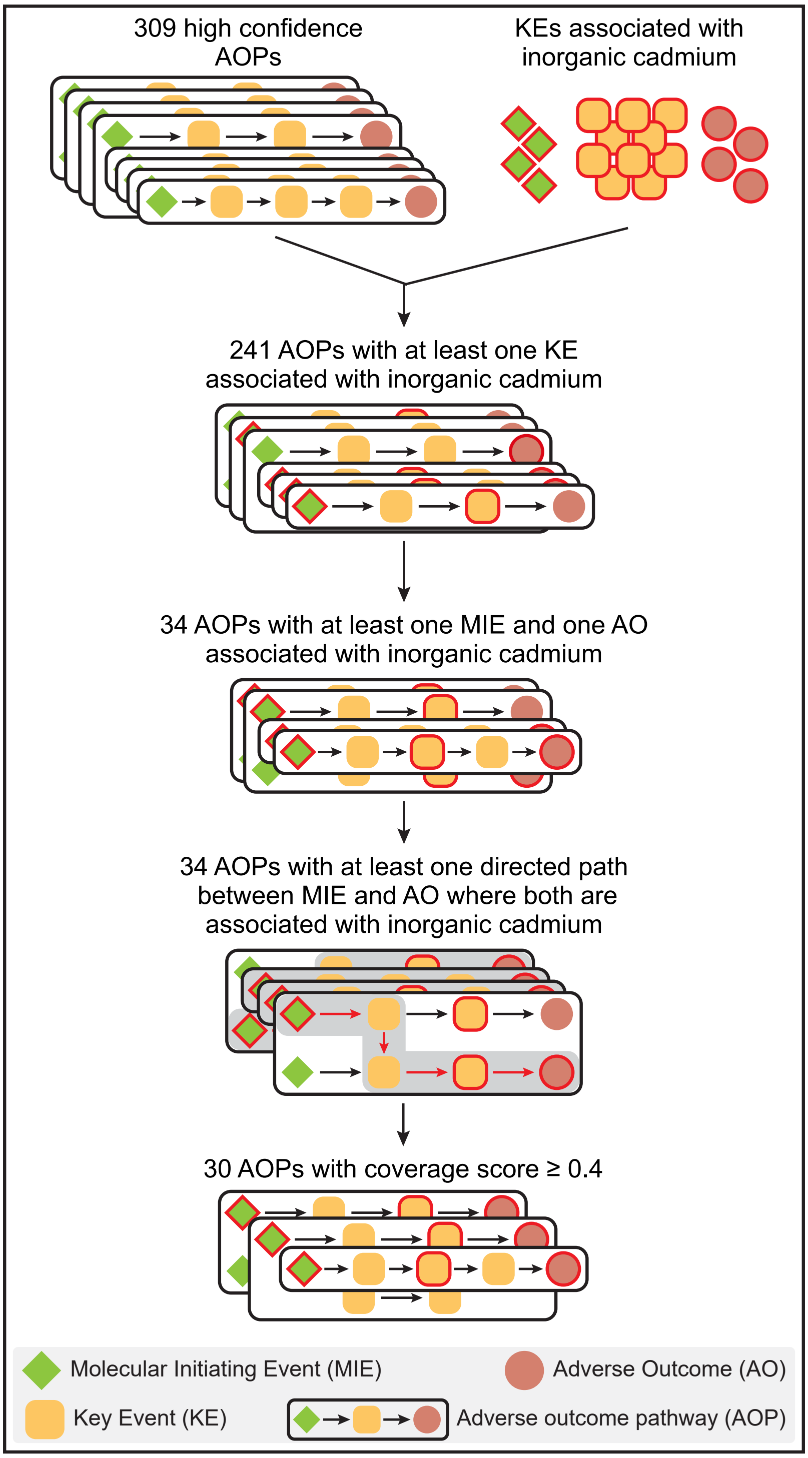

309 high confidence AOPs
KEs associated with inorganic cadmium
241 AOPs with at least one KE associated with inorganic cadmium
34 AOPs with at least one MIE and one AO associated with inorganic cadmium
34 AOPs with at least one directed path between MIE and AO where both are associated with inorganic cadmium
30 AOPs with coverage score ≥ 0.4
Molecular Initiating Event (MIE)
Adverse Outcome (AO)
Key Event (KE)
Adverse outcome pathway (AOP)

Figure 4.3: Workflow to identify AOPs relevant for cadmium-induced toxicity i.e., cadmium-AOPs, from the curated list of high confidence AOPs.

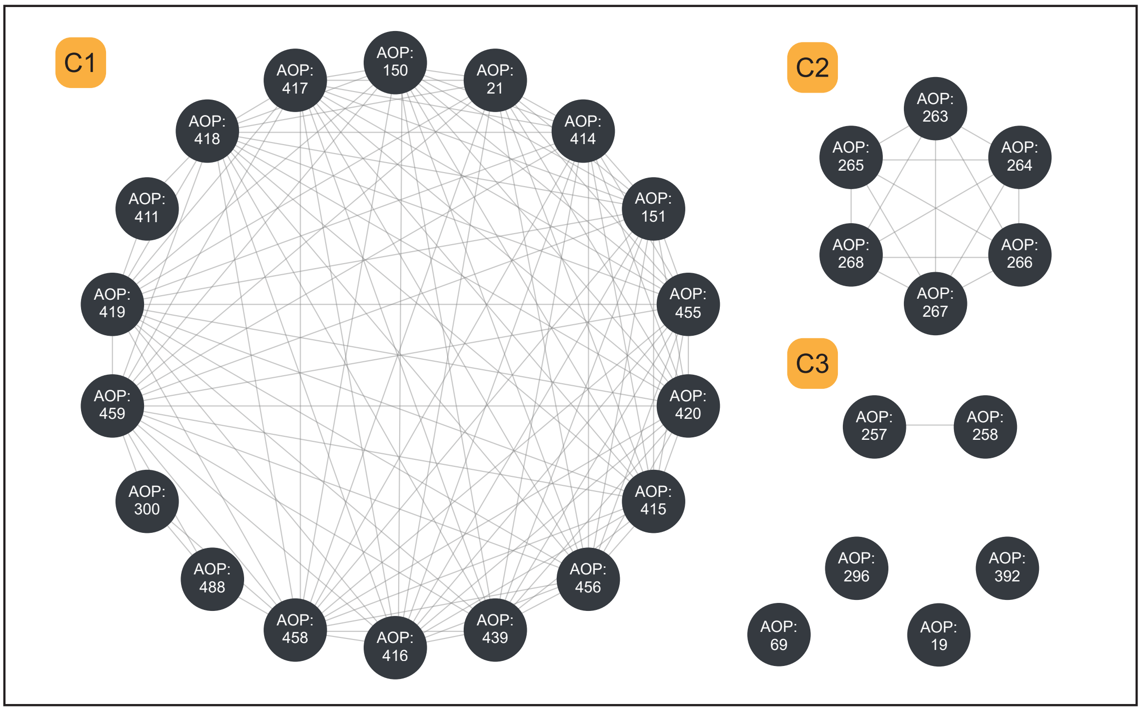


Figure 4.4: Undirected network of 30 cadmium-AOPs. Each node corresponds to a cadmium-AOP, and an edge between two nodes means that those two cadmium-AOPs share at least one KE. This network has 3 connected components (with two or more nodes) which are labeled as C1, C2 and C3, and 4 isolated nodes.

relevant databases such as AOP-Wiki, CTD, ToxCast, DEDuCT and NeurotoxKb, we systematically identified 312 KEs present in AOP-Wiki to be associated with inorganic cadmium-induced toxicity (Figure 4.2; Supplementary Table S4.5). Additionally, we curated the study type and dosages from each of these sources, and observed that the majority of the KEs have *in vivo* evidence for varying dosages of cadmium-induced toxicity (Supplementary Table S4.6). Finally, by applying various criteria on the specialized KEs such as MIEs and AOs, and considering a coverage score cut-off of 0.4, we identified 30 high confidence AOPs relevant for inorganic cadmium-induced toxicity (which are designated as ‘cadmium-AOPs’) (Figure 4.3). Importantly, we emphasize that AOP-Wiki had linked only 2 AOPs (AOP:257 and AOP:296) to inorganic cadmium stressor, whereas our systematic integration of heterogeneous data from diverse sources led to the identification of 28 additional AOPs within AOP-Wiki to be relevant for inorganic cadmium-induced toxicity.

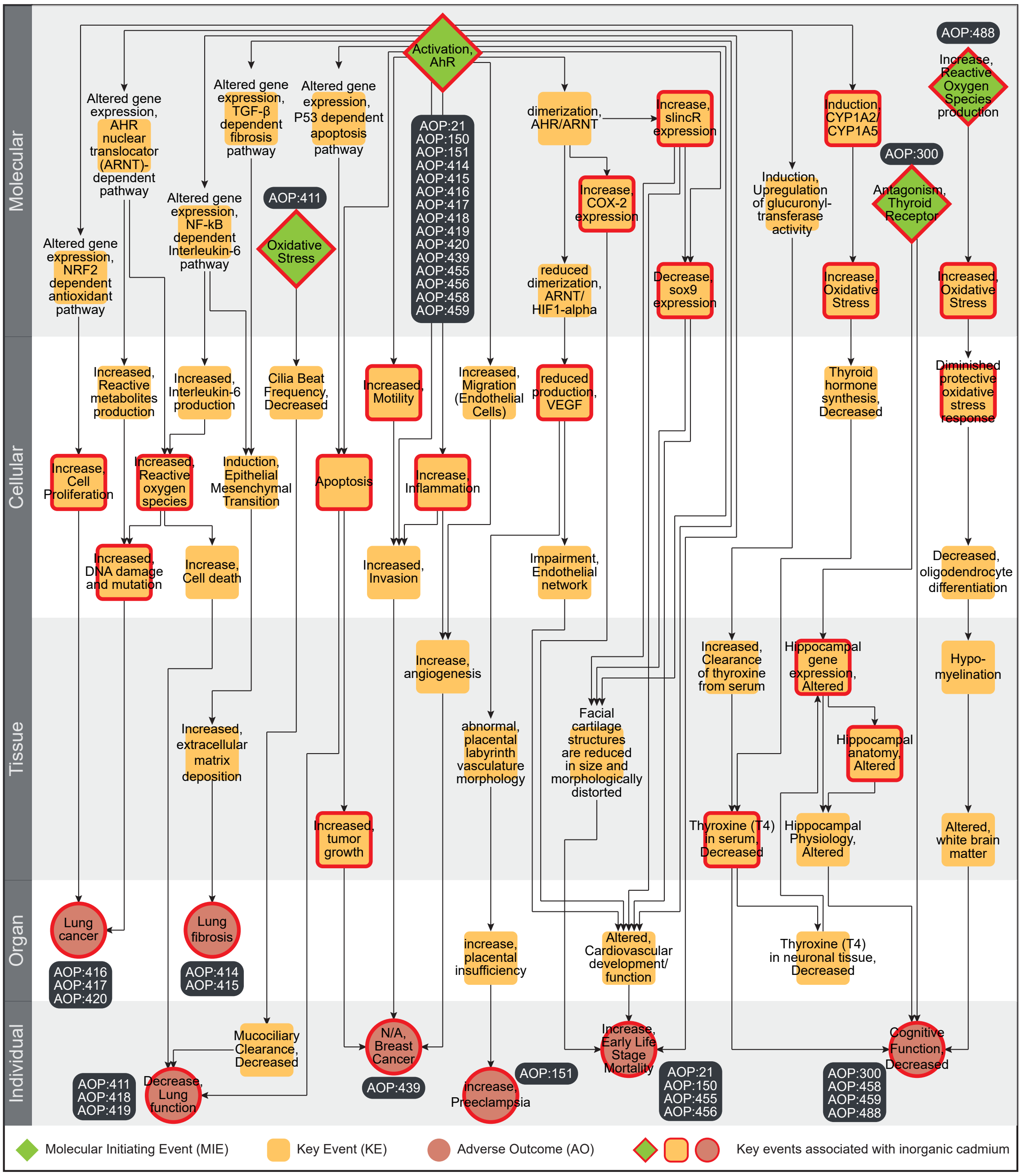

Molecular
Cellular
Tissue
Organ
Individual
AOP:488
Activation, AhR
Altered gene expression, AHR nuclear translocator (ARNT)-dependent pathway
Altered gene expression, TGF-β dependent fibrosis pathway
Altered gene expression, P53 dependent apoptosis pathway
dimerization, AHR/ARNT
Increase, slincR expression
Induction, CYP1A2/ CYP1A5
Increase, Reactive Oxygen Species production
AOP:21
AOP:150
AOP:151
AOP:414
AOP:415
AOP:416
AOP:417
AOP:418
AOP:419
AOP:420
AOP:439
AOP:455
AOP:456
AOP:458
AOP:459
AOP:300
Antagonism, Thyroid Receptor
Induction, Upregulation of glucuronyl-transferase activity
Increase, COX-2 expression
AOP:411
Oxidative Stress
Altered gene expression, NF-kB dependent Interleukin-6 pathway
Altered gene expression, NRF2 dependent antioxidant pathway
reduced dimerization, ARNT/ HIF1-alpha
Decrease, sox9 expression
Increase, Oxidative Stress
Increased, Oxidative Stress
Increased, Reactive metabolites production
Increased, Interleukin-6 production
Cilia Beat Frequency, Decreased
Increased, Motility
Increased, Migration (Endothelial Cells)
reduced production, VEGF
Thyroid hormone synthesis, Decreased
Diminished protective oxidative stress response
Increase, Cell Proliferation
Increased, Reactive oxygen species
Induction, Epithelial Mesenchymal Transition
Apoptosis
Increase, Inflammation
Increased, DNA damage and mutation
Increase, Cell death
Increased, Invasion
Impairment, Endothelial network
Decreased, oligodendrocyte differentiation
Increase, angiogenesis
Increased, Clearance of thyroxine from serum
Hippocampal gene expression, Altered
Hypo-myelination
Increased, extracellular matrix deposition
abnormal, placental labyrinth vasculature morphology
Facial cartilage structures are reduced in size and morphologically distorted
Hippocampal anatomy, Altered
Increased, tumor growth
Thyroxine (T4) in serum, Decreased
Hippocampal Physiology, Altered
Altered, white brain matter
Lung cancer
AOP:416
AOP:417
AOP:420
Lung fibrosis
AOP:414
AOP:415
increase, placental insufficiency
Altered, Cardiovascular development/ function
Thyroxine (T4) in neuronal tissue, Decreased
Mucociliary Clearance, Decreased
N/A, Breast Cancer
AOP:439
AOP:151
Increase, Early Life Stage Mortality
AOP:21
AOP:150
AOP:455
AOP:456
Cognitive Function, Decreased
AOP:300
AOP:458
AOP:459
AOP:488
AOP:411
AOP:418
AOP:419
Decrease, Lung function
increase, Preeclampsia
Molecular Initiating Event (MIE)
Key Event (KE)
Adverse Outcome (AO)
Key events associated with inorganic cadmium

Figure 4.5: Directed network corresponding to the largest component in the undirected cadmium-AOP network comprising 59 KEs and 82 KERs. Among the 59 KEs, 4 are categorized as MIEs (denoted as diamond), 7 are categorized as AOs (denoted as circle), and the remaining 48 are categorized as KEs (denoted as rounded square). The 29 KEs (including MIEs and AOs) associated with inorganic cadmium are marked in 'red'. In this figure, the 59 KEs are arranged vertically according to their level of biological organization.

Among the 30 cadmium-AOPs, we observed 26 cadmium-AOPs have a coverage score ≥ 0.5 signifying that at least half of their KEs have published evidence of being associated with inorganic cadmium (Table 4.1). Notably, 24 of these 26 cadmium-AOPs with high coverage score are identified through our integrative data-centric approach to be relevant for inorganic cadmium-induced toxicity. Further, we observed 9 cadmium-AOPs have 'High' cumulative WoE and 6 cadmium-AOPs have 'Moderate' cumulative WoE, highlighting the significance of the identified AOPs for inorganic cadmium-induced toxicity (Table 4.1). Based on the domain of taxonomic applicability mentioned in AOP-Wiki, we observed 17 out of 30 cadmium-AOPs are applicable across diverse group of species such as humans, animals like rats, mice and chicken, and aquatic species like zebrafish and *Lemna minor* (Supplementary Table S4.8). Furthermore, based on the domain of life-stage applicability mentioned in AOP-Wiki, we find that the 30 cadmium-AOPs capture the potential of inorganic cadmium to induce toxicity in various developmental stages (Supplementary Table S4.8). In addition, this underscores the relevance of the identified cadmium-AOPs in the assessment of inorganic cadmium toxicity in humans and in ecologically relevant species.

Figure 4.4 shows an undirected network representation of the 30 cadmium-AOPs (i.e., cadmium-AOP network) constructed by considering the cadmium-AOPs as nodes and the existence of shared KEs between any two cadmium-AOPs as edges. We found 3 connected components with two or more nodes (labeled C1, C2 and C3) and 4 isolated nodes in the cadmium-AOP network. The connected component C1 is the largest connected component (LCC) comprising 18 cadmium-AOPs, followed by C2 comprising 6 cadmium-AOPs and C3 comprising 2 cadmium-AOPs. We emphasize that only one (AOP:257) of the two cadmium-AOPs linked to inorganic cadmium stressor in AOP-Wiki is part of C3, while the C1 and C2 exclusively comprise cadmium-AOPs identified through our integrative data-centric approach. In other words, all cadmium-AOPs except one that comprise the three clusters C1, C2 and C3 in the cadmium-AOP network were identified to be relevant for inorganic cadmium-induced toxicity by our integrative analysis. Supplementary Table S4.9 provides the list of cadmium-AOPs, and their corresponding MIEs and AOs associated with each of the connected components.

Among the 18 cadmium-AOPs in LCC, 8 AOPs are related with lung diseases (AOP:411, AOP:414, AOP:415, AOP:416, AOP:417, AOP:418, AOP:419 and AOP:420), 4 AOPs with developmental disorders (AOP:21, AOP:150, AOP:455 and AOP:456), another 4 AOPs with cognitive disorders (AOP:300, AOP:458, AOP:459 and AOP:488), 1 AOP with breast cancer (AOP:439), and 1 AOP with pregnancy disorder namely, preeclampsia (AOP:151) (Supplementary Table S4.9). 'Activation, AhR' (KE:18) is the most common MIE in C1, and is shared across 15 AOPs related to all the adverse outcomes (Supplementary Table S4.9). The 6 cadmium-AOPs in C2 (AOP:263, AOP:264, AOP:265, AOP:266, AOP:267, and AOP:268) have 'Decrease, Coupling of oxidative phosphorylation' (KE:1446) as MIE and 'Decrease, Growth' (KE:1521) as AO (Supplementary Table S4.9). Upon closer inspection, we observed that these AOPs are developed by a single research group to address ecotoxicological effects of stressor induced mitochondrial dysfunction on growth and development of organisms. We remark that these 6 cadmium-AOPs, sharing both MIE and AO, can be considered as variations of a single toxicological pathway and represent potential alternate strategies to understand the same process. The 2 cadmium-AOPs in C3 (AOP:257 and AOP:258) have 'Occurrence, Kidney toxicity' (KE:814) as their AO (Supplementary Table S4.9). Upon closer inspection, we observed that AOP:257 is already documented in AOP-Wiki to be linked to inorganic cadmium stressor, and 4 of the 5 KEs in AOP:258 are associated with inorganic cadmium-induced toxicity.

Furthermore, we observed that C1 comprises 59 unique KEs and 82 unique KERs, C2 comprises 11 unique KEs and 15 unique KERs, and C3 comprises 8 unique KEs and 7 unique KERs. Similar to coverage score of AOPs, we computed the coverage score for each of the connected components based on the fraction of KEs associated with inorganic cadmium. We observed that C3 had the highest coverage score of 0.88 with 7 of the 8 KEs associated with inorganic cadmium, which can be attributed to the presence of cadmium stressor linked AOP:257. C1 has a coverage score of 0.49 with 29 of 59 KEs associated with inorganic cadmium, and C2 has a coverage score of 0.45 with 5 of 11 KEs associated with inorganic cadmium.

In sum, the undirected cadmium-AOP network highlighted the connectedness of different AOPs relevant for inorganic cadmium-induced toxicity. Connected component C1 is the largest, has roughly half the KEs associated with inorganic cadmium, and is the most diverse in terms of MIEs and AOs. We therefore considered C1 for further network-based analysis in this study.

4.2.2 Characterization and network-based analysis of the largest component in cadmium-AOP network

A directed network representation of connected AOPs, with nodes as KEs and directed edges as KERs, has the potential to elucidate AOP interactions and reveal connections among toxicity pathways [58]. In this study, we visualized the directed network of 18 cadmium-AOPs in LCC (C1), and analyzed the LCC using network measures to elucidate cadmium-induced toxicity pathways. Notably, all the 18 cadmium-AOPs were identified through our integrative data-centric approach. Furthermore, these 18 cadmium-AOPs comprise AOs that are not linked to inorganic cadmium related stressors in AOP-Wiki.

The LCC C1 comprises 59 unique KEs, of which 4 are MIEs and 7 are AOs (Figure 4.5). We observed that ‘Activation, AhR’ (KE:18) is the most common MIE, and is shared among 15 AOPs (Figure 4.5; Supplementary Table S4.9). ‘Increase, Early Life Stage Mortality’ (KE:947) and ‘Cognitive Function, Decreased’ (KE:402) are the most common AOs, and are shared among 4 AOPs each (Figure 4.5; Supplementary Table S4.9). Besides MIEs and AOs, ‘dimerization, AHR/ARNT’ (KE:944) is the most common KE, and is shared amongst 5 AOPs (Figure 4.5). Furthermore, there are 82 unique KERs in the LCC, of which 8 KERs are labeled as ‘Non-adjacent’ and others are labeled as ‘Adjacent’ (Supplementary Table S4.7).

We characterized the 59 KEs in LCC based on additional information available in AOP-Wiki. We observed that 39 of the 59 KEs are applicable across a diverse range of taxonomies. We also observed that 35 of 59 KEs are applicable across all stages of development. Moreover, based on the KE titles, we observed incompleteness and duplications among the 59 KEs. For instance, the AO 'N/A, Breast Cancer' (KE:1193) has incomplete action information. The 2 KEs, 'Increase, Oxidative stress' (KE:1969) and 'Increased, Oxidative stress' (KE:1088), have the same underlying process and action but are reported as two different KEs in AOP-Wiki. Further, we noted that while AOP-Wiki provides stressor information for each AOP, it does not provide associations between stressors and KEs. Therefore, by following a systematic workflow (Figure 4.2), we identified 29 of the 59 KEs to be associated with inorganic cadmium-induced toxicity by leveraging compiled information in five resources namely, AOP-Wiki, CTD, ToxCast, DEDuCT and NeurotoxKb (Figure 4.5).

Furthermore, we computed different node-centric network measures (in-degree, out-degree, eccentricity, betweenness centrality, convergence and divergence) for the constructed directed network of 18 cadmium-AOPs (Supplementary Table S4.10). We observed that KE 'Altered, Cardiovascular development/function' (KE:317) has the maximum in-degree of 5, while the MIE 'Activation, AhR' (KE:18) has the maximum out-degree of 17. Additionally, based on the in-degree and out-degree values of each KE, we identified 14 convergent (i.e., in-degree > out-degree) KEs and 10 divergent (i.e., in-degree < out-degree) KEs (Supplementary Table S4.10). Among the convergent KEs, we observed that 'Altered, Cardiovascular development/function' (KE:317) has the maximum in-degree value of 5. This KE links different toxicity pathways that originate from activation of AhR and lead to early life-stage mortality (Figure 4.5). The convergent AO, 'Cognitive Function, Decreased', with in-degree value of 4, is the anchor of different toxicity pathways originating from MIEs such as 'Activation, AhR' (KE:18), 'Antagonism, Thyroid Receptor' (KE:1656), and 'Increase, Reactive Oxygen Species production' (KE:257) (Figure 4.5; Supplementary Table S4.10). Among the divergent KEs, ‘Activation, AhR’ (KE:18) with the maximum out-degree of 17, is the origin of different toxicity pathways leading to the 7 AOs ‘Lung cancer’ (KE:1670), ‘Lung fibrosis’ (KE:1276), ‘Decrease, Lung function’ (KE:1250), ‘N/A, Breast Cancer’ (KE:1193), ‘Increase, Preeclampsia’ (KE:1893), ‘Increase, Early Life Stage Mortality’ (KE:947), and ‘Cognitive Function, Decreased’ (KE:402) (Figure 4.5; Supplementary Table S4.10).

Eccentricity of a node indicates the distance of a node to all the other nodes in the network [181]. A large eccentricity value denotes remotely positioned nodes, whereas the low eccentricity value denotes a more centrally positioned node [181]. We computed the eccentricity of each KE present in the directed AOP network, and observed that 2 MIEs namely, ‘Activation, AhR’ (KE:18), ‘Increase, Reactive Oxygen Species production’ (KE:257), and one KE, ‘Induction, CYP1A2/CYP1A5’ (KE:850) have the maximum eccentricity value of 6 (Figure 4.6; Supplementary Table S4.10).

Betweenness centrality of a node indicates the proportion of the shortest paths that pass through it to the total number of shortest paths present between all pairs of nodes excluding that node in the network [59]. We computed the betweenness centrality for each of the 59 KEs in the directed AOP network, and observed that ‘Thyroxine (T4) in serum, Decreased’ (KE:281) has the highest value (Figure 4.7; Supplementary Table S4.10). This KE is present in 2 different toxicity pathways and serves as a critical control event [59] in induced activation of AhR leading to decreased cognitive function.

In sum, the directed AOP network (for LCC) highlighted the diversity of the interconnected inorganic cadmium-induced toxicity pathways. Further, a detailed network analysis highlighted the role of KEs across different toxicity pathways. In the following subsection, we compile auxiliary evidence and provide detailed explanation of novel association of KEs in LCC with inorganic cadmium-induced toxicity.

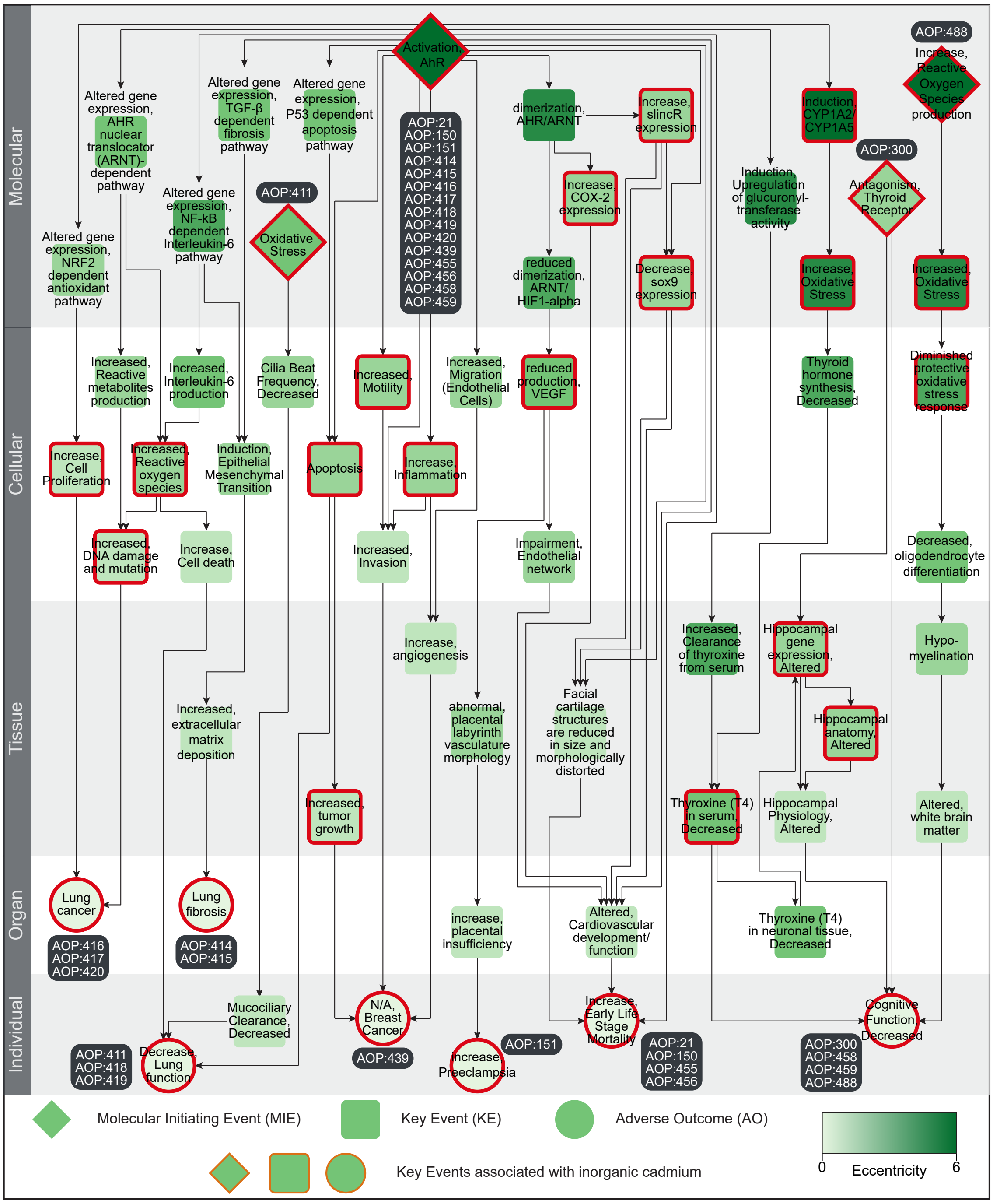

Molecular
Cellular
Tissue
Organ
Individual
Activation, AhR
AOP:21 AOP:150 AOP:151 AOP:414 AOP:415 AOP:416 AOP:417 AOP:418 AOP:419 AOP:420 AOP:439 AOP:455 AOP:456 AOP:458 AOP:459
AOP:488
Increase, Reactive Oxygen Species production
Altered gene expression, AHR nuclear translocator (ARNT)-dependent pathway
Altered gene expression, TGF-β dependent fibrosis pathway
Altered gene expression, P53 dependent apoptosis pathway
dimerization, AHR/ARNT
Increase, slincR expression
Induction, CYP1A2/CYP1A5
AOP:300
Antagonism, Thyroid Receptor
Induction, Upregulation of glucuronyl-transferase activity
Increase, COX-2 expression
AOP:411
Oxidative Stress
Altered gene expression, NF-kB dependent Interleukin-6 pathway
Altered gene expression, NRF2 dependent antioxidant pathway
reduced dimerization, ARNT/ HIF1-alpha
Decrease, sox9 expression
Increase, Oxidative Stress
Increased, Oxidative Stress
Increased, Reactive metabolites production
Increased, Interleukin-6 production
Cilia Beat Frequency, Decreased
Increased, Motility
Increased, Migration (Endothelial Cells)
reduced production, VEGF
Thyroid hormone synthesis, Decreased
Diminished protective oxidative stress response
Increase, Cell Proliferation
Increased, Reactive oxygen species
Induction, Epithelial Mesenchymal Transition
Apoptosis
Increase, Inflammation
Increased, DNA damage and mutation
Increase, Cell death
Increased, Invasion
Impairment, Endothelial network
Decreased, oligodendrocyte differentiation
Increase, angiogenesis
Increased, Clearance of thyroxine from serum
Hippocampal gene expression, Altered
Hypo-myelination
Increased, extracellular matrix deposition
abnormal, placental labyrinth vasculature morphology
Facial cartilage structures are reduced in size and morphologically distorted
Hippocampal anatomy, Altered
Increased, tumor growth
Thyroxine (T4) in serum, Decreased
Hippocampal Physiology, Altered
Altered, white brain matter
Lung cancer
AOP:416 AOP:417 AOP:420
Lung fibrosis
AOP:414 AOP:415
increase, placental insufficiency
Altered, Cardiovascular development/ function
Thyroxine (T4) in neuronal tissue, Decreased
Mucociliary Clearance, Decreased
N/A, Breast Cancer
AOP:439
AOP:151
Increase, Early Life Stage Mortality
Cognitive Function, Decreased
AOP:411 AOP:418 AOP:419
Decrease, Lung function
Increase, Preeclampsia
AOP:21 AOP:150 AOP:455 AOP:456
AOP:300 AOP:458 AOP:459 AOP:488
Molecular Initiating Event (MIE)
Key Event (KE)
Adverse Outcome (AO)
Key Events associated with inorganic cadmium
0
Eccentricity
6

Figure 4.6: Directed network corresponding to the LCC (C1) in the cadmium-AOP network, where the KEs (including MIEs and AOs) are colored based on their eccentricity values. The 29 KEs (including MIEs and AOs) associated with inorganic cadmium are marked in ‘red’. In this figure, the 59 KEs are arranged vertically according to their level of biological organization.

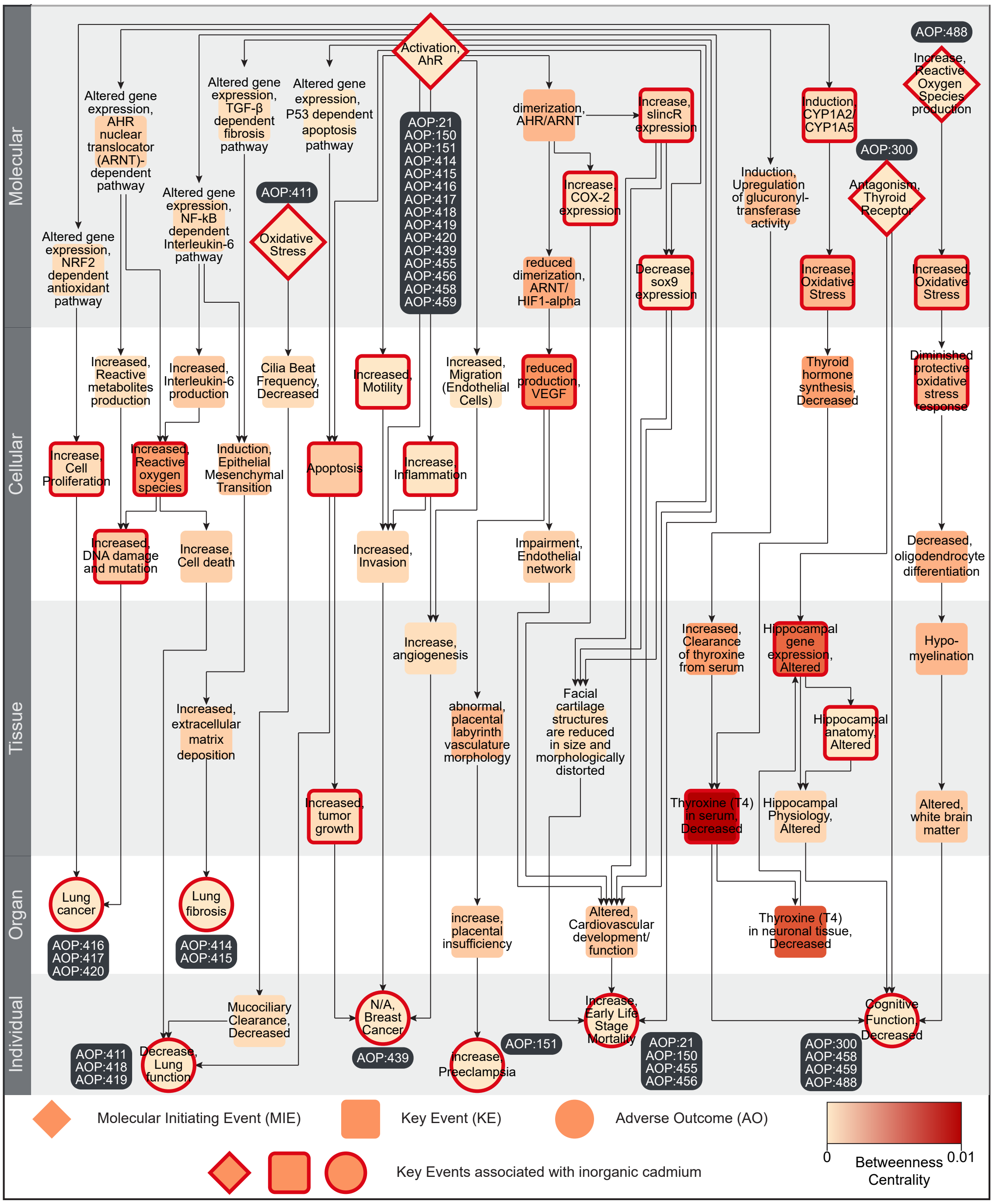

Molecular
Cellular
Tissue
Organ
Individual
Activation, AhR
AOP:21 AOP:150 AOP:151 AOP:414 AOP:415 AOP:416 AOP:417 AOP:418 AOP:419 AOP:420 AOP:439 AOP:455 AOP:456 AOP:458 AOP:459
Altered gene expression, AHR nuclear translocator (ARNT)-dependent pathway
Altered gene expression, TGF-β dependent fibrosis pathway
Altered gene expression, P53 dependent apoptosis pathway
Altered gene expression, NRF2 dependent antioxidant pathway
Altered gene expression, NF-kB dependent Interleukin-6 pathway
AOP:411
Oxidative Stress
dimerization, AHR/ARNT
Increase, slincR expression
Increase, COX-2 expression
reduced dimerization, ARNT/HIF1-alpha
Decrease, sox9 expression
Induction, Upregulation of glucuronyl-transferase activity
Induction, CYP1A2/CYP1A5
AOP:300
Antagonism, Thyroid Receptor
AOP:488
Increase, Reactive Oxygen Species production
Increase, Oxidative Stress
Increased, Oxidative Stress
Increased, Reactive metabolites production
Increased, Interleukin-6 production
Cilia Beat Frequency, Decreased
Increased, Motility
Increased, Migration (Endothelial Cells)
reduced production, VEGF
Thyroid hormone synthesis, Decreased
Diminished protective oxidative stress response
Increase Cell Proliferation
Increased, Reactive oxygen species
Induction, Epithelial Mesenchymal Transition
Apoptosis
Increase, Inflammation
Increased, DNA damage and mutation
Increase, Cell death
Increased, Invasion
Impairment, Endothelial network
Decreased, oligodendrocyte differentiation
Increase, angiogenesis
Increased, Clearance of thyroxine from serum
Hippocampal gene expression, Altered
Hypo-myelination
Increased, extracellular matrix deposition
abnormal, placental labyrinth vasculature morphology
Facial cartilage structures are reduced in size and morphologically distorted
Hippocampal anatomy, Altered
Increased, tumor growth
Thyroxine (T4) in serum, Decreased
Hippocampal Physiology, Altered
Altered, white brain matter
Lung cancer
AOP:416 AOP:417 AOP:420
Lung fibrosis
AOP:414 AOP:415
increase, placental insufficiency
Altered, Cardiovascular development/function
Thyroxine (T4) in neuronal tissue, Decreased
Mucociliary Clearance, Decreased
N/A, Breast Cancer
AOP:439
AOP:151
Increase, Early Life Stage Mortality
Cognitive Function, Decreased
AOP:411 AOP:418 AOP:419
Decrease, Lung function
increase Preeclampsia
AOP:21 AOP:150 AOP:455 AOP:456
AOP:300 AOP:458 AOP:459 AOP:488
Molecular Initiating Event (MIE)
Key Event (KE)
Adverse Outcome (AO)
0
0.01
Betweenness Centrality
Key Events associated with inorganic cadmium

Figure 4.7: Directed network corresponding to the LCC (C1) in the cadmium-AOP network, where the KEs (including MIEs and AOs) are colored based on their betweenness centrality values. The 29 KEs (including MIEs and AOs) associated with inorganic cadmium are marked in 'red'. In this figure, the 59 KEs are arranged vertically according to their level of biological organization.

4.2.3 Auxiliary evidence for cadmium-induced toxicity pathways in the largest component

In this study, we systematically integrated heterogeneous datasets from AOP-Wiki, CTD, ToxCast, DEDuCT and NeurotoxKb to identify KEs associated with inorganic cadmium. This data-centric approach enabled us to identify 29 of the 59 KEs present in the directed network of LCC (C1) to be associated with inorganic cadmium (Figure 4.5). Notably, the 29 KEs associated with inorganic cadmium comprise 4 MIEs and 7 AOs (Figure 4.5). The toxicity pathway originating from MIE 'Activation, AhR' (KE:18), passing through KEs 'Apoptosis' (KE:1262) and 'Increased, tumor growth' (KE:1971), and eventually terminating at AO 'N/A, Breast Cancer' (KE:1193) contains 4 of the 29 KEs associated with inorganic cadmium. Additionally, we noted that this toxicity pathway is part of AOP 'Activation of the AhR leading to breast cancer' (AOP:439), which was systematically developed through extensive literature review [182]. Moreover, AOP:439 has a cumulative WoE of 'High' (Table 4.1) and is applicable to human adults with high level of evidence (Supplementary Table S4.8). Therefore, the pathway originating from activation of AhR and terminating in breast cancer is a potential toxicity pathway of cadmium-induced breast cancer outcome in humans.

Subsequently, to assess the rationality of associations between the remaining 30 KEs in LCC and inorganic cadmium, we relied on the toxicity data in published literature. We leveraged an artificial intelligence (AI) based tool, AOP-helpFinder [183, 184] to screen existing literature and identify associations of KEs with inorganic cadmium. In addition, we relied on Abstract Sifter [185] to filter published literature from PubMed [186] that are relevant for cadmium-induced toxicity. This extensive literature curation helped us identify novel associations between the remaining 30 KEs in LCC and inorganic cadmium (Supplementary Table S4.11). We also identified auxiliary evidence for the 29 KEs in LCC that were already associated with inorganic cadmium through our data-centric approach (Supplementary Table S4.11). Additionally, we curated chemical name, study type, and dosage information from these auxiliary evidences (Supplementary Table S4.11).

To conclude, we performed two case studies pertaining to a human relevant AOP and ecotoxicity relevant AOPs in LCC to explore the rationale and highlight the relevance of the directed AOP network for cadmium-induced toxicity.

AOP linking inorganic cadmium exposure to preeclampsia

Preeclampsia is a chronic human pregnancy complication, and is emerging as a leading cause of neonatal mortality [187]. We noted that a preeclampsia specific AOP in AOP-Wiki, 'AhR activation leading to preeclampsia' (AOP:151), is identified by this study as cadmium-AOP (Table 4.1) and is part of LCC (C1) (Supplementary Table S4.9). According to AOP-Wiki, AOP:151 is currently under development, but is included in the OECD work plan (Supplementary Table S4.1). Different published studies have found significant correlation between environmental exposure to cadmium and preeclampsia in pregnant women [188, 189]. Therefore, we leveraged AOP:151 to explore and verify the rationale behind the cadmium-induced toxicity in preeclampsia.

It has been shown that cadmium exposure causes modulation in AhR downstream genes through cross-talk between AhR and estrogen receptors in rat uterine tissue [190]. AhR is a cytosolic protein, which upon binding with a ligand relocates into the nucleus, where it dimerizes with aryl hydrocarbon receptor nuclear translocator (ARNT) to transcribe its downstream genes [191]. Simultaneously, cadmium has been observed to degrade the activity of hypoxia inducible factor 1 (HIF-1) protein, thereby hindering the dimerization of ARNT with HIF-1 [192]. HIF-1 is a master regulator of hypoxia induced responses, and upon dimerization, induces downstream genes such as vascular endothelial growth factor (VEGF) that enables angiogenesis [193, 194]. It has been observed that cadmium exposure leads to reduction in VEGF levels of human placental trophoblasts [195] and in human umbilical vein endothelial cells (HUVEC) [196]. Several *in vivo* experiments showed that cadmium exposed pregnant rats have reduced levels of vasculature in placenta, resulting in placental insufficiency [195, 197–199]. Alternatively, cadmium exposure has also been seen to cause placental insufficiency by inducing oxidative stress in placenta [200]. Such abrupt vascularization ultimately results in showing key features of preeclampsia in both pregnant rats [201] and in human cell line studies [202]. In conclusion, by leveraging published evidence of cadmium-induced toxicity, we were able to explore a potential toxicity pathway in AOP:151 that links cadmium exposure to preeclampsia.

AOPs linking inorganic cadmium exposure to aquatic ecotoxicity

Aquatic ecotoxicity is of primary regulatory concern as it is one of the major determinants in the well-being of terrestrial and aquatic species alike [203]. We identified 2 cadmium-AOPs namely 'Aryl hydrocarbon receptor activation leading to early life stage mortality, via increased COX-2' (AOP:21) and 'Aryl hydrocarbon receptor activation leading to early life stage mortality, via reduced VEGF' (AOP:150), that are part of LCC (C1), and have 'High' evidence of applicability in aquatic species (Supplementary Table S4.8). Moreover, these AOPs have a cumulative WoE of 'High' and are endorsed by Working Group of the National Coordinators of the Test Guidelines Programme (WNT) and the Working Party on Hazard Assessment (WPHA) under the OECD AOP development programme (Supplementary Table S4.1). Additionally, these 2 AOPs share the same MIE ('Activation, AhR') and AO ('Increase, Early Life Stage Mortality'). AOP:150 also shares 4 of its KEs (including MIE) with AOP:151 (Figure 4.5). Therefore, we leveraged these 2 AOPs to explore and verify the rationale behind cadmium toxicity in aquatic ecosystems.

Zebrafish larvae showed dose-dependent response to cadmium toxicity through the upregulation of AhR downstream genes [204]. In zebrafish, the AhR gets activated upon being bound to a ligand, and is transported into the nucleus where it dimerizes with ARNT to enable the transcription of downstream genes [205]. One such group of downstream genes, cyclooxygenase-2 (COX-2) has been observed to be upregulated in common carp spleens upon exposure to inorganic cadmium [138]. Alternatively, cadmium exposure affects HIF-1 activity through AhR mediated pathways, and this results in reduced levels of VEGF [192, 195, 196]. Various *in vitro* experiments showed that cadmium exposure impairs endothelial cell function and promotes their apoptosis [206–208]. Consequently, cadmium exposure has also been observed to induce cardiovascular developmental disorders by hindering the process of cardiomyocyte differentiation [209, 210]. Ultimately, cadmium exposure has been observed to promote early life-stage mortality in aquatic species by hindering their developmental processes [204, 211].

4.3 Discussion

Cadmium, a heavy metal, is considered to be a priority environmental pollutant due to its abundance and considerable toxicity to humans and aquatic species. In the past, the concept of AOP network had enabled elucidation of complex toxicity pathways and aided in regulatory decision making. To this end, we present an integrative data-centric approach for derivation and characterization of the AOP network relevant to cadmium-induced toxicity (Figure 4.8). We describe a detailed computational workflow to curate high quality and complete AOPs within AOP-Wiki. Further, by systematically integrating heterogeneous data from different exposome-relevant databases, we uncover novel associations between the inorganic cadmium and the existing AOPs (Figure 4.8). Notably, our integrative data-centric approach revealed 28 novel cadmium-AOPs associated with various adverse outcomes such as pulmonary disorders, reproductive and developmental disorders, breast cancer and cognitive disorders which were not otherwise linked to cadmium stressor in AOP-Wiki. Importantly, this study also highlights the use of different AOP networks namely, an undirected network of cadmium-AOPs and a directed AOP network of cadmium-AOPs by utilizing their KE and KER information to explore cadmium-induced toxicities in human and aquatic species.

However, among the curated list of cadmium-AOPs, we observed that many are still under development. Consequently, the information associated with these AOPs may be

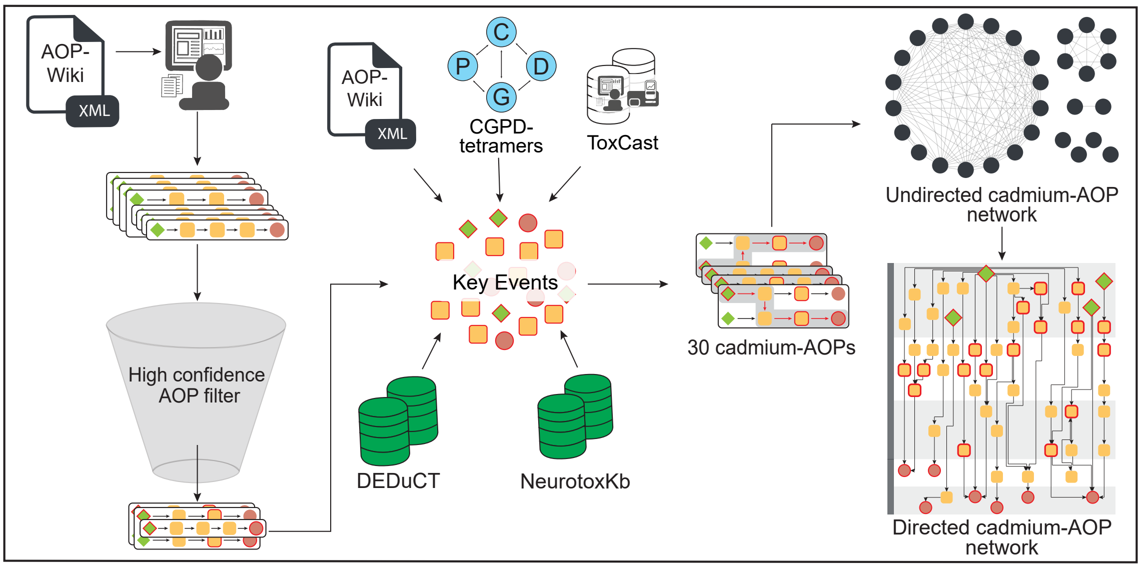


Figure 4.8: Schematic summary of the data-centric approach for derivation and characterization of the AOP network relevant to cadmium-induced toxicity.

incomplete, and importantly, we noted inconsistencies in the KE information provided by AOP-Wiki. We also observed that AOP-Wiki does not exhaustively capture all possible adverse outcomes induced by inorganic cadmium which were otherwise revealed by the CGPD-tetramers constructed from CTD data.

Nonetheless, we present the first AOP network relevant to cadmium-induced toxicity by integrating heterogeneous data from different resources. Moreover, the compiled dosage information might provide empirical support for the dose-response relationship and thereby enable the development of quantitative AOPs that will aid in regulatory decision-making with respect to cadmium-induced toxicities [212, 213]. We expect that the observations from this study will aid in regulation of cadmium and its inorganic compounds in future.

Supplementary Information

Supplementary Tables S4.1-S4.11 associated with this chapter are available for download from the GitHub repository: `https://github.com/asamallab/PhDThesis-Ajaya_Kumar_Sahoo/blob/main/SI/ST_Chapter4.xlsx`.

Code Availability

The computer programs used to perform the computations reported in this chapter are available in the following GitHub repository:
https://github.com/asamallab/PhDThesis-Ajaya_Kumar_Sahoo/tree/main/Codes.

Serial number	AOP identifier	AOP title	Coverage score	Cumulative WoE
1	19	Androgen receptor antagonism leading to adverse effects in the male foetus (mammals)	0.6	Not Specified
2	21	Aryl hydrocarbon receptor activation leading to early life stage mortality, via increased COX-2	0.6	High
3	69	Modulation of Adult Leydig Cell Function Subsequent to Decreased Cholesterol Synthesis or Transport in the Adult Leydig Cell	0.6	Not Specified
4	150	Aryl hydrocarbon receptor activation leading to early life stage mortality, via reduced VEGF	0.43	High
5	151	AhR activation leading to preeclampsia	0.43	Not Specified
6	257	Receptor mediated endocytosis and lysosomal overload leading to kidney toxicity	1	High
7	258	Renal protein alkylation leading to kidney toxicity	0.8	High
8	263	Uncoupling of oxidative phosphorylation leading to growth inhibition via decreased cell proliferation	1	Moderate
9	264	Uncoupling of oxidative phosphorylation leading to growth inhibition via ATP depletion associated cell death	1	Moderate
10	265	Uncoupling of oxidative phosphorylation leading to growth inhibition via increased cytosolic calcium	0.75	Moderate
11	266	Uncoupling of oxidative phosphorylation leading to growth inhibition via decreased Na-K ATPase activity	0.67	Not Specified
12	267	Uncoupling of oxidative phosphorylation leading to growth inhibition via glucose depletion	0.6	Not Specified
13	268	Uncoupling of oxidative phosphorylation leading to growth inhibition via mitochondrial swelling	0.75	Not Specified
14	296	Oxidative DNA damage leading to chromosomal aberrations and mutations	1	High
15	300	Thyroid Receptor Antagonism and Subsequent Adverse Neurodevelopmental Outcomes in Mammals	0.8	Moderate
16	392	Decreased fibrinolysis and activated bradykinin system leading to hyperinflammation	0.8	Not Specified
17	411	Oxidative stress Leading to Decreased Lung Function	0.5	High
18	414	Aryl hydrocarbon receptor activation leading to lung fibrosis through TGF-β dependent fibrosis toxicity pathway	0.4	Not Specified
19	415	Aryl hydrocarbon receptor activation leading to lung fibrosis through IL-6 toxicity pathway	0.4	Not Specified
20	416	Aryl hydrocarbon receptor activation leading to lung cancer through IL-6 toxicity pathway	0.67	Not Specified

21	417	Aryl hydrocarbon receptor activation leading to lung cancer through AHR-ARNT toxicity pathway	0.6	Not Specified
22	418	Aryl hydrocarbon receptor activation leading to impaired lung function through AHR-ARNT toxicity pathway	0.6	Not Specified
23	419	Aryl hydrocarbon receptor activation leading to impaired lung function through P53 toxicity pathway	0.75	Not Specified
24	420	Aryl hydrocarbon receptor activation leading to lung cancer through sustained NRF2 toxicity pathway	0.75	Not Specified
25	439	Activation of the AhR leading to breast cancer	0.67	High
26	455	Aryl hydrocarbon receptor activation leading to early life stage mortality via impeded craniofacial development	0.67	Moderate
27	456	Aryl hydrocarbon receptor activation leading to early life stage mortality via cardiovascular toxicity	0.67	High
28	458	AhR activation in the liver leading to Subsequent Adverse Neurodevelopmental Outcomes in Mammals	0.5	High
29	459	AhR activation in the thyroid leading to Subsequent Adverse Neurodevelopmental Outcomes in Mammals	0.67	Moderate
30	488	Increased reactive oxygen species production leading to decreased cognitive function	0.57	Not Specified

Table 4.1: The curated list of 30 cadmium-AOPs and their corresponding AOP identifiers, AOP titles, computed coverage scores, and cumulative WoE.

Chapter 5

Leveraging integrative toxicogenomic approach towards development of stressor-centric adverse outcome pathway networks for plastic additives

Plastics are the most widely produced synthetic chemicals, roughly constituting about 10% of solid waste generated globally [19, 214]. Extensive usage followed by improper waste management of plastics have made them ubiquitous pollutants in atmosphere, terrestrial and aquatic environments [215–217]. Plastics comprise various chemicals including polymers, solvents, additives and unintentional chemical residues resulting from the manufacturing process [19]. In particular, additives are chemicals that are intentionally added during the plastic manufacturing process to achieve specific desirable properties such as flexibility, reduced flammability, pigmentation, and make up nearly 50% by weight of the plastics [20, 218, 219]. These plastic additives are not covalently bonded to plastic, and thus can be potentially released into the environment throughout the plastic life cycle [20, 21, 220]. Environmental exposure to such plastic additives has been observed to elicit various adverse health effects such as cancer, developmental defects, endocrine disruptions, and metabolic disruptions in humans and other species alike [6, 21, 221–223], but the lack of information on their presence throughout the plastics life cycle hampered their risk assessment and eventually the product safety [20]. Additionally, these plastic additives can persist in the environment, bioaccumulate in various organisms, and have long-lasting ecological impacts [224, 225]. Therefore, it is imperative to identify these plastic additives and perform their risk assessment to achieve a toxic-free circular economy for plastics.

Previously, Aguayo-Orozco *et al.* [93] had utilized biological endpoint data of chemicals screened through several high throughout toxicity assays in ToxCast [16] to construct stressor-adverse outcome pathway (AOP) network linking chemicals in ToxCast to several developed AOPs within AOP-Wiki [53]. Such a construction enabled exploration of the adverse effects associated with this chemical space from a mechanistic perspective [93]. Furthermore, a data integrative approach, similar to the study reported in Chapter 4, can help identify AOPs within AOP-Wiki that are relevant for plastic additives-induced toxicity. Such plastic additive-AOP associations can aid in the development of stressor-centric AOP network for plastic additives that will provide a holistic view of plastic additives-induced adverse effects.

In Chapter 4, we discussed a data integrative approach to derivation and characterization of AOP network relevant to a single chemical stressor namely, inorganic cadmium. In this chapter, we extend our efforts to linking plastic additives (multiple stressors) with the AOPs compiled within AOP-Wiki. We systematically curate a list of plastic additives from chemicals documented to be found in plastics in a published report. Notably, we employ a toxicogenomic approach and integrate heterogeneous biological endpoint data from various exposome-relevant resources to develop stressor-centric AOP networks for plastic additives, thereby facilitating the exploration of their toxicities. **The work reported in this chapter is contained in the published manuscript [96].**

5.1 Methods

5.1.1 Compilation and curation of plastic additives

Recently, the United Nations Environment Programme (UNEP) published a report titled 'Chemicals in Plastics – A Technical Report' [20] that provides an annex cataloging over 13000 chemicals found in plastics and plastic manufacturing processes that were systematically curated by Aurisano *et al.* [226] and Wiesinger *et al.* [227]. Among the various chemicals in plastics, the compounds termed as plastic additives define the desirable properties in the final plastic product [20]. Plastic additives are intentionally added, constituting anywhere between 4%-50% by weight in plastics [20], and can potentially leach into the environment as they are not covalently bonded to the plastic polymers, thus posing a risk to human health and environment [21]. Here, we relied on the annex provided by the UNEP report to identify the different plastic additives (Figure 5.1).

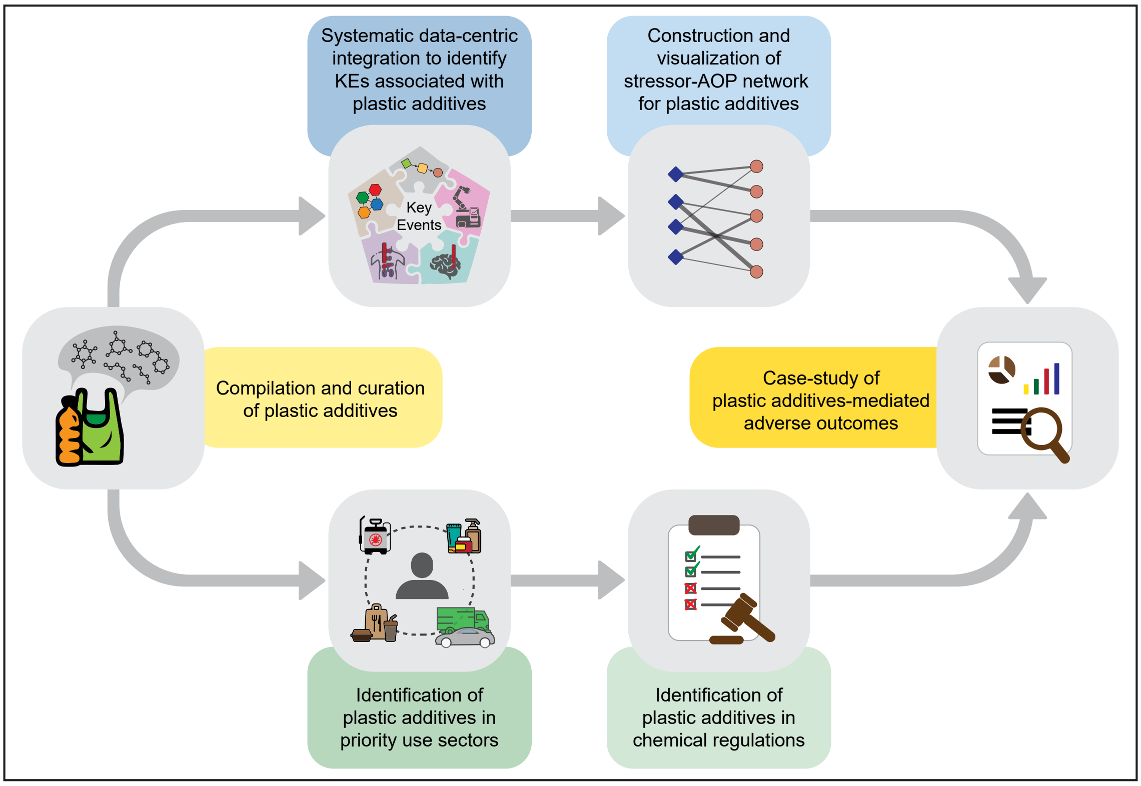

Systematic data-centric integration to identify KEs associated with plastic additives
Construction and visualization of stressor-AOP network for plastic additives
Key Events
Compilation and curation of plastic additives
Case-study of plastic additives-mediated adverse outcomes
Identification of plastic additives in priority use sectors
Identification of plastic additives in chemical regulations

Figure 5.1: Summary of the workflow followed to identify plastic additives from chemicals found in plastics, followed by the exploration of their toxicity pathways through the construction of stressor-centric AOP networks.

We first compiled the chemicals and their corresponding Chemical Abstracts Service (CAS) registry numbers provided by the UNEP report. Thereafter, we relied on the CAS common chemistry web portal [228] to identify synonymous CAS registry numbers and mapped them to their latest identifiers to remove redundancy and duplications in chemical identifiers listed in the UNEP report. In case the CAS identifier provided by the UNEP report is not present in the portal, we used the identifier provided by the UNEP report, and finally compiled 13640 unique chemicals.

Next, we observed that various terms were used in the UNEP report to identify functions of different chemicals in plastics. Notably, some of these chemicals were annotated only as non-intentionally added substances (NIAS), which we excluded from our analysis. Thereafter, we standardized the vocabulary of the associated functions by relying on various published sources [19, 20, 218, 229–232]. To identify the functions associated with plastic additives, we relied on several published sources and documents [19, 20, 218, 229–234]. The details of the curated functions associated with plastic additives, and their descriptions is provided in Supplementary Table S5.1. We consider a chemical as a plastic additive if it has at least one annotated function that is associated with plastic additives. Through this extensive manual effort, we finally curated a list of 6470 plastic additives (Supplementary Table S5.2) from the chemicals compiled in the UNEP report, and leveraged them for further analysis. Figure 5.2 illustrates the steps taken to curate a list of 6470 plastic additives from chemicals documented to be found in plastics.

5.1.2 Compilation of AOPs within AOP-Wiki

The AOP-Wiki [53] is the largest publicly accessible repository, hosted by the Society for the Advancement of Adverse Outcome Pathways (SAAOP), which compiles and organizes various AOPs developed globally. In order to access the latest information available within AOP-Wiki, we downloaded the XML file (released on 1 January 2024) from ‘Project Downloads’ page in AOP-Wiki. Then, we utilized an in-house python script

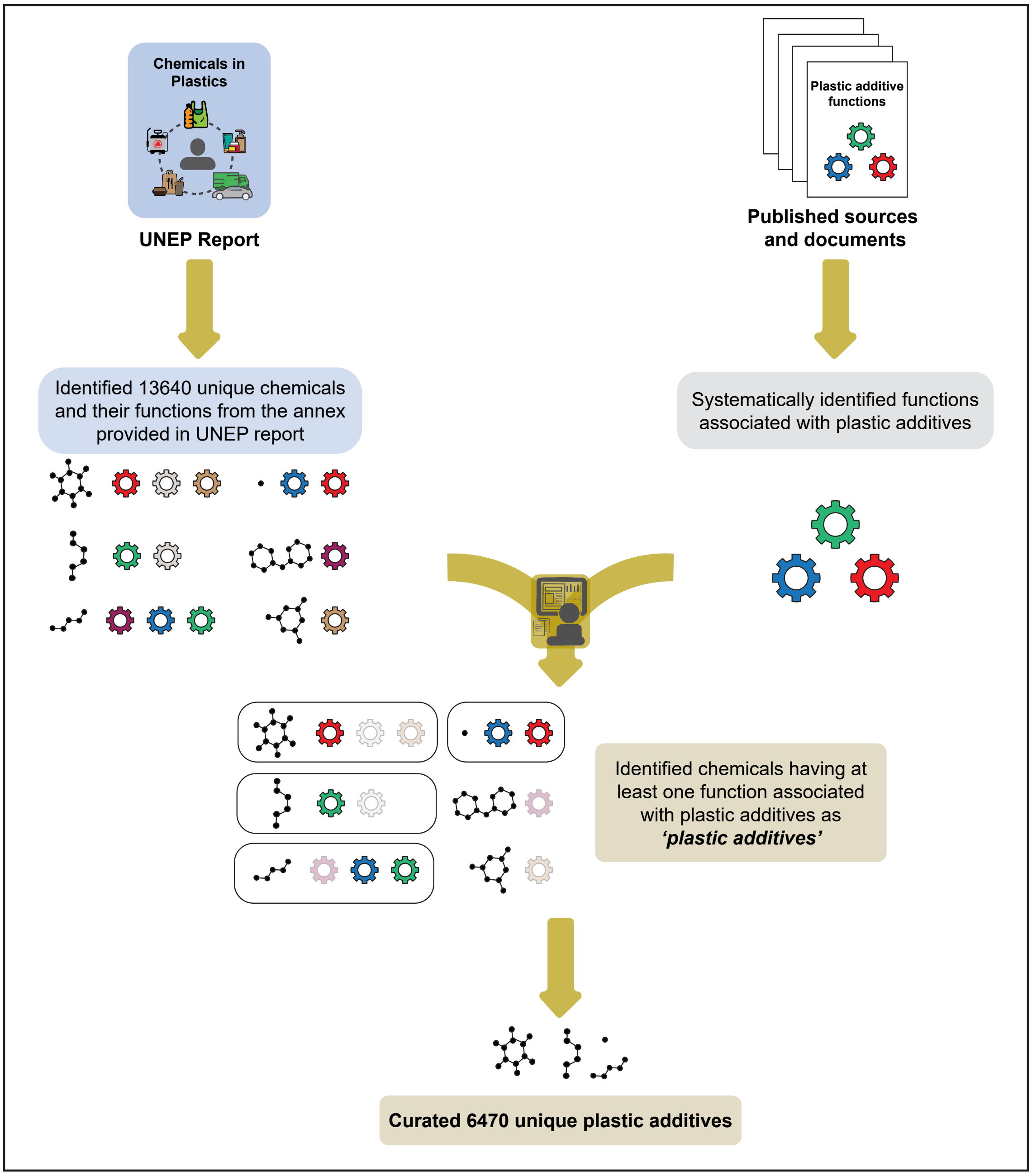

Chemicals in Plastics
UNEP Report
Plastic additive functions
Published sources and documents
Identified 13640 unique chemicals and their functions from the annex provided in UNEP report
Systematically identified functions associated with plastic additives
Identified chemicals having at least one function associated with plastic additives as 'plastic additives'
Curated 6470 unique plastic additives

Figure 5.2: Workflow to identify 6470 unique plastic additives from chemicals documented in the UNEP report.

to parse the XML file and extract various information associated with AOPs like AOP identifier, AOP title, associated key events (KEs) (molecular initiating events - MIEs and adverse outcomes - AOs) and key event relationships (KERs), linked stressors, status according to Organisation for Economic Co-operation and Development (OECD) and SAAOP, and biological applicability information such as taxonomy, sex and life-stage of the organism, and their corresponding weight of evidence (WoE). Additionally, we also extracted information associated with KEs like KE title, KE identifier, level of biological organization, action name, object name, object identifiers and process name, and information associated with KERs like upstream/downstream KEs, evidence for biological plausibility of KER, adjacency, and the extent of quantitative understanding of KER.

5.1.3 Identification of 'high confidence AOPs' within AOP-Wiki

Within AOP-Wiki, the AOPs are continuously updated based on current understanding and availability of novel experimental data, and thus, the AOPs are living documents [171]. Therefore, we relied on a systematic workflow developed in our previous work [95], to filter high quality and complete AOPs within AOP-Wiki (Figure 5.3). First, we filtered out AOPs with SAAOP status as 'archived'. Next, we manually checked and removed AOPs that have KE title as 'unknown' or lacked any KEs or KERs (Figure 5.3). Next, we checked for the presence of disconnected components in AOPs using NetworkX library [173] in python, and manually updated and filtered out AOPs that contained disconnected components. Lastly, we checked the remaining AOPs for presence of MIEs, AOs and a directed path between MIE and AO, and filtered out AOPs that did not contain any such path (Figure 5.3). This combined computational and manual effort led to the identification of 328 complete, connected and high quality AOPs within AOP-Wiki (last accessed on 15 February 2024) which we designate as 'high confidence AOPs' (Figure 5.3; Supplementary Table S5.3). The 328 high confidence AOPs comprise 1107 unique KEs (Supplementary Table S5.4) and 1717 unique KERs (Supplementary Table S5.5).

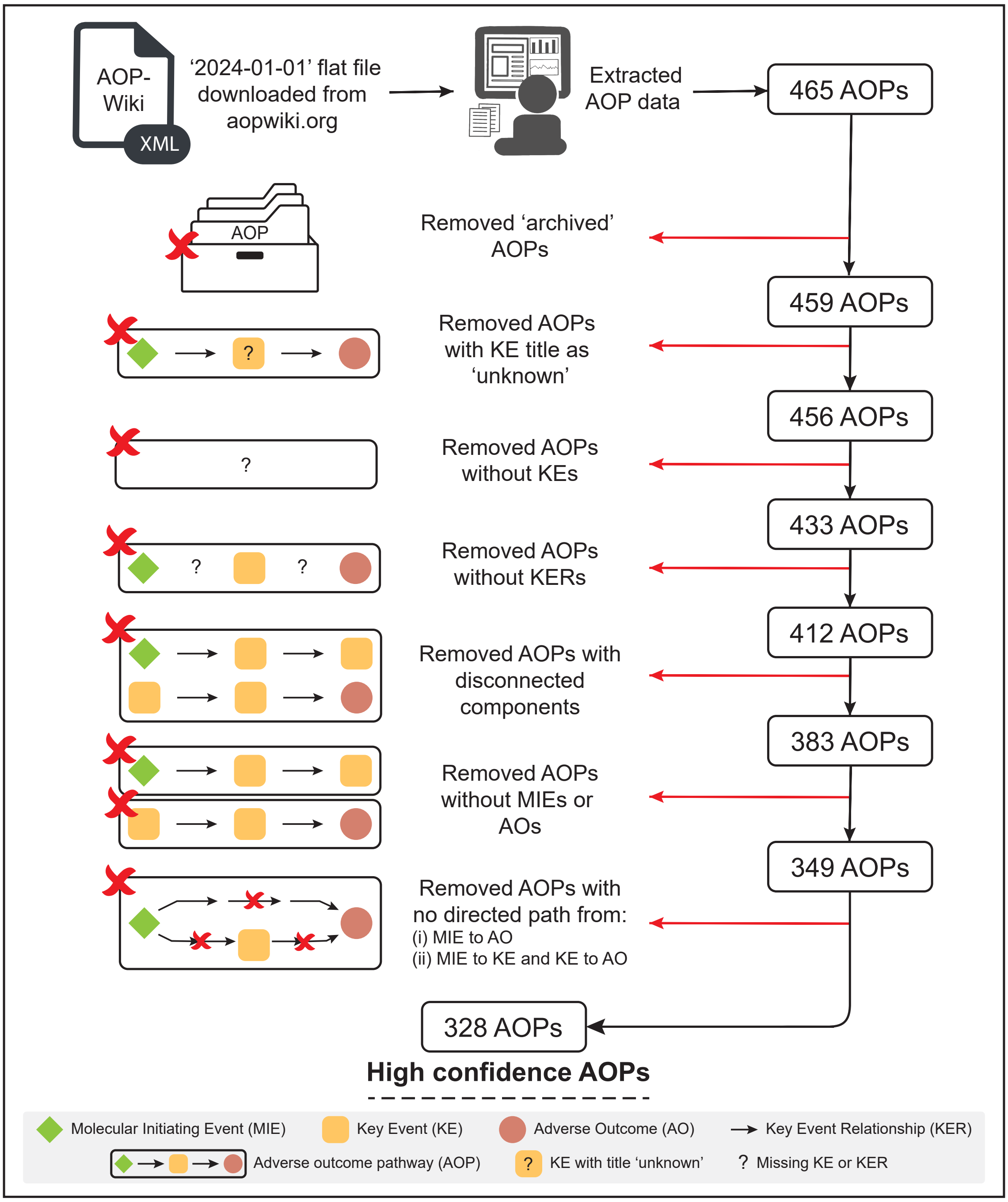

AOP-Wiki
XML
'2024-01-01' flat file downloaded from aopwiki.org
Extracted AOP data
465 AOPs
AOP
Removed 'archived' AOPs
459 AOPs
Removed AOPs with KE title as 'unknown'
456 AOPs
Removed AOPs without KEs
433 AOPs
Removed AOPs without KERs
412 AOPs
Removed AOPs with disconnected components
383 AOPs
Removed AOPs without MIEs or AOs
349 AOPs
Removed AOPs with no directed path from:
(i) MIE to AO
(ii) MIE to KE and KE to AO
328 AOPs
High confidence AOPs
Molecular Initiating Event (MIE)
Key Event (KE)
Adverse Outcome (AO)
Key Event Relationship (KER)
Adverse outcome pathway (AOP)
KE with title 'unknown'
? Missing KE or KER

Figure 5.3: Workflow to filter high confidence AOPs from AOP-Wiki by employing computation and manual curation in conjunction.

5.1.4 Identification of KEs associated with plastic additives

Based on our previous work [95], we relied on a systematic and comprehensive data-centric integration method to identify the KEs within AOP-Wiki that are associated with plastic additives by utilizing toxicogenomics and biological endpoints data from five exposome-relevant resources: ToxCast [16], Comparative Toxicogenomics Database (CTD) [27], DEDuCT [8, 15], NeurotoxKb [29] and AOP-Wiki [53].

Using ToxCast

US EPA's ToxCast program provides high throughput *in vitro* bioactivity assay data for thousands of chemicals tested across several assays [16]. Importantly, the ToxCast data includes assay annotations and information on associated bioprocess and genes that can aid in the identification of KEs (specifically MIEs) associated with the corresponding active chemical [58, 93, 95, 235]. First, we downloaded the latest ToxCast invitrodb version 4.1 dataset from the US EPA repository [178]. Next, we retrieved chemicals and their corresponding assay endpoints from the
'mc5-6_winning_model_fits-flags_invitrodb_v4_1_SEPT2023.csv' file and filtered chemicals with active assay endpoints ('hitc' ≥ 0.9) [179]. Furthermore, we retrieved the 'activatory' or 'inhibitory' response of these active chemicals by relying on the 'top' value of the corresponding winning model from the
'mc4_all_model_fits_invitrodb_v4_1_SEPT2023.csv' file [179].

Sometimes, chemicals exhibit their activity in a narrow range of concentrations that coincides with that of cell stress and cytotoxicity, thereby leading to non-specific activation of reporter genes. Such phenomena are termed as 'cytotoxicity-associated bursts' and can lead to inaccurate assay endpoint readings [236]. Therefore, in this study, we identified such cytotoxicity-associated bursts for plastic additives tested within ToxCast, and did not consider those endpoints for mapping with KEs within AOP-Wiki (Figure 5.4).

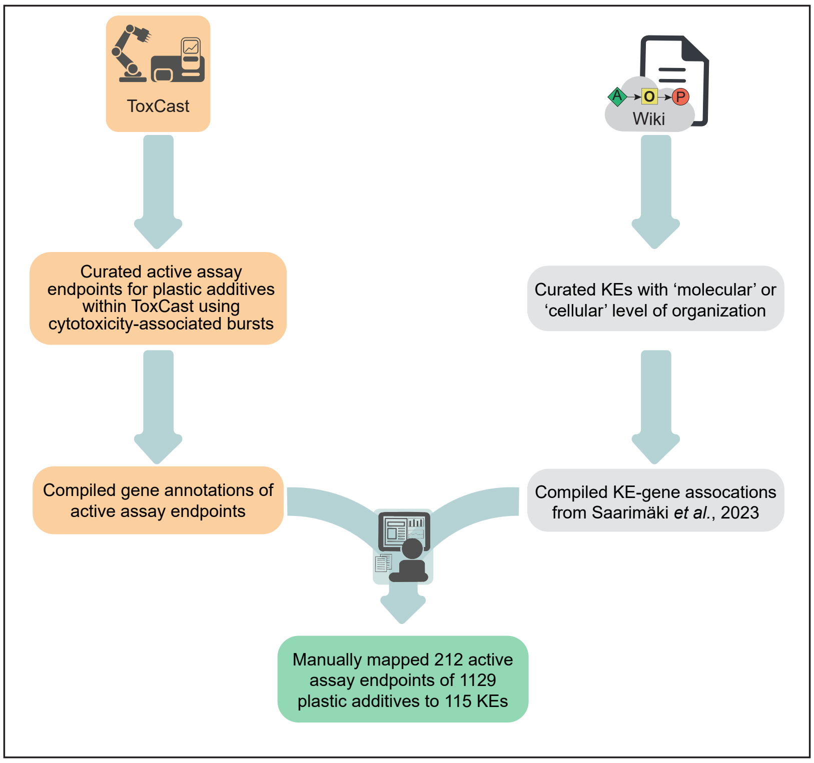

ToxCast
A
O
P
Wiki
Curated active assay endpoints for plastic additives within ToxCast using cytotoxicity-associated bursts
Curated KEs with 'molecular' or 'cellular' level of organization
Compiled gene annotations of active assay endpoints
Compiled KE-gene assocations from Saarimäki *et al.*, 2023
Manually mapped 212 active assay endpoints of 1129 plastic additives to 115 KEs

Figure 5.4: Workflow to identify KEs from AOP-Wiki which are mapped to the active assay endpoints of plastic additives within ToxCast.

To identify cytotoxicity-associated bursts within ToxCast, Judson *et al.* [236] proposed the following Z-score metric:

$$Z(\text{chemical, assay}) = \frac{-\log AC_{50}(\text{chemical, assay}) - median\left[-\log AC_{50}(\text{chemical, cytotox})\right]}{\text{global cytotoxicity MAD}}$$

wherein, 'logAC$_{50}$(chemical, assay)' is the logarithm of the AC$_{50}$ value of the chemical in the assay, 'logAC$_{50}$(chemical, cytotox)' is the logarithm of the AC$_{50}$ value of the chemical in the corresponding cytotoxicity assay and the 'global cytotoxicity MAD' is the median of the MAD (median average deviations) of the logAC$_{50}$(chemical, cytotox) distributions across all chemicals [236]. For a given chemical, assays having Z-score values lying between +3 and -3 were considered as cytotoxicity-associated bursts [236]. Here, we relied on this Z-score metric by Judson *et al.* to identify cytotoxicity-associated bursts corresponding to the plastic additives tested within ToxCast.

The 'cytotox_invitrodb_v4_1_SEPT2023.xlsx' file in ToxCast invitrodb version 4.1

provides the global cytotoxicity MAD and $\log AC_{50}(\text{chemical, cytotox})$ ('cytotox_median_log') for chemicals across various cytotoxicity assays [179]. We retrieved these values for the plastic additives tested within ToxCast, computed their Z-scores and discarded assays that had a Z-score value lying between +3 and -3. Through this process, we identified 1108 assay endpoints associated with 1327 plastic additives, and proceeded to map them to KEs within AOP-Wiki (Figure 5.4).

First, we retrieved the genes associated with these 1108 assay endpoints from the 'assay_annotations_invitrodb_v4_1_SEPT2023.xlsx' file and the details of assay endpoint-gene mappings from 'assay_gene_mappings_invitrodb_v4_1_SEPT2023' file. Next, we leveraged KE-gene annotations provided by Saarimäki *et al.* [237] to identify gene sets associated with KEs having biological level of organization as either molecular or cellular. Thereafter, we mapped these KEs to ToxCast assay endpoints based on gene overlaps, and manually filtered the mappings based on the assay endpoint descriptions. Through this extensive toxicogenomics based manual curation, we obtained 212 assay endpoints mapped to 115 KEs for 1129 of the 6470 curated plastic additives (Figure 5.4).

Using CTD

CTD [27] is one of the largest toxicogenomics resources that compiles data on chemical-gene/protein, chemical-phenotype, chemical-disease and gene-disease associations from published literature. The concept of chemical (C), gene (G), phenotype (P) and disease (D) tetramers, i.e., CGPD-tetramers, was proposed to understand the phenotypes and diseases that result from the interaction of chemicals with genes [83, 174]. Based on our previous work [95], we retrieved the CGPD-tetramers associated with plastic additives within CTD, and leveraged them to identify the associated KEs within AOP-Wiki.

First, we downloaded the CTD's January 2024 release and constructed the CGPD-tetramers for the plastic additives based on the workflow proposed in our previous work [95]. This resulted in the identification of 124496 tetramers comprising 258 chemicals, 2932 genes, 1489 phenotypes and 690 diseases (Supplementary Table S5.6). Furthermore, we generated the immediate neighbor GO terms for the CGPD-tetramer phenotype GO terms using the GOSim package [177] available in R programming language. Thereafter, we overlapped the GO terms with the process identifiers of KEs within AOP-Wiki, and manually inspected to identify 307 KEs associated with 266 phenotypes for 241 of the 6470 curated plastic additives (Supplementary Table S5.7). We also manually inspected the disease terms to identify 157 KEs associated with 315 diseases for 232 of the 6470 curated plastic additives (Supplementary Table S5.7).

Using DEDuCT and NeurotoxKb

DEDuCT [8, 15] is one of the largest databases that compiles curated information on endocrine disrupting chemicals (EDCs) and their corresponding endocrine-mediated endpoints from published literature. Therefore, we compiled the endocrine-mediated endpoints corresponding to plastic additives within DEDuCT, and considered them to find associated KEs within AOP-Wiki. We manually inspected the endpoints and titles of KEs within AOP-Wiki, and identified 165 KEs that are associated with 188 endocrine-mediated endpoints for 203 of the 6470 curated plastic additives (Supplementary Table S5.7).

NeurotoxKb [29] is a manually curated resource on mammalian neurotoxicity associated endpoints of environmental chemicals curated from published literature. Therefore, we compiled the neurotoxic endpoints corresponding to plastic additives within NeurotoxKb, and considered them to find associated KEs within AOP-Wiki. We manually inspected the neurotoxic endpoints and KEs within AOP-Wiki, and identified 25 KEs that are associated with 24 neurotoxic endpoints for 92 of the 6470 curated plastic additives (Supplementary Table S5.7).

Using AOP-Wiki

AOP-Wiki also catalogs the stressor information for each AOP, where there exists well documented evidence of such stressor(s) showing response at multiple KEs, including MIEs [171]. Therefore, we relied on the stressor information within AOP-Wiki to identify

KEs associated with plastic additives. We retrieved information on stressors associated with each AOP, and identified 33 AOPs to be associated with 42 of the 6470 curated plastic additives (Supplementary Table S5.7). Thereafter, we identified 178 KEs in these 33 AOPs that are associated with plastic additives.

Overall, we identified 688 KEs that are associated with 1314 plastic additives (out of the 6470 plastic additives in our curated list) through the integration of heterogeneous toxicogenomics and biological endpoints data from five exposome-relevant resources: ToxCast, CTD, DEDuCT, NeurotoxKb and AOP-Wiki (Supplementary Table S5.7).

5.1.5 Compilation of chemical lists for priority use sectors of plastic additives

Globally, plastics are used across different scales and for various applications. The UNEP report [20] has identified 10 priority use sectors, based on the likelihood of exposure of chemicals in plastic products in these sectors to humans and environment. The 10 priority use sectors include ‘Toys and other children’s products’, ‘Furniture’, ‘Packaging including food contact materials’, ‘Electrical and electronic equipment’, ‘Transport’, ‘Personal care and household products’, ‘Medical devices’, ‘Building materials’, ‘Synthetic textiles’, and ‘Agriculture, aquaculture and fisheries’. To identify the plastic additives being used in each of the priority use sectors, we first compiled the list of chemicals in use in each of these sectors.

Chemical and Products Database (CPDat) [148] is among the largest resources that catalogs the presence of chemicals in various consumer products. For each product, CPDat assigns a Product Use Category (PUC) based on the general category and product type mentioned in the original data source [148]. Here, we relied on CPDat to identify the chemicals in use in each of the 10 priority use sectors. We accessed the CPDat data file [238] (last accessed on 15 February 2024) to compile the list of chemicals associated with different PUCs, and identified 20 PUCs to be grouped under the different priority use

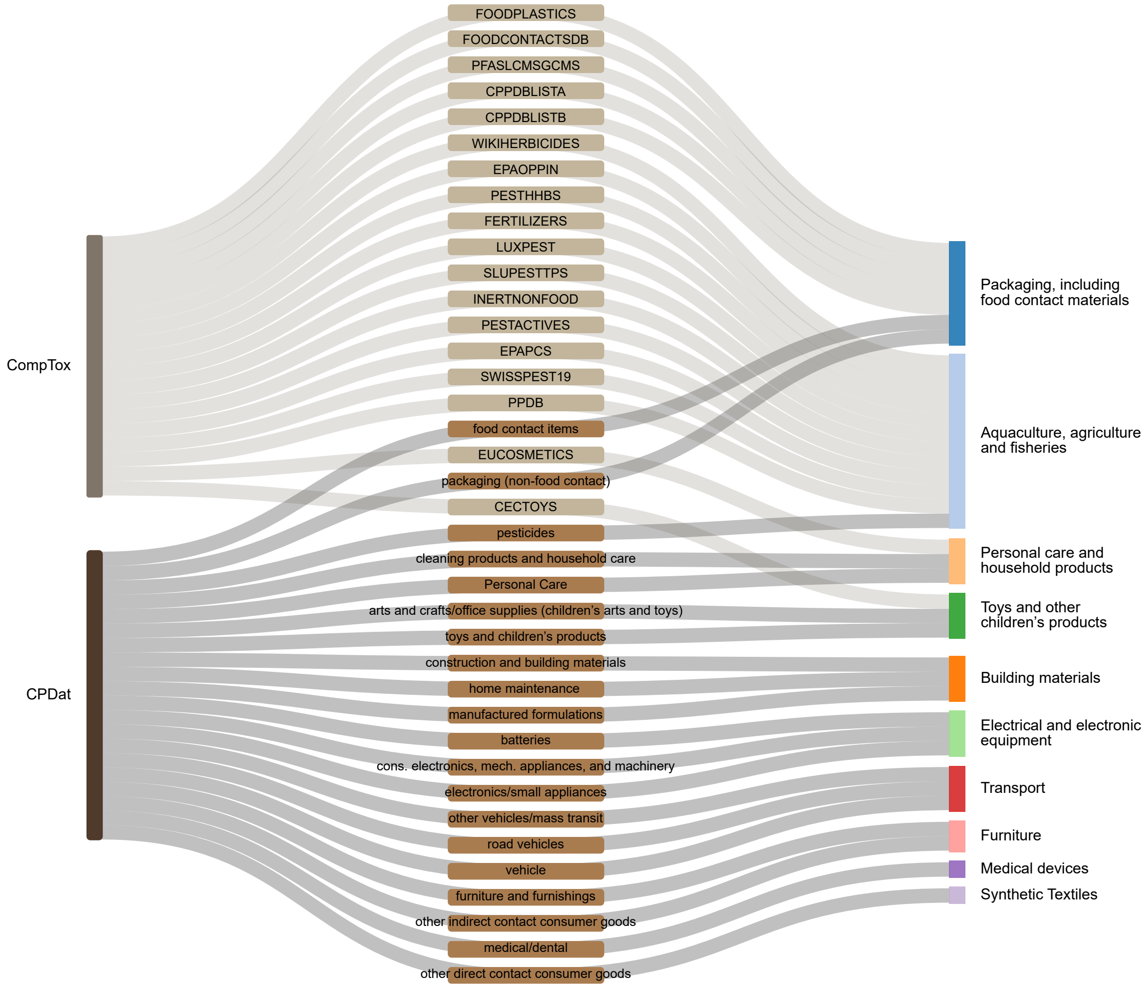

FOODPLASTICS
FOODCONTACTSDB
PFASLCMSGCMS
CPPDBLISTA
CPPDBLISTB
WIKIHERBICIDES
EPAOPPIN
PESTHHBS
FERTILIZERS
LUXPEST
SLUPESTTPS
INERTNONFOOD
PESTACTIVES
EPAPCS
SWISSPEST19
PPDB
food contact items
EUCOSMETICS
packaging (non-food contact)
CECTOYS
pesticides
cleaning products and household care
Personal Care
arts and crafts/office supplies (children's arts and toys)
toys and children's products
construction and building materials
home maintenance
manufactured formulations
batteries
cons. electronics, mech. appliances, and machinery
electronics/small appliances
other vehicles/mass transit
road vehicles
vehicle
furniture and furnishings
other indirect contact consumer goods
medical/dental
other direct contact consumer goods
CompTox
CPDat
Packaging, including food contact materials
Aquaculture, agriculture and fisheries
Personal care and household products
Toys and other children's products
Building materials
Electrical and electronic equipment
Transport
Furniture
Medical devices
Synthetic Textiles

Figure 5.5: Mapping of chemical or category lists from CompTox Chemicals Dashboard and CPDat with the 10 priority use sectors of plastic additives.

sectors (Figure 5.5; Supplementary Table S5.8).

The CompTox Chemicals Dashboard [239, 240] is one of the largest public repositories that provides access to different lists of chemicals associated with projects, publications, source databases or collections. Here, we queried the chemical lists based on their description, and identified chemical lists associated with the different priority use sectors (Figure 5.5; Supplementary Table S5.8). The chemical lists from CompTox which were used included food contact chemicals [241, 242], chemicals associated with plastic packaging [243], chemicals associated with pesticides [244, 245] and chemicals associated with plastic toys [246]. Furthermore, we compiled chemicals from an in-house repository namely, Fragrance Chemicals in Children's Products (FCCP) [247] as chemicals found in the use sector ‘Toys and other children’s products’. Finally, we compiled the chemicals in use in each of the priority use sectors, and identified plastic additives present in each sector (Supplementary Table S5.8).

5.1.6 Construction and visualization of the stressor-AOP network

Stressor-AOP network provides a holistic view of chemical perturbances across different AOPs [93]. To better understand the perturbances caused by the different plastic additives, we constructed a stressor-AOP network as a bipartite graph that linked various plastic additives to different AOPs within AOP-Wiki. In order to obtain high confidence associations between plastic additives and AOPs, we relied only on the curated list of 328 high confidence AOPs (Supplementary Table S5.3).

To obtain the stressor-AOP network for plastic additives, we initially linked plastic additives to AOPs if they share at least one associated KE, and thereafter, characterized each link between a stressor and an AOP based on the coverage score and level of relevance. Coverage score of a stressor-AOP link is defined as the ratio of number of KEs within that AOP associated with the stressor to the total number of KEs within that AOP [64]. Coverage score is a real number that takes a value between 0 and 1 and we denote this score as the edge weight of linkage between a stressor and an AOP in our stressor-AOP network. Next, we realized that the plastic additives were associated with different AOPs with varying levels of relevance. Therefore, we propose the following five-level criterion to qualitatively understand the relevance of associations:

- *Level 1*: The stressor is associated with at least one KE within an AOP, where the KE is neither MIE nor AO within that AOP.
- *Level 2*: The stressor is associated with at least one AO within an AOP, but not associated with any MIE within that AOP.
- *Level 3*: The stressor is associated with at least one MIE within an AOP, but not associated with any AO within that AOP.

- *Level 4*: The stressor is associated with at least one MIE and one AO within an AOP.
- *Level 5*: The stressor is associated with at least one MIE and one AO within an AOP and there exists a directed path between the associated MIE and AO.

Supplementary Table S5.9 contains all the data on the stressor-AOP network constructed for plastic additives, including the coverage score and level of relevance for each of the stressor-AOP links. We visualized this stressor-AOP network of plastic aditives using Cytoscape [180].

5.2 Results

5.2.1 Exploration of the curated list of plastic additives

Plastic additives are chemicals that are added to plastics to achieve specific desirable properties in the end product [218, 230]. External stress on such products can cause the separation of these additives, thereby leading to their release into the environment and eventually posing risks to humans and ecosystems [21]. In this study, we curated a list of plastic additives from chemicals cataloged in the UNEP report [20] (Supplementary Table S5.2) and explored their potential risks by systematically integrating the associated heterogeneous biological endpoints within the context of AOP framework (Figure 5.1).

The UNEP report provides functional annotations for each of these chemicals based on two independent studies by Aurisano *et al.* [226] and Wiesinger *et al.* [227]. Notably, Wiesinger *et al.* observed that their text mining approach for identifying these functional annotations lacked context sensitivity, leading to some inaccuracies. Despite this limitation, the UNEP report remains the most comprehensive source cataloging chemicals found in plastics and their associated functions. Therefore, we relied on the functional annotations provided by the report to identify 6470 chemicals with reported functions, which we designate as ‘plastic additives’ in this study (Supplementary Table S5.2).

Among the 6470 plastic additives, we observed that many chemicals (3217 of 6470) provide a variety of functions to the plastics, with colorants being the most frequently associated function (3675 of 6470) (Supplementary Table S5.2). Further, we observed that majority of plastic additives (4309 of 6470) are found in products made by different priority use sectors, of which 3963 additives are found in the use sector 'Packaging, including food contact materials' (Supplementary Table S5.2).

Next, we relied on the United States High Production Volume (USHPV) [147] chemical list and Organisation for Economic Co-operation and Development High Production Volume (OECD HPV) [146] chemical list and identified 2084 of 6470 plastic additives to be HPV chemicals (Supplementary Table S5.2). Notably, among these HPV plastic additives, we found 154 additives to be known endocrine disrupting chemicals (EDCs) with experimental evidence for endocrine disruption in humans or rodents from DEDuCT [8,15] and 101 additives as potential carcinogens based on International Agency for Research on Cancer (IARC) monographs on identification of carcinogenic hazards to humans [248] (Supplementary Table S5.2). Furthermore, we observed that 215 additives are identified as substances of very high concern (SVHC) [126] by European Chemicals Agency (ECHA) and 412 additives are prohibited for use as per REACH regulation [125] (Supplementary Table S5.2). Figure 5.6a shows the distribution of HPV, SVHC and REACH prohibited plastic additives across the 10 priority use sectors.

5.2.2 Plastic additives are accumulated in various human biospecimens

Humans are exposed to various plastic additives via direct contact, inhalation or ingestion, which can eventually accumulate in different human tissues and potentially lead to various adverse health effects [6, 20]. In order to explore the plastic additives detected in various human biospecimens, we relied on two databases namely, Tissue-specific Exposome Atlas (TExAs) [249] and Exposome-Explorer [250] which have compiled the

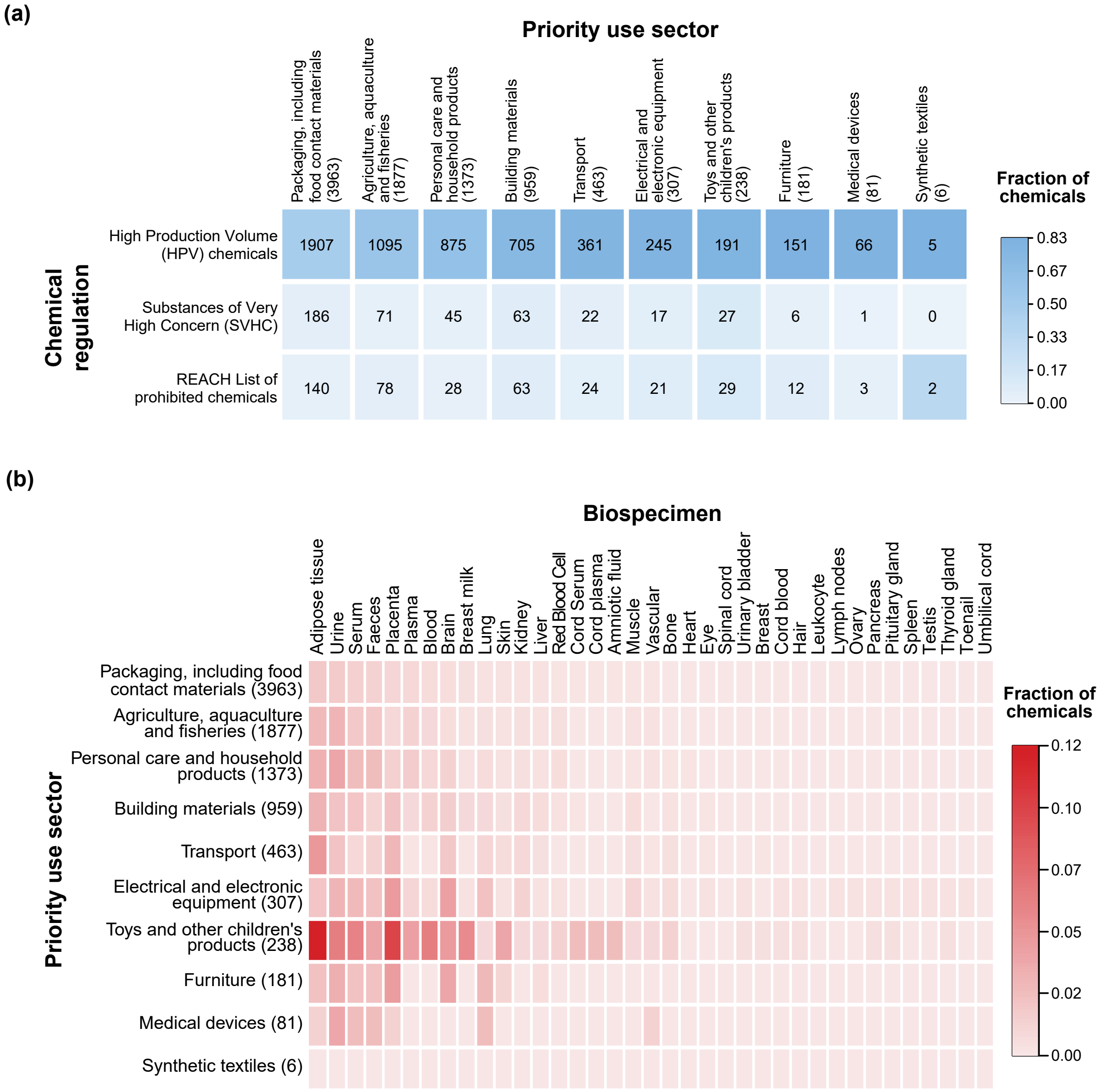


Figure 5.6: Identification of plastic additives in different chemical regulations and human biospecimens. (a) Heatmap depicting the presence of plastic additives from 10 priority use sectors in chemical regulations. The number of the plastic additives from the priority use sector in each of the chemical regulations is denoted in the heatmap. (b) Heatmap depicting the presence of plastic additives from 10 priority use sectors in different human biospecimens based on published exposure studies.

presence of environmental chemicals as xenobiotics in different human tissues from published exposure studies. Although, these two databases have compiled information from limited human exposure studies, they have documented 204 of the 6470 plastic additives to be accumulated as xenobiotics in 37 different human biospecimens (Figure 5.6b; Supplementary Table S5.2). Moreover, we observed that plastic additives from 9 of the 10 priority use sectors have been documented as xenobiotics in human biospecimens namely, faeces, serum, urine, lung, placenta and adipose tissue (Figure 5.6b). Note, the use sector 'Synthetic textiles' comprised the least number of additives (6 chemicals) in our curated list of 6470 additives, and there are no published exposure studies wherein their presence was detected in different human biospecimens.

5.2.3 Stressor-AOP network for plastic additives

Stressor-AOP networks provide a panoramic visualization of the different AOPs associated with stressors of interest, and help in understanding the stressor-induced adverse biological effects [93]. In this study, we therefore constructed stressor-AOP network to understand the various adverse effects induced by plastic additives. First, we followed a systematic approach that involved data-centric integration of heterogeneous toxicogenomics and biological endpoints data from five exposome-relevant resources namely, ToxCast, CTD, DEDuCT, NeurotoxKb and AOP-Wiki, and identified 688 KEs within AOP-Wiki to be associated with 1314 of the 6470 plastic additives (Supplementary Table S5.7). Thereafter, we curated 328 high confidence AOPs within AOP-Wiki and mapped them to plastic additives if at least one KE within that AOP is associated with the plastic additive. Based on these plastic additive-AOP associations, we constructed a plastic additives-centric bipartite stressor-AOP network comprising two types of nodes namely, 1287 plastic additives and 322 high confidence AOPs, and 46243 stressor-AOP links as edges between the two types of nodes, and we designate this bipartite network as plastic additives-AOP network. Notably, we observed that AOP-Wiki documented only 37 of the 1287 plastic additives in the constructed stressor-AOP network to be associated with 27 of the 322 high confidence AOPs in the network.

Next, we leveraged the KEs associated with plastic additives to compute the coverage score for the stressor-AOP links in the plastic additives-AOP network and observed that 20 plastic additives are associated with all the KEs (coverage score = 1.0) in 15 high confidence AOPs, and these stressor-AOP links were otherwise not documented in AOP-Wiki (Supplementary Table S5.9). Moreover, we calculated the levels of relevance for the stressor-AOP links in the plastic additives-AOP network and observed that 27189 links between 1155 plastic additives and 288 AOPs are classified as Level 1, 4236 links between 345 plastic additives and 241 AOPs are classified as Level 2, 14187 links between 1152 plastic additives and 139 AOPs are classified as Level 3, and 631 links between 118 plastic additives and 98 AOPs are classified as Level 5 (Supplementary Table S5.9). Note, the stressor-AOP links with Level 4 relevance were also satisfied by Level 5 criterion, and therefore, there are no stressor-AOP links with Level 4 relevance in the constructed network (Supplementary Table S5.9).

Next, we relied on the standardized disease ontology provided in Disease Ontology [251] database to classify the AOPs based on their AOs. Based on the standardized ontology, we classified 322 AOPs into 26 disease classes based on their AOs (Supplementary Tables S5.9 and S5.10). Note that 125 of the 322 AOPs could not be classified under any standardized ontology provided by Disease Ontology, and we therefore marked them as 'unclassified'. Importantly, we observed that cancer is the most represented disease category comprising 40 of the 322 AOPs in the plastic additives-AOP network (Supplementary Table S5.9). Finally, we have linked the plastic additives to their corresponding priority use sectors and the AOPs to their corresponding disease categories in the plastic additives-AOP network (Supplementary Table S5.9).

We relied on the graph visualization software Cytoscape [180] to visualize the plastic additives-AOP network for each of the 1287 plastic additives, and make them available on a dedicated website: https://cb.imsc.res.in/saopadditives/. In the website, the plastic additives are grouped based on their priority use sectors. For instance, the priority use sector ‘Toys and other children’s products’ consists of 162 plastic additives, 301 AOPs and 8300 stressor-AOP links, wherein 4696 links between 148 plastic additives and 265 AOPs are classified as Level 1, 1460 links between 75 plastic additives and 170 AOPs are classified as Level 2, 1928 links between 144 plastic additives and 117 AOPs are classified as Level 3, and 216 links between 30 plastic additives and 58 AOPs are classified as Level 5. Figure 5.7 shows a portion of the plastic additives-AOP network, comprising Level 5 stressor-AOP links for plastic additives in the use sector ‘Toys and other children’s products’.

Additionally, the constructed plastic additives-AOP network can highlight the adverse outcomes induced by plastic additives across different use sectors. Figure 5.8 shows the disease categories linked with 10 priority use sectors for plastic additives in the plastic additives-AOP network with Level 5 relevance, where cancer is the most represented disease category.

5.2.4 Stressor-AOP network reveals highly relevant AOPs associated with plastic additives

A stressor-AOP network can help identify most relevant AOPs associated with each stressor which can further highlight the complexity and diversity among toxicity pathways induced by that stressor [58]. Here, we considered stressor-AOP links from the constructed plastic additives-AOP network with Level 5 relevance and coverage score threshold of 0.4, and identified 107 of the 1287 plastic additives to be associated with 88 of the 322 AOPs through 526 stressor-AOP links (Supplementary Table S5.9). Note the coverage score threshold of 0.4 denotes that at least 40% of the KEs in that AOP are linked with the stressor [64, 95]. Notably, 15 of these 107 plastic additives are associated with more than 10 AOPs (Table 5.1). Among these 15 plastic additives, 14 are documented as EDCs in DEDuCT, and 10 are documented as carcinogens in IARC monographs (Supplementary Table S5.2).

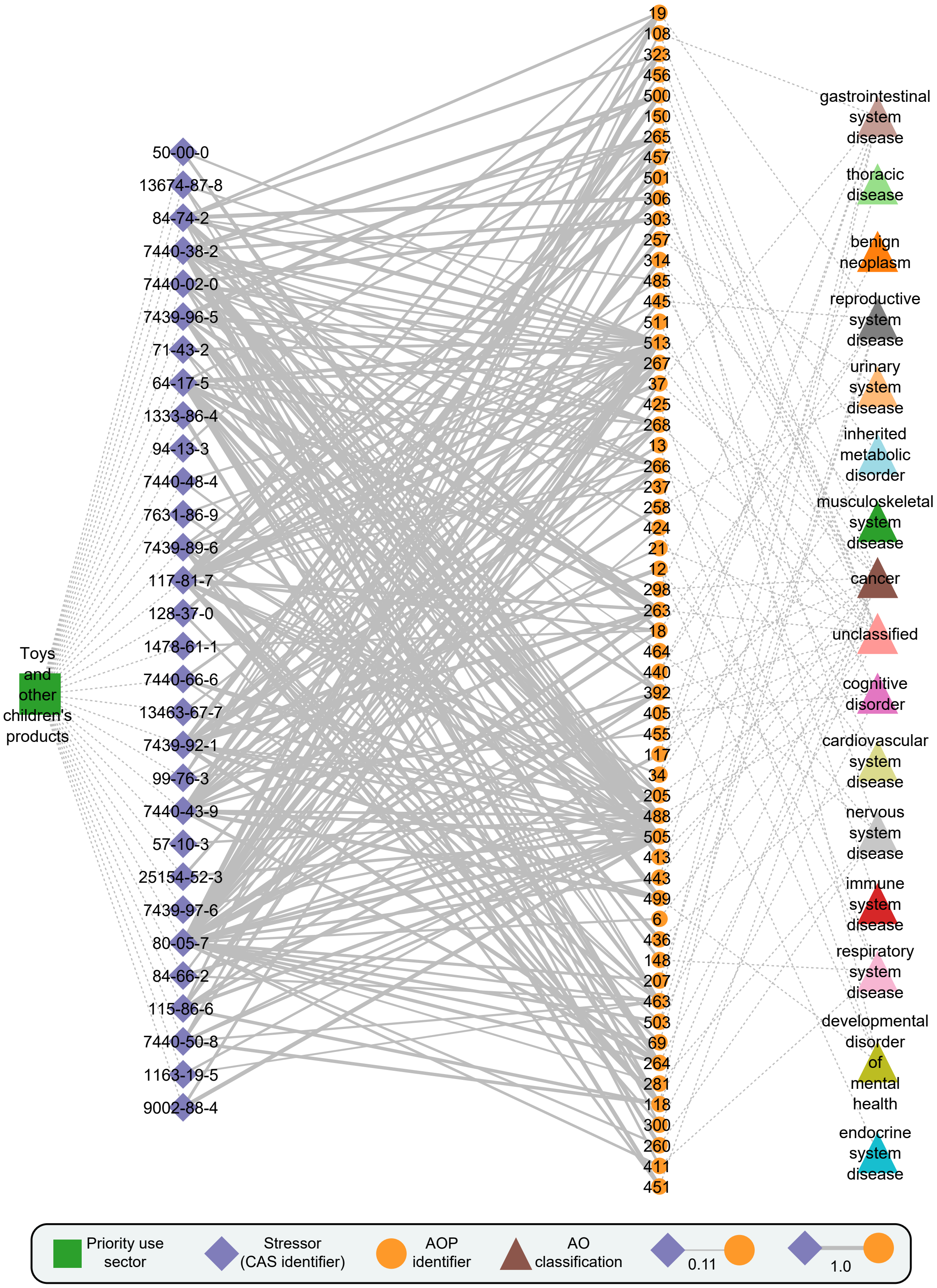


Figure 5.7: Visualization of a stressor-centric AOP network for plastic additives in the priority use sector ‘Toys and other children’s products’, along with the disease classifications of AOs in AOPs. In the stressor-AOP network, only edges or stressor-AOP links with Level 5 relevance are shown. Further, the edges or stressor-AOP links are weighted based on their coverage score, i.e., the fraction of KEs within AOP that are linked with the plastic additives.

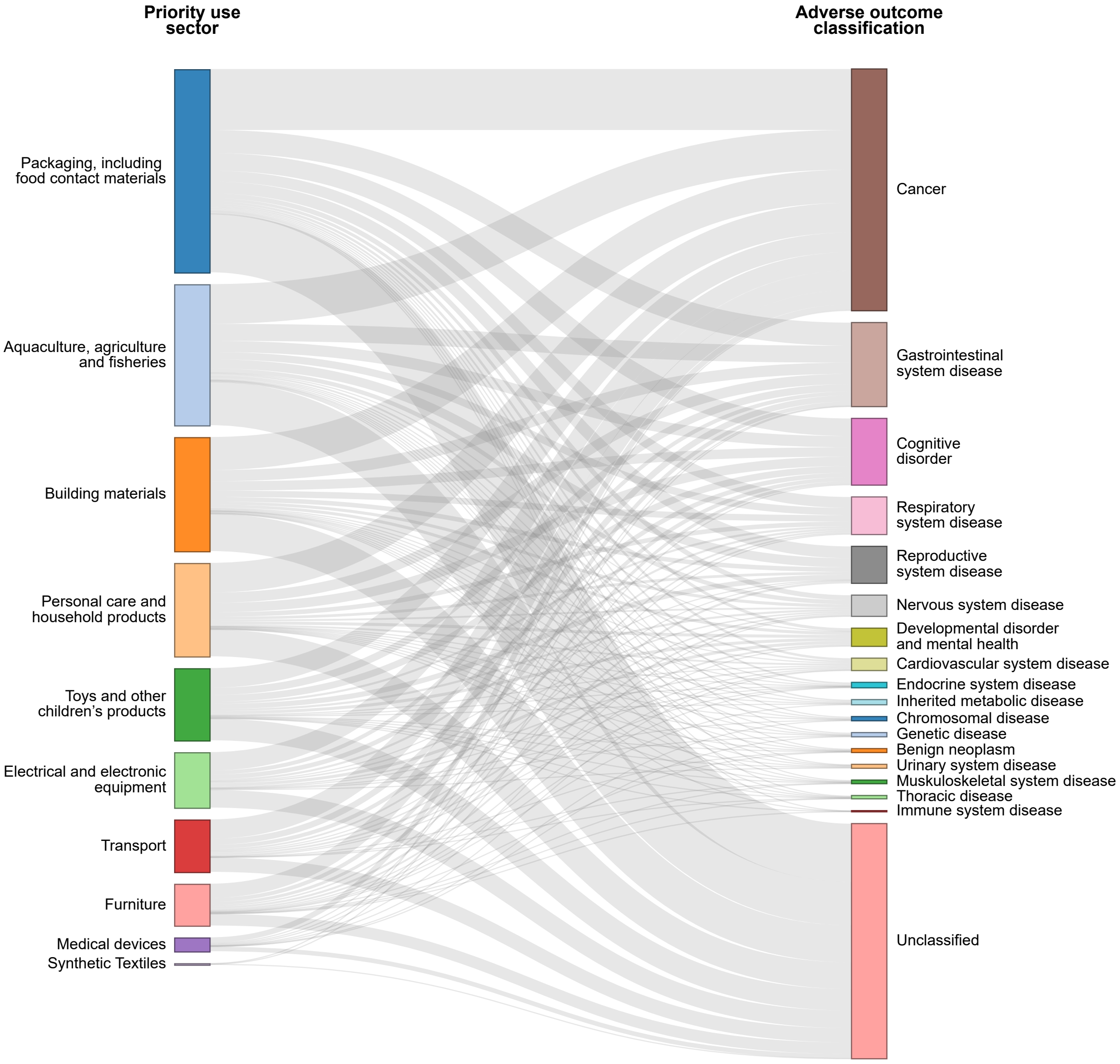


Figure 5.8: Aggregate visualization of the disease categories associated with AOPs in plastic additives stressor-AOP network with Level 5 relevance for each of the 10 priority use sectors.

Among the AOPs associated with these 15 plastic additives, we observed that majority of the AOPs are identified through our systematic data integrative approach. Notably, we observed that AOP:263, AOP:264, AOP:265, AOP:267 and AOP:268 are shared among all 15 plastic additives (Table 5.1). Moreover, we observed that these five AOPs share the same MIE 'Decrease, Coupling of oxidative phosphorylation' (KE:1446) and AO 'Decrease, Growth' (KE:1521), while AOP:263 titled 'Uncoupling of oxidative phosphorylation leading to growth inhibition via decreased cell proliferation' is endorsed by Working Group of the National Coordinators of the Test Guidelines Programme (WNT) and the Working Party on Hazard Assessment (WPHA) under the OECD AOP development programme.

An AOP network constructed from stressor-specific AOPs can highlight interactions among the associated AOPs, thereby aiding in the assessment of stressor-induced toxicity [63,64,67,95,252,253]. Among the 15 plastic additives, we observed that Benzo[a]pyrene (28 associated AOPs), Bisphenol A (27 associated AOPs), and Bis(2-ethylhexyl) phthalate (19 associated AOPs) are the top three chemicals based on the number of associated AOPs (Table 5.1). Although Benzo[a]pyrene has been annotated as a plastic additive in this study, evidence suggests that it is more likely a contaminant or a byproduct resulting from the use of extender oils or carbon black in plastic production [227, 254, 255]. Nonetheless, all these three chemicals are well-known pollutants. Therefore, we constructed AOP networks for each of these chemicals and explored their potential human-relevant and ecotoxicology-relevant toxicity pathways.

5.2.5 Exploration of toxicity pathways in AOP network constructed from Benzo[a]pyrene-relevant AOPs

Benzo[a]pyrene (B[a]P or CAS:50-32-8) has the largest number of associated AOPs (28 AOPs), all of which were solely identified through our systematic data integrative approach (Table 5.1). These 28 AOPs are classified under various disease categories namely, cancer, gastrointestinal system disease, reproductive system disease, respiratory system disease, cognitive disorder, thoracic disease and musculoskeletal system disease (Supplementary Table S5.9). Previously, Yang *et al.* [253] had constructed an AOP network for B[a]P-induced toxicity, but they relied only on CTD to identify KEs associated with B[a]P and focused only on B[a]P-induced male reproductive damages. Therefore, we relied on 28 AOPs associated with B[a]P-induced toxicity (which we designate as B[a]P-AOPs) and constructed an AOP network to explore various adverse effects associated with B[a]P.

Next, we computed the cumulative WoE for each of these 28 B[a]P-AOPs based on their KER information to assess the biological plausibility [77, 95]. We observed that 9 of these 28 B[a]P-AOPs have 'High' cumulative WoE and 6 B[a]P-AOPs have 'Moderate' cumulative WoE (Supplementary Table S5.11). Moreover, we observed that many of these 28 B[a]P-AOPs are applicable across various species and developmental stages (Supplementary Table S5.11). Figure 5.9 shows the undirected AOP network representation of the 28 B[a]P-AOPs, where nodes represent B[a]P-AOPs and the edges represent the existence of shared KEs between two AOPs. We observed that the B[a]P-AOPs form three connected components (with two or more AOPs) and two isolated nodes, where the largest connected component (LCC) (labeled C1) comprises 18 B[a]P-AOPs.

We constructed and visualized a directed AOP network to explore the interactions among the B[a]P-AOPs present in the LCC C1 (Figure 5.10). We observed that the directed network comprised 66 unique KEs (including 7 MIEs and 11 AOs) and 99 unique KERs (Figure 5.10; Supplementary Table S5.12). Among the 66 KEs, 36 KEs were associated with B[a]P-induced toxicity through our systematic data integrative approach, of which 5 are MIEs and 10 are AOs (Figure 5.10). Notably, we observed that the toxicity pathway originating from MIE 'Activation, AhR' (KE:18), passing through KEs 'Altered gene expression, NRF2 dependent antioxidant pathway' (KE:1917) and 'Increase, Cell Proliferation' (KE:870), and eventually terminating at AO 'Lung Cancer' (KE:1670) consists of 4 of the 36 KEs associated with B[a]P-induced toxicity (Figure 5.10). Upon further inspection, we identified that this toxicity pathway was captured in the AOP:420 ti-

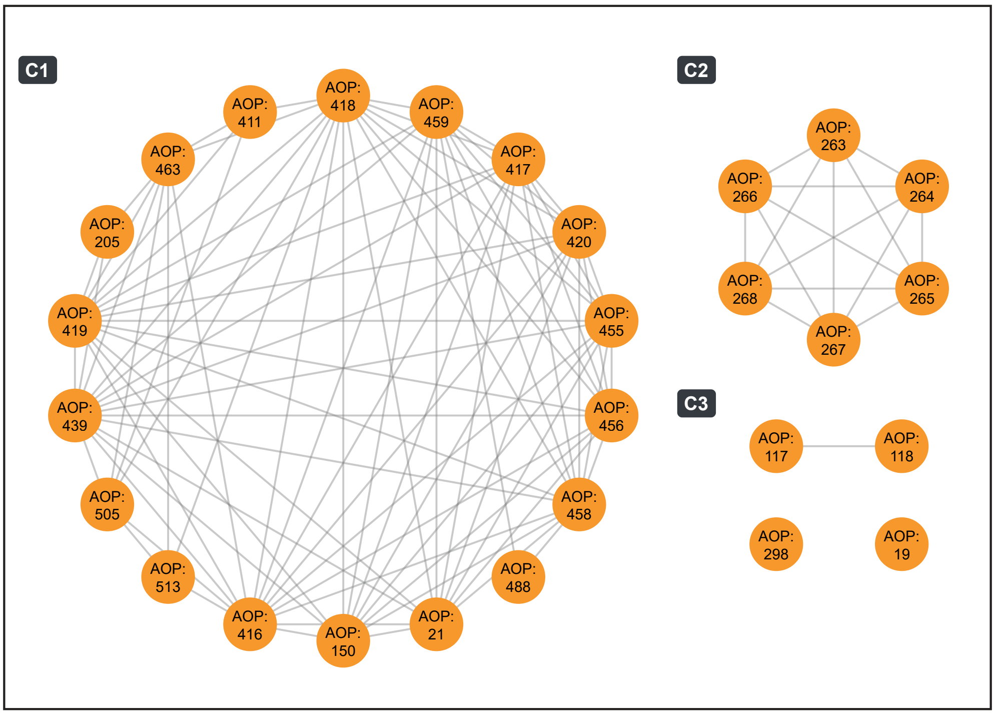


Figure 5.9: Undirected network of B[a]P-AOPs. Each node corresponds to B[a]P-AOP and an edge between two nodes denotes that the two AOPs share at least one KE. This undirected network has 3 connected components (with two or more nodes) which are labeled as C1, C2 and C3, and 2 isolated nodes.

tled 'Aryl hydrocarbon receptor activation leading to lung cancer through sustained NRF2 toxicity pathway', which was systematically built and supported by extensive literature survey and experimental data on B[a]P [51].

Next, we computed different node-centric network measures to explore various features of this directed network. We observed that the MIE 'Activation, AhR' (KE:18) has the highest out-degree of 16, while the KE 'Altered, Cardiovascular development/function' (KE:317) and AO 'N/A, Liver fibrosis' (KE:344) have the highest in-degree of 5 (Supplementary Table S5.12). The MIE 'Increased, Reactive oxygen species' (KE:1115) has the highest betweenness centrality value, denoting that several toxicity pathways are passing through it in this network (Figure 5.11) [59]. The KEs 'Induction, CYP1A2/CYP1A5' (KE:850) and 'Altered gene expression, NF-kB dependent Interleukin-6 pathway' (KE:1921), and MIEs 'Activation, AhR' (KE:18) and 'Increase, Reactive Oxygen Species production' (KE:257) have the highest eccentricity, denoting that they are the most re-

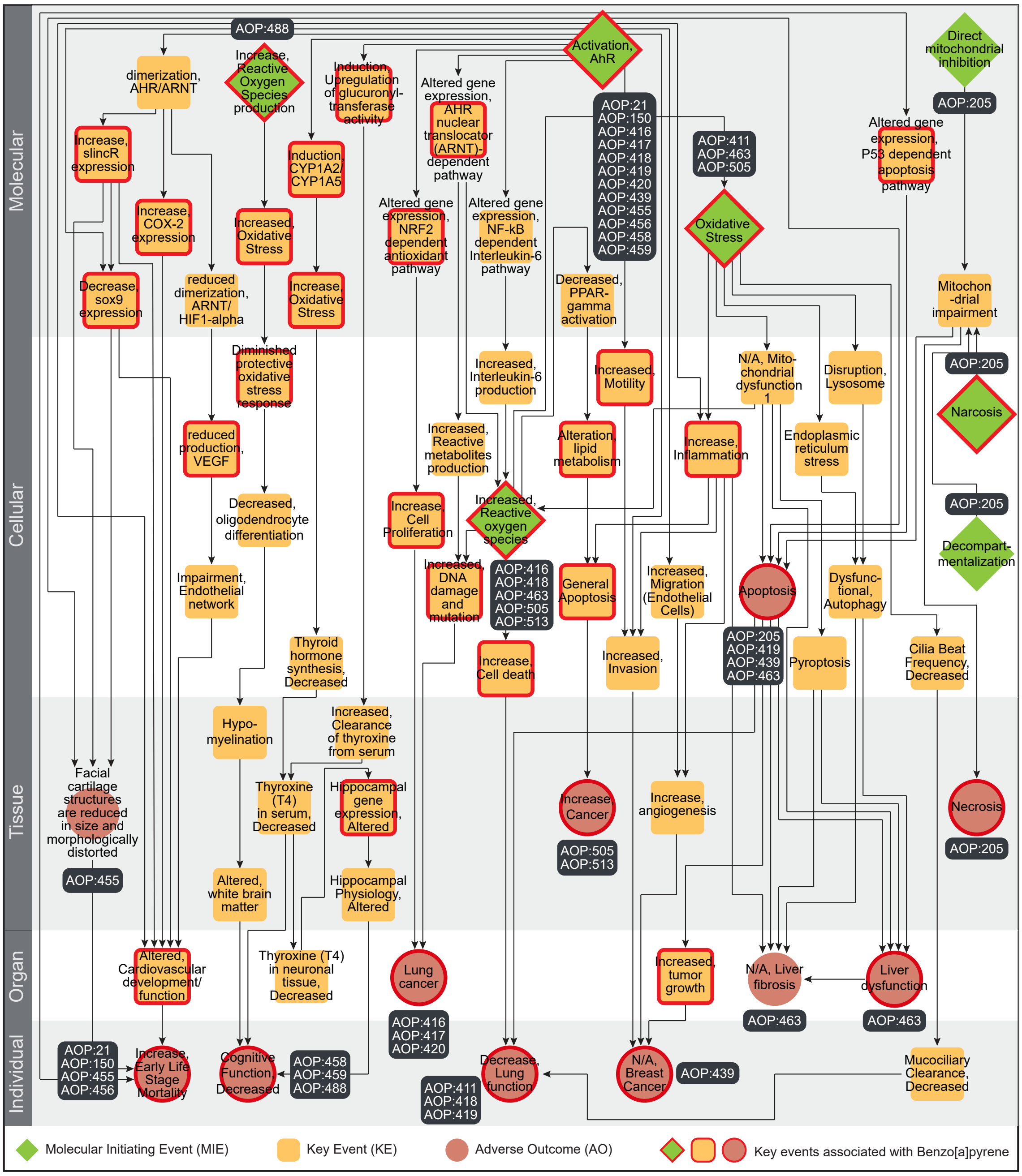

Molecular
Cellular
Tissue
Organ
Individual
AOP:488
dimerization, AHR/ARNT
Increase, Reactive Oxygen Species production
Induction, Upregulation of glucuronyl-transferase activity
Activation, AhR
Direct mitochondrial inhibition
Altered gene expression, AHR nuclear translocator (ARNT)-dependent pathway
AOP:205
Increase, slincR expression
Induction CYP1A2/ CYP1A5
AOP:21 AOP:150 AOP:416 AOP:417 AOP:418 AOP:419 AOP:420 AOP:439 AOP:455 AOP:456 AOP:458 AOP:459
AOP:411 AOP:463 AOP:505
Altered gene expression, P53 dependent apoptosis pathway
Increase, COX-2 expression
Increased, Oxidative Stress
Altered gene expression, NRF2 dependent antioxidant pathway
Altered gene expression, NF-kB dependent Interleukin-6 pathway
Oxidative Stress
Decrease, sox9 expression
reduced dimerization, ARNT/ HIF1-alpha
Increase, Oxidative Stress
Decreased, PPAR-gamma activation
Mitochondrial impairment
Diminished protective oxidative stress response
Increased, Interleukin-6 production
Increased, Motility
N/A, Mitochondrial dysfunction 1
Disruption, Lysosome
AOP:205
Narcosis
reduced production, VEGF
Increased, Reactive metabolites production
Alteration, lipid metabolism
Increase, Inflammation
Endoplasmic reticulum stress
Decreased, oligodendrocyte differentiation
Increase, Cell Proliferation
Increased, Reactive oxygen species
AOP:205
Decompartmentalization
Impairment, Endothelial network
Increased, DNA damage and mutation
AOP:416 AOP:418 AOP:463 AOP:505 AOP:513
General Apoptosis
Increased, Migration (Endothelial Cells)
Apoptosis
Dysfunctional, Autophagy
Thyroid hormone synthesis, Decreased
Increase, Cell death
Increased, Invasion
AOP:205 AOP:419 AOP:439 AOP:463
Pyroptosis
Cilia Beat Frequency, Decreased
Hypo-myelination
Increased, Clearance of thyroxine from serum
Facial cartilage structures are reduced in size and morphologically distorted
Thyroxine (T4) in serum, Decreased
Hippocampal gene expression, Altered
Increase, Cancer
Increase, angiogenesis
Necrosis
AOP:505 AOP:513
AOP:205
AOP:455
Altered, white brain matter
Hippocampal Physiology, Altered
Altered, Cardiovascular development/ function
Thyroxine (T4) in neuronal tissue, Decreased
Lung cancer
Increased, tumor growth
N/A, Liver fibrosis
Liver dysfunction
AOP:416 AOP:417 AOP:420
AOP:463
AOP:463
AOP:21 AOP:150 AOP:455 AOP:456
Increase, Early Life Stage Mortality
Cognitive Function, Decreased
AOP:458 AOP:459 AOP:488
Decrease, Lung function
N/A, Breast Cancer
AOP:439
Mucociliary Clearance, Decreased
AOP:411 AOP:418 AOP:419
Molecular Initiating Event (MIE)
Key Event (KE)
Adverse Outcome (AO)
Key events associated with Benzo[a]pyrene

Figure 5.10: Directed network corresponding to the largest component in the undirected B[a]P-AOP network comprising 66 KEs and 99 KERs. Among the 66 KEs, 7 are categorized as MIEs (denoted as diamond), 11 are categorized as AOs (denoted as circle), and the remaining 48 are categorized as KEs (denoted as rounded square). The 36 KEs (including MIEs and AOs) associated with B[a]P are marked in 'red'. In this figure, the 66 KEs are arranged vertically according to their level of biological organization.

motely placed KEs in this network (Figure 5.12) [181].

Finally, we relied on artificial intelligence (AI) based tool, AOP-helpFinder [183,184] and Abstract Sifter [185] to screen published literature and manually identified novel associations between the B[a]P-induced toxicities and the remaining 30 KEs in the directed AOP network (Supplementary Table S5.13). Additionally, we compiled auxiliary evidence for the 36 KEs that were associated with B[a]P-induced toxicity through our systematic data integrative approach. Furthermore, we compiled information on the type of evidence and the reported toxicity dosage values of B[a]P exposure from these published evidence (Supplementary Table S5.13). To conclude, we performed two case studies to explore both the human-relevant and ecotoxicology-relevant B[a]P-induced toxicity pathways from this directed AOP network.

Toxicity pathway linking B[a]P exposure to liver fibrosis in humans

Liver fibrosis, which results from chronic damage to the liver, is a characteristic of many chronic liver diseases [256]. Previously, exposome-based studies had found a significant association between environmental chemicals such as B[a]P and different liver diseases [257, 258]. Here, we observed an emergent B[a]P-induced toxicity pathway originating from MIE 'Activation, AhR' (KE:18) and terminating at AO 'N/A, Liver fibrosis' (KE:344). Therefore, we relied on this emergent toxicity pathway to understand the rationale behind B[a]P-induced liver fibrosis in humans.

Various *in vivo* and *in vitro* experiments in human cell lines and rodents had shown that B[a]P induces different downstream processes through the activation of AhR [51, 259–261]. Subsequently, B[a]P exposure has been observed to induce oxidative stress in cells through increased interleukin 6 (IL-6) production as a result of activated NF-κB signaling pathway [51,262,263]. The oxidative stress caused by B[a]P exposure has been studied as a cause for disruption of lysosomes, eventually leading to dysfunctional autophagy [264,265]. Finally, it has been shown that B[a]P exposure can induce different fibrotic pathways, including dysfunctional autophagy, in human hepatic models [266,267].

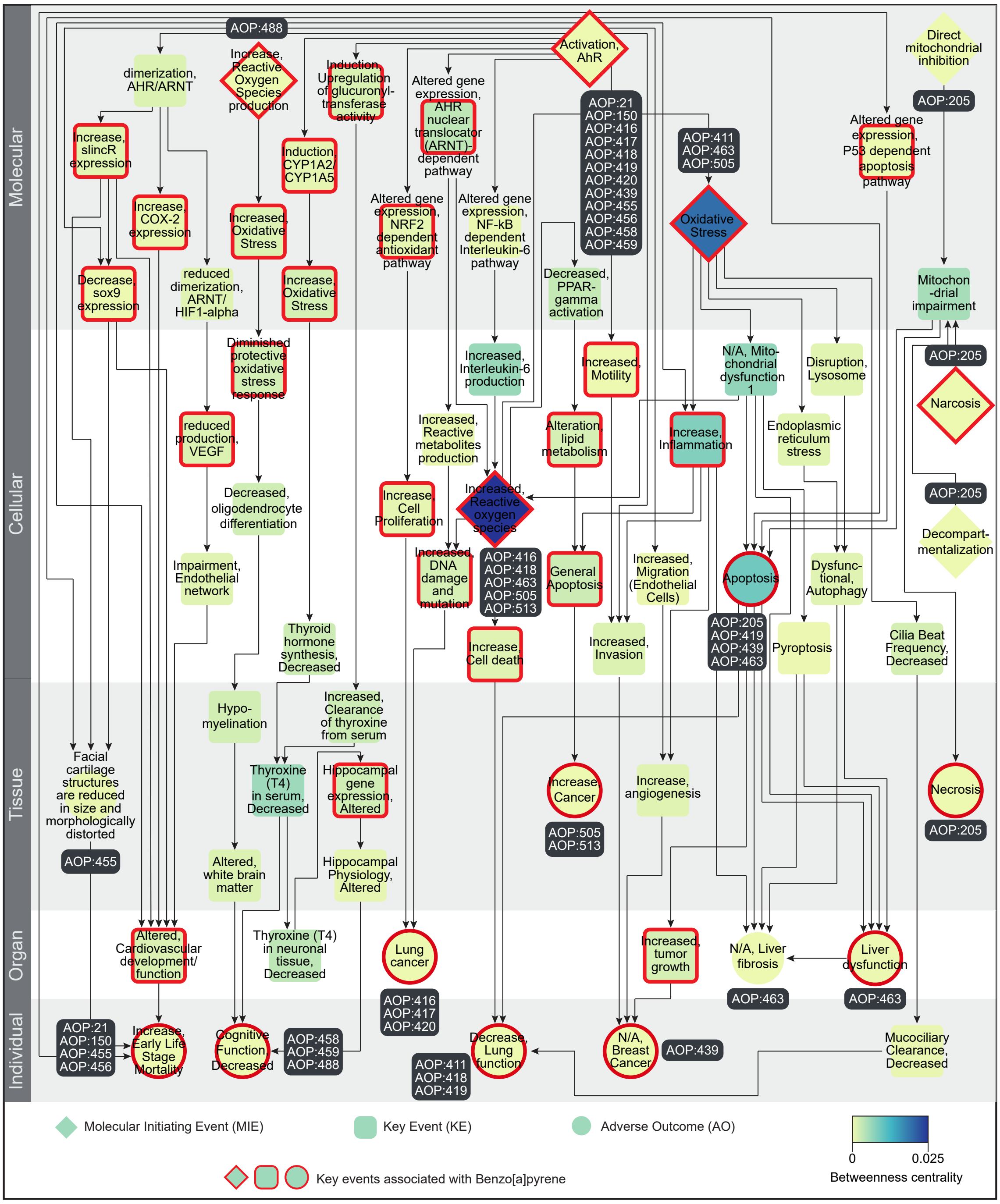

Molecular
Cellular
Tissue
Organ
Individual
AOP:488
Increase, Reactive Oxygen Species production
Induction, Upregulation of glucuronyl-transferase activity
Activation, AhR
Direct mitochondrial inhibition
AOP:205
dimerization, AHR/ARNT
Altered gene expression, AHR nuclear translocator (ARNT)-dependent pathway
AOP:21 AOP:150 AOP:416 AOP:417 AOP:418 AOP:419 AOP:420 AOP:439 AOP:455 AOP:456 AOP:458 AOP:459
AOP:411 AOP:463 AOP:505
Altered gene expression, P53 dependent apoptosis pathway
Increase, slincR expression
Induction, CYP1A2/ CYP1A5
Increase COX-2 expression
Increased, Oxidative Stress
Altered gene expression, NRF2 dependent antioxidant pathway
Altered gene expression, NF-kB dependent Interleukin-6 pathway
Oxidative Stress
Decrease, sox9 expression
reduced dimerization, ARNT/ HIF1-alpha
Increase Oxidative Stress
Decreased, PPAR-gamma activation
Mitochon-drial impairment
Diminished protective oxidative stress response
Increased, Interleukin-6 production
Increased, Motility
N/A, Mito-chondrial dysfunction 1
Disruption, Lysosome
AOP:205
Narcosis
reduced production, VEGF
Increased, Reactive metabolites production
Alteration, lipid metabolism
Increase, Inflammation
Endoplasmic reticulum stress
Decreased, oligodendrocyte differentiation
Increase Cell Proliferation
Increased, Reactive oxygen species
AOP:205
Decompart-mentalization
Impairment, Endothelial network
Increased, DNA damage and mutation
AOP:416 AOP:418 AOP:463 AOP:505 AOP:513
General Apoptosis
Increased, Migration (Endothelial Cells)
Apoptosis
Dysfunc-tional, Autophagy
Thyroid hormone synthesis, Decreased
Increase Cell death
Increased, Invasion
AOP:205 AOP:419 AOP:439 AOP:463
Pyroptosis
Cilia Beat Frequency, Decreased
Hypo-myelination
Increased, Clearance of thyroxine from serum
Facial cartilage structures are reduced in size and morphologically distorted
AOP:455
Thyroxine (T4) in serum, Decreased
Hippocampal gene expression, Altered
Increase Cancer
AOP:505 AOP:513
Increase, angiogenesis
Necrosis
AOP:205
Altered, white brain matter
Hippocampal Physiology, Altered
Altered, Cardiovascular development/ function
Thyroxine (T4) in neuronal tissue, Decreased
Lung cancer
Increased, tumor growth
N/A, Liver fibrosis
Liver dysfunction
AOP:416 AOP:417 AOP:420
AOP:463
AOP:463
AOP:21 AOP:150 AOP:455 AOP:456
Increase, Early Life Stage Mortality
Cognitive Function, Decreased
AOP:458 AOP:459 AOP:488
AOP:411 AOP:418 AOP:419
Decrease, Lung function
N/A, Breast Cancer
AOP:439
Mucociliary Clearance, Decreased
Molecular Initiating Event (MIE)
Key Event (KE)
Adverse Outcome (AO)
0
0.025
Betweenness centrality
Key events associated with Benzo[a]pyrene

Figure 5.11: Directed network corresponding to the LCC (C1) in the B[a]P-AOP network, where the KEs (including MIEs and AOs) are colored based on their betweenness centrality values. The 36 KEs (including MIEs and AOs) associated with B[a]P are marked in 'red'. In this figure, the 66 KEs are arranged vertically according to their level of biological organization.

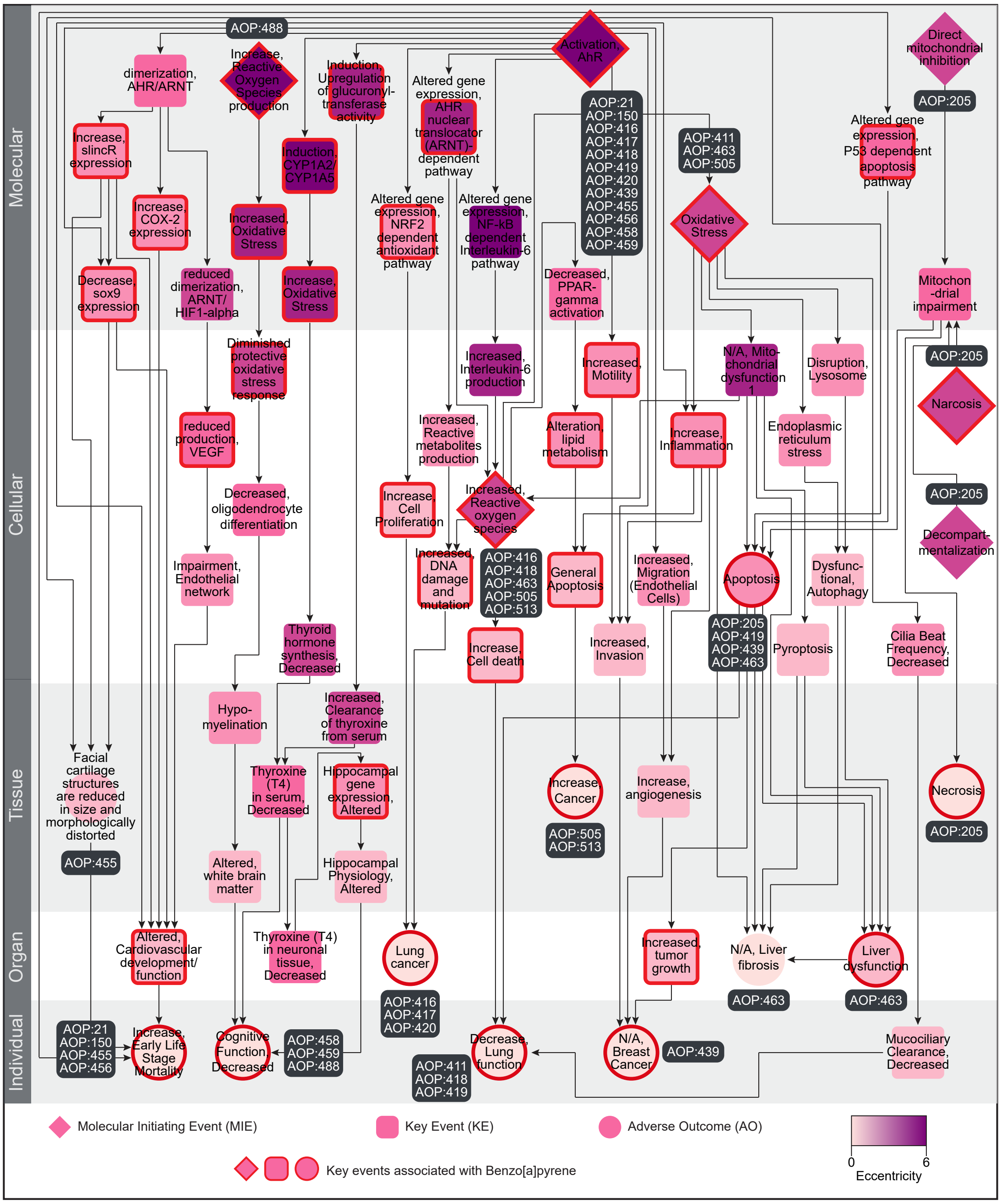

Molecular
Cellular
Tissue
Organ
Individual
AOP:488
dimerization, AHR/ARNT
Increase, Reactive Oxygen Species production
Induction, Upregulation of glucuronyl-transferase activity
Altered gene expression, AHR nuclear translocator (ARNT)-dependent pathway
Activation, AhR
Direct mitochondrial inhibition
AOP:205
AOP:21 AOP:150 AOP:416 AOP:417 AOP:418 AOP:419 AOP:420 AOP:439 AOP:455 AOP:456 AOP:458 AOP:459
AOP:411 AOP:463 AOP:505
Altered gene expression, P53 dependent apoptosis pathway
Increase, slincR expression
Induction, CYP1A2/ CYP1A5
Increase, COX-2 expression
Increased, Oxidative Stress
Altered gene expression, NRF2 dependent antioxidant pathway
Altered gene expression, NF-kB dependent Interleukin-6 pathway
Oxidative Stress
Decrease, sox9 expression
reduced dimerization, ARNT/ HIF1-alpha
Increase, Oxidative Stress
Decreased, PPAR-gamma activation
Mitochondrial impairment
Diminished protective oxidative stress response
Increased, Interleukin-6 production
Increased, Motility
N/A, Mitochondrial dysfunction 1
Disruption, Lysosome
AOP:205
Narcosis
reduced production, VEGF
Increased, Reactive metabolites production
Alteration, lipid metabolism
Increase, Inflammation
Endoplasmic reticulum stress
Decreased, oligodendrocyte differentiation
Increase, Cell Proliferation
Increased, Reactive oxygen species
AOP:205
Decompartmentalization
Impairment, Endothelial network
Increased, DNA damage and mutation
AOP:416 AOP:418 AOP:463 AOP:505 AOP:513
General Apoptosis
Increased, Migration (Endothelial Cells)
Apoptosis
Dysfunctional, Autophagy
Thyroid hormone synthesis, Decreased
Increase, Cell death
Increased, Invasion
AOP:205 AOP:419 AOP:439 AOP:463
Pyroptosis
Cilia Beat Frequency, Decreased
Hypo-myelination
Increased, Clearance of thyroxine from serum
Facial cartilage structures are reduced in size and morphologically distorted
Thyroxine (T4) in serum, Decreased
Hippocampal gene expression, Altered
Increase, Cancer
Increase, angiogenesis
Necrosis
AOP:455
AOP:505 AOP:513
AOP:205
Altered, white brain matter
Hippocampal Physiology, Altered
Altered, Cardiovascular development/ function
Thyroxine (T4) in neuronal tissue, Decreased
Lung cancer
Increased, tumor growth
N/A, Liver fibrosis
Liver dysfunction
AOP:416 AOP:417 AOP:420
AOP:463
AOP:463
AOP:21 AOP:150 AOP:455 AOP:456
Increase, Early Life Stage Mortality
Cognitive Function, Decreased
AOP:458 AOP:459 AOP:488
Decrease, Lung function
AOP:411 AOP:418 AOP:419
N/A, Breast Cancer
AOP:439
Mucociliary Clearance, Decreased
Molecular Initiating Event (MIE)
Key Event (KE)
Adverse Outcome (AO)
Key events associated with Benzo[a]pyrene
0
6
Eccentricity

Figure 5.12: Directed network corresponding to the LCC (C1) in the B[a]P-AOP network, where the KEs (including MIEs and AOs) are colored based on their eccentricity values. The 36 KEs (including MIEs and AOs) associated with B[a]P are marked in 'red'. In this figure, the 66 KEs are arranged vertically according to their level of biological organization.

In conclusion, by leveraging various published evidence, we were able to explore a potential toxicity pathway that links B[a]P-induced toxicity with liver fibrosis in humans.

Toxicity pathway linking B[a]P exposure to early life-stage mortality in aquatic organisms

B[a]P is found in large quantities in different aquatic environments due to various anthropogenic activities and waste discharges from both household and industries [268, 269]. B[a]P is a toxic pollutant, and drastically affects various aquatic organisms, including economically relevant fish [269–271]. Here, we observed that the AOP titled 'Aryl hydrocarbon receptor activation leading to early life stage mortality via sox9 repression induced impeded craniofacial development' (AOP:455), with biological applicability for developmental effects in aquatic species, has been identified as a B[a]P-AOP (Supplementary Table S5.11). Moreover, this B[a]P-AOP is part of the largest connected component, and is currently included in the OECD work plan (Supplementary Table S5.3). Therefore, we relied on this AOP to understand the rationale behind B[a]P-induced ecotoxicological effects in aquatic organisms.

Independent *in vivo* experiments in zebrafish and clam have shown that B[a]P exposure alters gene expression patterns through activation of AhR and subsequent dimerization of AhR and ARNT in affected tissues [272, 273]. It has been shown that B[a]P exposure in zebrafish facilitates the recruitment of AhR-dependent long noncoding RNA (slincR) to *sox9b* 5' UTR, eventually repressing its transcription [274]. *sox9b* is an important transcription factor involved in chondrocyte differentiation during zebrafish development [275]. Subsequently it has been shown that B[a]P exposure induces alteration in expression patterns of genes involved in chondrogenesis, thereby leading to improper craniofacial skeleton development and eventually early life stage mortality in zebrafish [271, 276]. In conclusion, by leveraging various published evidence, we were able to explore a potential toxicity pathway that links B[a]P-induced toxicity with early life stage mortality in aquatic species.

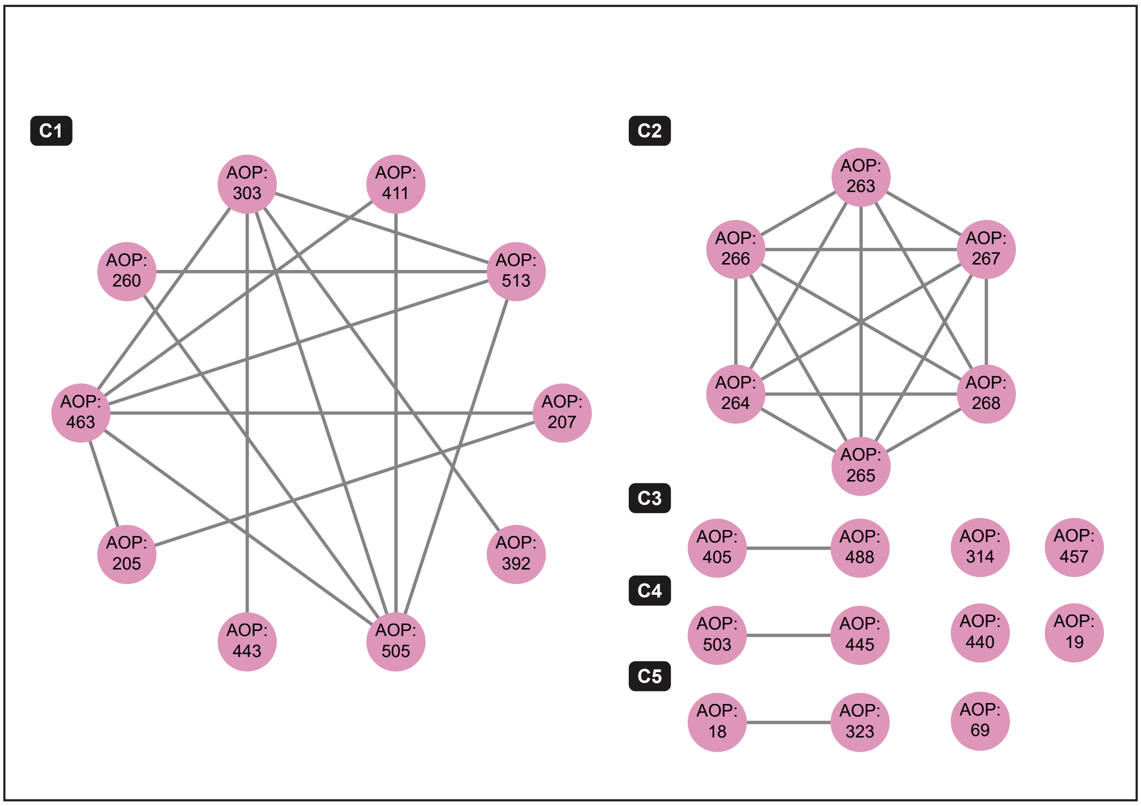


Figure 5.13: Undirected network of BPA-AOPs. Each node corresponds to BPA-AOP and an edge between two nodes denotes that the two AOPs share at least one KE. This undirected network has 5 connected components (with two or more nodes) which are labeled as C1, C2, C3, C4 and C5, and 5 isolated nodes.

5.2.6 Exploration of toxicity pathways in AOP network constructed from Bisphenol A-relevant AOPs

Similar to the construction of B[a]P-AOP network, we constructed the AOP network for Bisphenol A (BPA or CAS:80-05-7) toxicity. We identified 27 highly relevant AOPs associated with BPA-induced toxicities (Table 5.1), which we designated as BPA-AOPs. Among the 27 BPA-AOPs, we observed that 10 have ‘High’ cumulative WoE and 6 have ‘Moderate’ cumulative WoE (Supplementary Table S5.14). In the undirected AOP network constructed from these 27 BPA-AOPs (Figure 5.13), we observed five connected components (with two or more AOPs) and five isolated nodes, where the LCC (labeled C1) comprises 10 BPA-AOPs.

We constructed and visualized a directed AOP network to explore the interactions among the BPA-AOPs present in the LCC C1 (Figure 5.13). We observed that the directed network comprised 55 unique KEs (including 12 MIEs and 11 AOs) and 72 unique KERs (Figure 5.14; Supplementary Table S5.15). Among the 55 KEs, 31 KEs were associated with BPA-induced toxicity through our systematic data integrative approach, of which 9 are MIEs and 10 are AOs (Figure 5.14).

Then, we computed several node-centric measures for the constructed BPA-AOP directed network. We observed that the AO 'Apoptosis' (KE:1262) and 'N/A, Liver fibrosis' (KE:344) have the highest in-degree of 5, while the MIE 'Oxidative Stress' (KE:1392) has the highest out-degree of 5 (Supplementary Table S5.15). The MIE 'Increased, Reactive oxygen species' (KE:1115) has the highest betweenness centrality value, denoting that several toxicity pathways are passing through it in this network (Figure 5.15; Supplementary Table S5.15). The MIEs 'Bradykinin system, hyperactivated' (KE:1867), 'Fibrinolysis, decreased' (KE:1866) and 'Frustrated phagocytosis' (KE:1668) have the highest eccentricity, denoting that they are the most remotely placed KEs in this network (Figure 5.16; Supplementary Table S5.15). Finally, following a similar approach taken in the B[a]P-AOP directed network analysis, we compiled auxiliary evidence from published literature for all these 55 KEs (Supplementary Table S5.16), and performed two case studies to explore both the human-relevant and ecotoxicology-relevant BPA-induced toxicity pathways from this directed BPA-AOP network.

Toxicity pathway linking BPA exposure to neurodegeneration in humans

Neurodegeneration refers to the progressive loss of structure or function of neurons, including their death [277, 278]. This process is a characteristic feature of various neurological diseases, such as Alzheimer's disease, Parkinson's disease and Huntington's disease [277, 278]. Here, we observed BPA-AOP titled 'CYP2E1 activation and formation of protein adducts leading to neurodegeneration' (AOP:260) having a cumulative WoE of 'High' (Supplementary Table S5.14) and taxonomical relevance to humans (Supplementary Table S5.14). Therefore, we relied on this AOP to understand the rationale behind

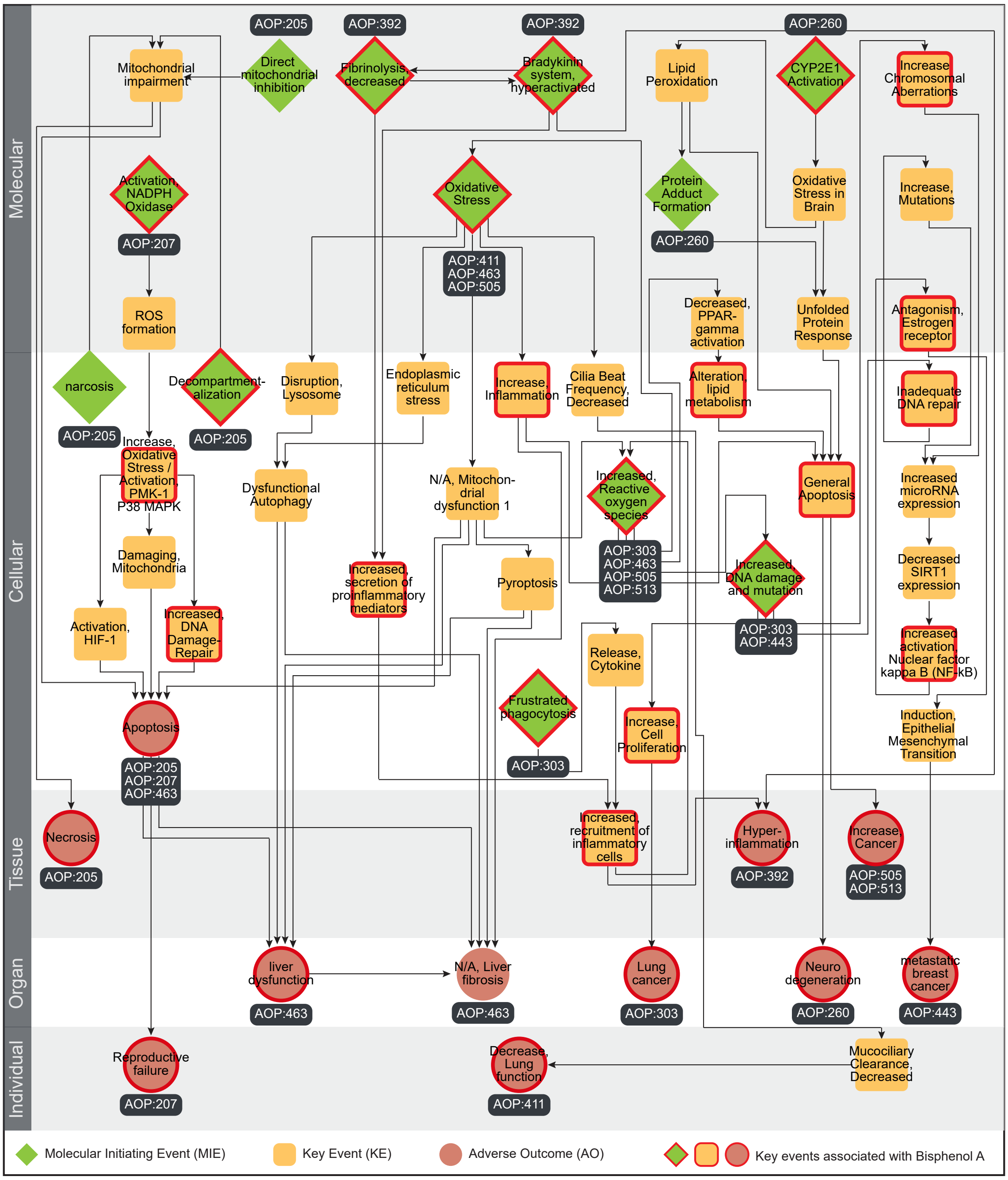

Molecular
Cellular
Tissue
Organ
Individual
AOP:205
AOP:392
AOP:392
AOP:260
Mitochondrial impairment
Direct mitochondrial inhibition
Fibrinolysis, decreased
Bradykinin system, hyperactivated
Lipid Peroxidation
CYP2E1 Activation
Increase Chromosomal Aberrations
Activation, NADPH Oxidase
AOP:207
Oxidative Stress
Protein Adduct Formation
AOP:260
Oxidative Stress in Brain
Increase, Mutations
AOP:411
AOP:463
AOP:505
ROS formation
Decreased, PPAR-gamma activation
Unfolded Protein Response
Antagonism, Estrogen receptor
narcosis
AOP:205
Decompartment-alization
AOP:205
Disruption, Lysosome
Endoplasmic reticulum stress
Increase, Inflammation
Cilia Beat Frequency, Decreased
Alteration, lipid metabolism
Inadequate DNA repair
Increase, Oxidative Stress / Activation, PMK-1 P38 MAPK
Dysfunctional Autophagy
N/A, Mitochon-drial dysfunction 1
Increased, Reactive oxygen species
General Apoptosis
Increased microRNA expression
Damaging, Mitochondria
AOP:303
AOP:463
AOP:505
AOP:513
Increased, secretion of proinflammatory mediators
Pyroptosis
Increased, DNA damage and mutation
Decreased SIRT1 expression
Activation, HIF-1
Increased, DNA Damage-Repair
AOP:303
AOP:443
Release, Cytokine
Increased activation Nuclear factor kappa B (NF-kB)
Apoptosis
AOP:205
AOP:207
AOP:463
Frustrated phagocytosis
AOP:303
Increase, Cell Proliferation
Induction, Epithelial Mesenchymal Transition
Necrosis
AOP:205
Increased, recruitment of inflammatory cells
Hyper-inflammation
AOP:392
Increase, Cancer
AOP:505
AOP:513
liver dysfunction
AOP:463
N/A, Liver fibrosis
AOP:463
Lung cancer
AOP:303
Neuro degeneration
AOP:260
metastatic breast cancer
AOP:443
Reproductive failure
AOP:207
Decrease, Lung function
AOP:411
Mucociliary Clearance, Decreased
Molecular Initiating Event (MIE)
Key Event (KE)
Adverse Outcome (AO)
Key events associated with Bisphenol A

Figure 5.14: Directed network corresponding to the largest component in the undirected BPA-AOP network comprising 55 KEs and 72 KERs. Among the 55 KEs, 12 are categorized as MIEs (denoted as diamond), 11 are categorized as AOs (denoted as circle), and the remaining 32 are categorized as KEs (denoted as rounded square). The 31 KEs (including MIEs and AOs) associated with BPA are marked in 'red'. In this figure, the 55 KEs are arranged vertically according to their level of biological organization.

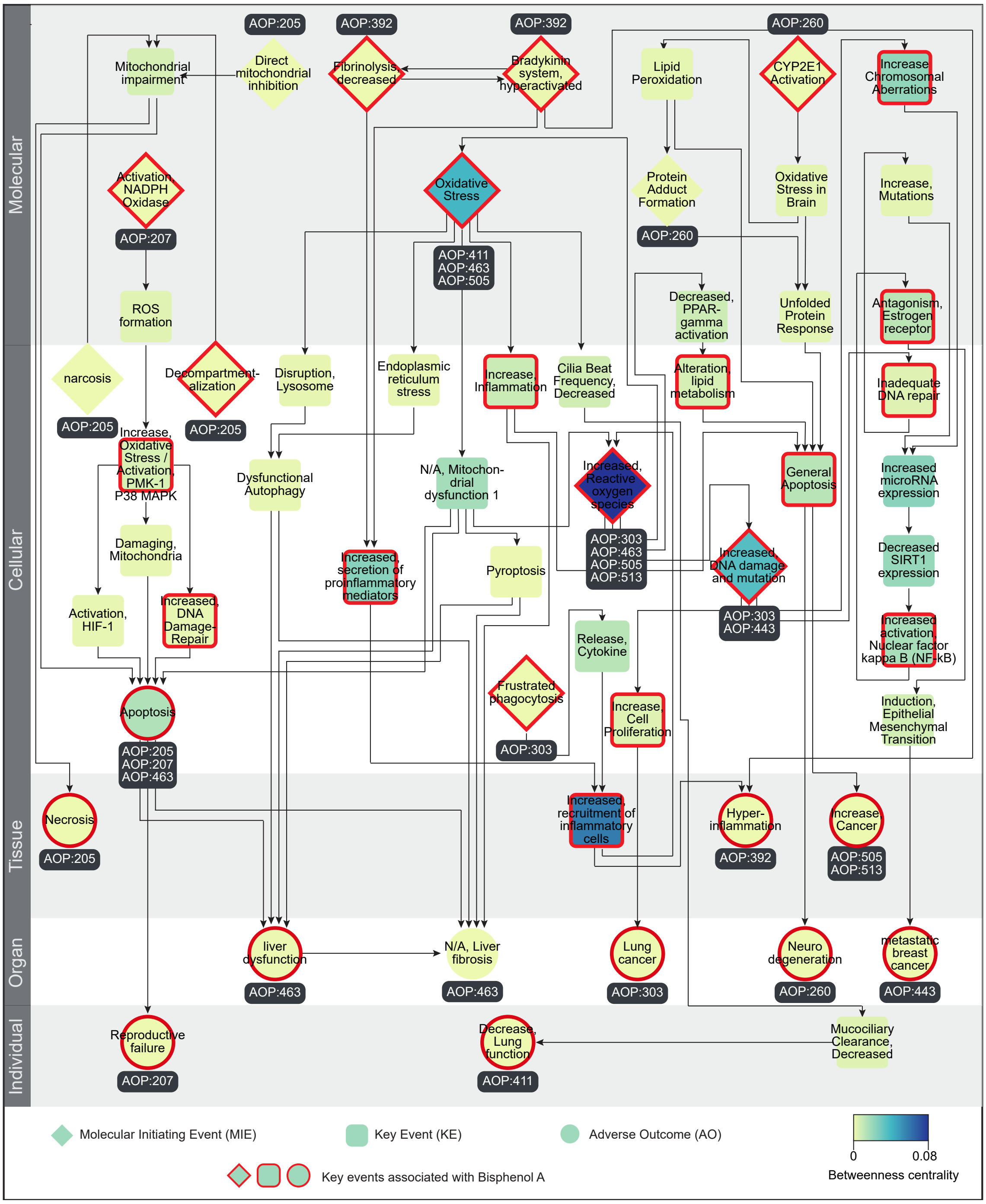


Figure 5.15: Directed network corresponding to the LCC (C1) in the BPA-AOP network, where the KEs (including MIEs and AOs) are colored based on their betweenness centrality values. The 31 KEs (including MIEs and AOs) associated with BPA are marked in 'red'. In this figure, the 55 KEs are arranged vertically according to their level of biological organization.

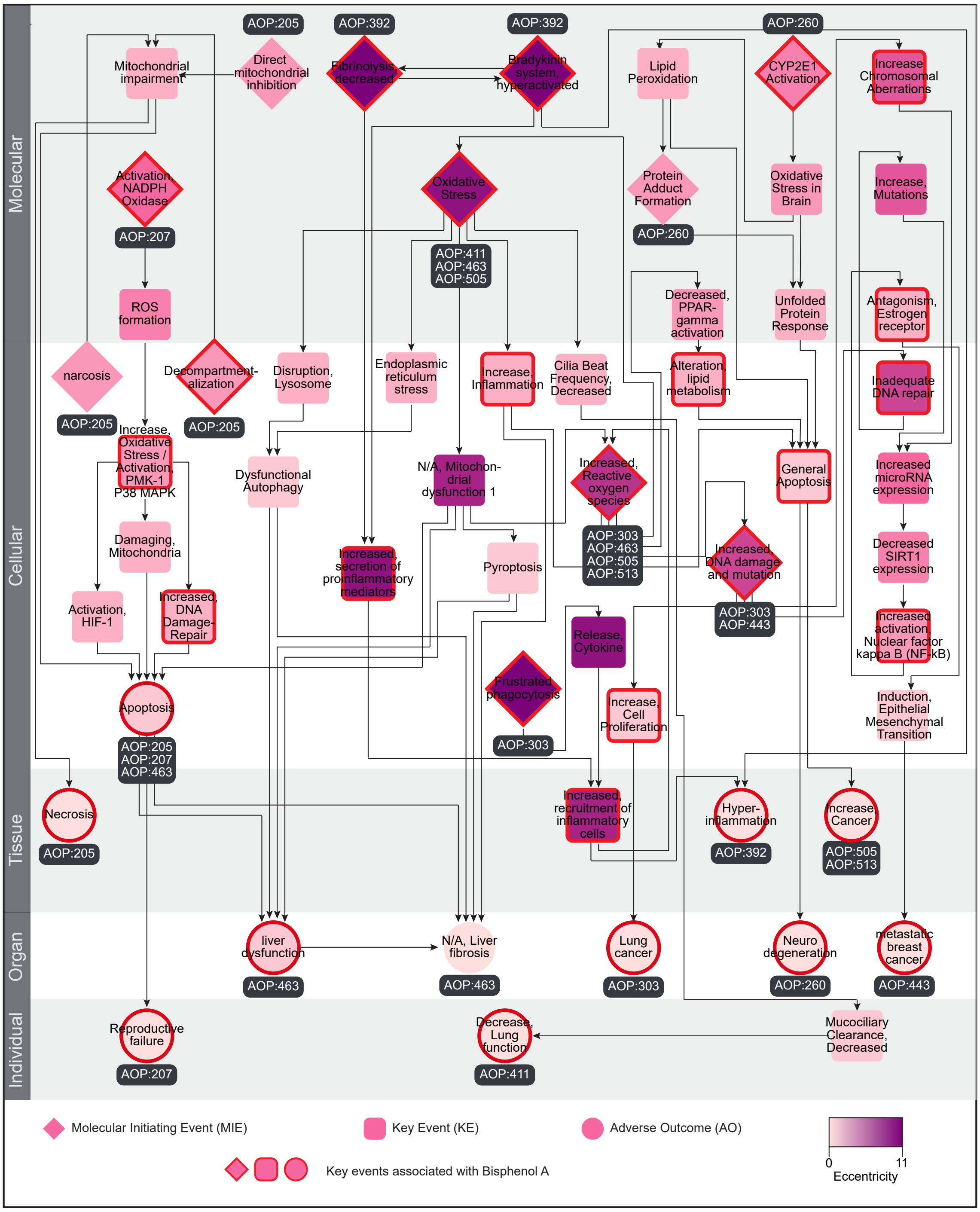

Molecular
Cellular
Tissue
Organ
Individual
AOP:205
AOP:392
AOP:392
AOP:260
Mitochondrial impairment
Direct mitochondrial inhibition
Fibrinolysis, decreased
Bradykinin system, hyperactivated
Lipid Peroxidation
CYP2E1 Activation
Increase Chromosomal Aberrations
Activation, NADPH Oxidase
AOP:207
Oxidative Stress
Protein Adduct Formation
AOP:260
Oxidative Stress in Brain
Increase, Mutations
AOP:411
AOP:463
AOP:505
ROS formation
Decreased, PPAR-gamma activation
Unfolded Protein Response
Antagonism, Estrogen receptor
narcosis
Decompartment-alization
Disruption, Lysosome
Endoplasmic reticulum stress
Increase, Inflammation
Cilia Beat Frequency, Decreased
Alteration, lipid metabolism
Inadequate DNA repair
AOP:205
AOP:205
Increase, Oxidative Stress / Activation, PMK-1 P38 MAPK
Dysfunctional Autophagy
N/A, Mitochon-drial dysfunction 1
Increased, Reactive oxygen species
General Apoptosis
Increased microRNA expression
Damaging, Mitochondria
Increased, secretion of proinflammatory mediators
Pyroptosis
AOP:303
AOP:463
AOP:505
AOP:513
Increased, DNA damage and mutation
Decreased SIRT1 expression
Activation, HIF-1
Increased, DNA Damage-Repair
AOP:303
AOP:443
Release, Cytokine
Increased activation, Nuclear factor kappa B (NF-kB)
Frustrated phagocytosis
Apoptosis
Increase Cell Proliferation
Induction, Epithelial Mesenchymal Transition
AOP:205
AOP:207
AOP:463
AOP:303
Necrosis
AOP:205
Increased recruitment of inflammatory cells
Hyper-inflammation
AOP:392
Increase, Cancer
AOP:505
AOP:513
liver dysfunction
AOP:463
N/A, Liver fibrosis
AOP:463
Lung cancer
AOP:303
Neuro degeneration
AOP:260
metastatic breast cancer
AOP:443
Reproductive failure
AOP:207
Decrease, Lung function
AOP:411
Mucociliary Clearance, Decreased
Molecular Initiating Event (MIE)
Key Event (KE)
Adverse Outcome (AO)
Key events associated with Bisphenol A
0
11
Eccentricity

Figure 5.16: Directed network corresponding to the LCC (C1) in the BPA-AOP network, where the KEs (including MIEs and AOs) are colored based on their eccentricity values. The 31 KEs (including MIEs and AOs) associated with BPA are marked in 'red'. In this figure, the 55 KEs are arranged vertically according to their level of biological organization.

BPA-induced neurodegenerative effects.

BPA has been shown to induce the cytochrome P450 enzyme CYP2E1 in rat liver and kidney cells in both *in vitro* and *in vivo* studies [279, 280]. Although there is no direct evidence of BPA inducing CYP2E1 in the brain, BPA is known to cross the blood-brain barrier and cause adverse effects [281]. In an earlier study, Valencia-Olvera *et al.* [282] identified that xenobiotic stress triggers CYP2E1 expression in cerebellar granule neurons, leading to oxidative stress through the generation of reactive oxygen species (ROS). This oxidative stress can activate apoptotic pathways through unfolded protein response in these neurons, which ultimately results in neurodegeneration [283, 284]. In conclusion, by leveraging various published evidence, we were able to explore a potential toxicity pathway that links BPA-induced toxicity with neurodegeneration in humans.

Toxicity pathway linking BPA exposure to reproductive failure in aquatic species

To understand the ecotoxicological effects of BPA exposure, we investigated the toxicity pathway outlined in the BPA-AOP 'NADPH oxidase and P38 MAPK activation leading to reproductive failure in Caenorhabditis elegans' (AOP:207). Although this AOP is documented as taxonomically applicable to *C. elegans* (Supplementary Table S5.14), we identified auxiliary evidence supporting this toxicity pathway in aquatic species. Various *in vitro* assays have demonstrated that BPA exposure can lead to oxidative stress due to increased ROS production through the activation of NADPH oxidase (NOX) [285,286]. Additionally, BPA exposure has been observed to cause a significant increase in p38 MAPK levels in zebrafish ovarian cells [287]. This increase triggers ovarian inflammation and activates apoptotic pathways in zebrafish oocytes, ultimately leading to reduced reproductive success [287]. In conclusion, by leveraging various published evidence, we were able to explore a potential toxicity pathway that links BPA-induced toxicity with reproductive failure in aquatic organisms.

5.2.7 Exploration of toxicity pathways in AOP network constructed from Bis(2-ethylhexyl) phthalate-relevant AOPs

Similar to the construction of B[a]P-AOP network, we constructed the AOP network for Bis(2-ethylhexyl) phthalate (CAS:117-81-7, commonly known as diethylhexyl phthalate or DEHP) toxicity. We identified 19 highly relevant AOPs associated with DEHP-induced toxicities (Table 5.1), which we designated as DEHP-AOPs. Among the 19 DEHP-AOPs, we observed that 6 have 'High' cumulative WoE and 6 have 'Moderate' cumulative WoE (Supplementary Table S5.17). In the undirected AOP network constructed from these 19 DEHP-AOPs (Figure 5.17), we observed four connected components (with two or more AOPs) and four isolated nodes, where the largest component (labeled C1) comprises 5 DEHP-AOPs.

We observed that the 5 DEHP-AOPs in the LCC C1 (AOP:263, AOP:264, AOP:265, AOP:267, and AOP:268) have 'Decrease, Coupling of oxidative phosphorylation' (KE:1446) as MIE and 'Decrease, Growth' (KE:1521) as AO. Notably, these 5 AOPs were developed by a single research group to address ecotoxicological effects on growth and development of organisms, and thus can be considered as potential alternate strategies to understand the same toxicity pathway [95]. The connected components C2 and C3 are the next largest components, each containing 4 DEHP-AOPs. Therefore, we constructed and visualized the directed AOP networks to explore the interactions among DEHP-AOPs present in both C2 and C3 connected components.

We observed that the directed AOP network of the connected component C2 comprised 20 unique KEs (including 4 MIEs and 5 AOs) and 29 unique KERs (Figure 5.18; Supplementary Table S5.18). Among the 20 KEs, 11 KEs were associated with DEHP-induced toxicity through our systematic data integrative approach, of which 2 are MIEs and 4 are AOs (Figure 5.18). Then we computed several node-centric network measures for this directed network. We observed that the AO 'N/A, Liver fibrosis' (KE:344) has the highest in-degree of 5, while the KEs 'N/A, Mitochondrial dysfunction 1' (KE:177)

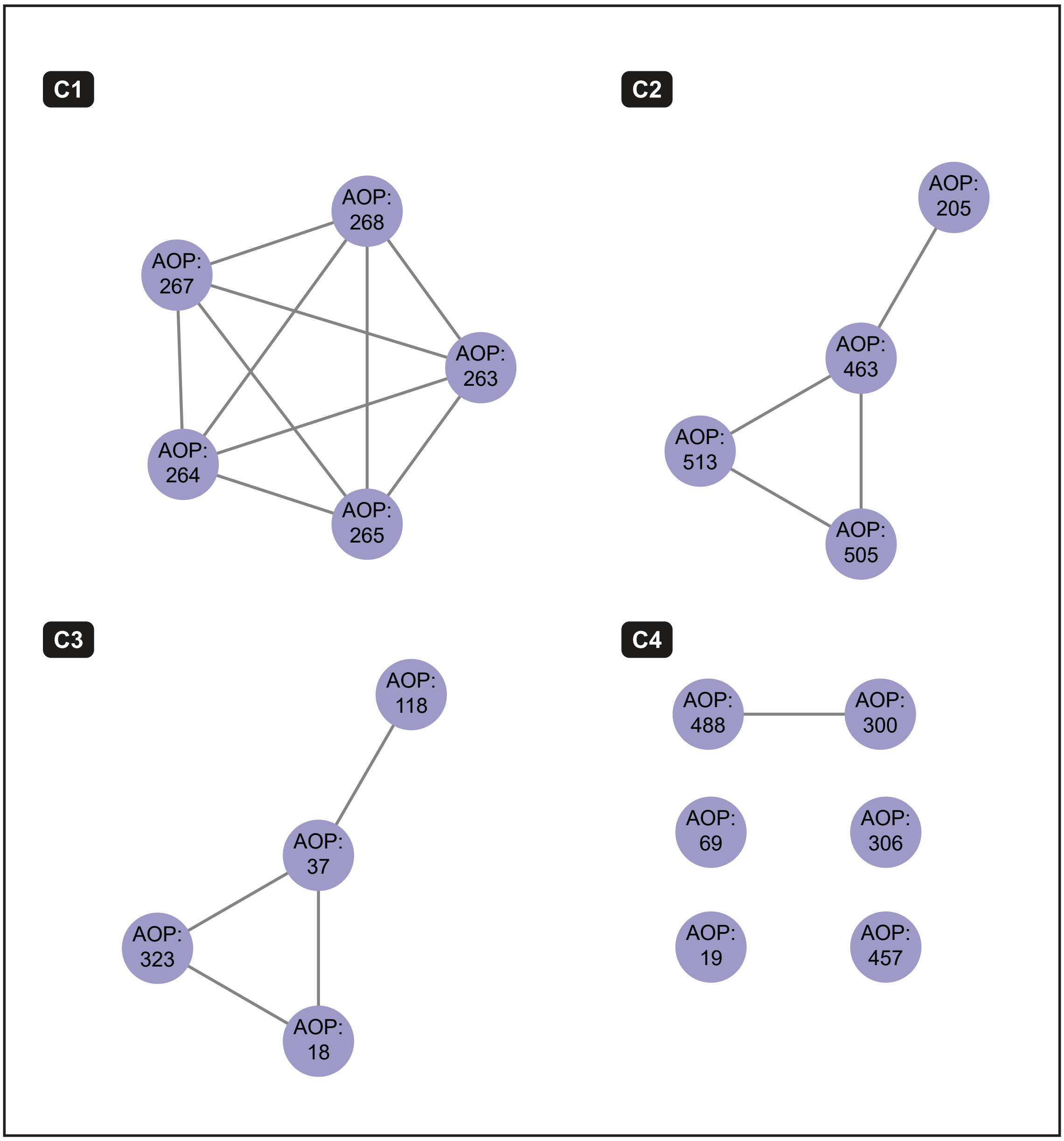


Figure 5.17: Undirected network of Bis(2-ethylhexyl) phthalate (DEHP)-AOPs. Each node corresponds to a DEHP-AOP and an edge between two nodes denotes that the two AOPs share at least one KE. This undirected network has 4 connected components (with two or more nodes) which are labeled as C1, C2, C3 and C4, and 4 isolated nodes.

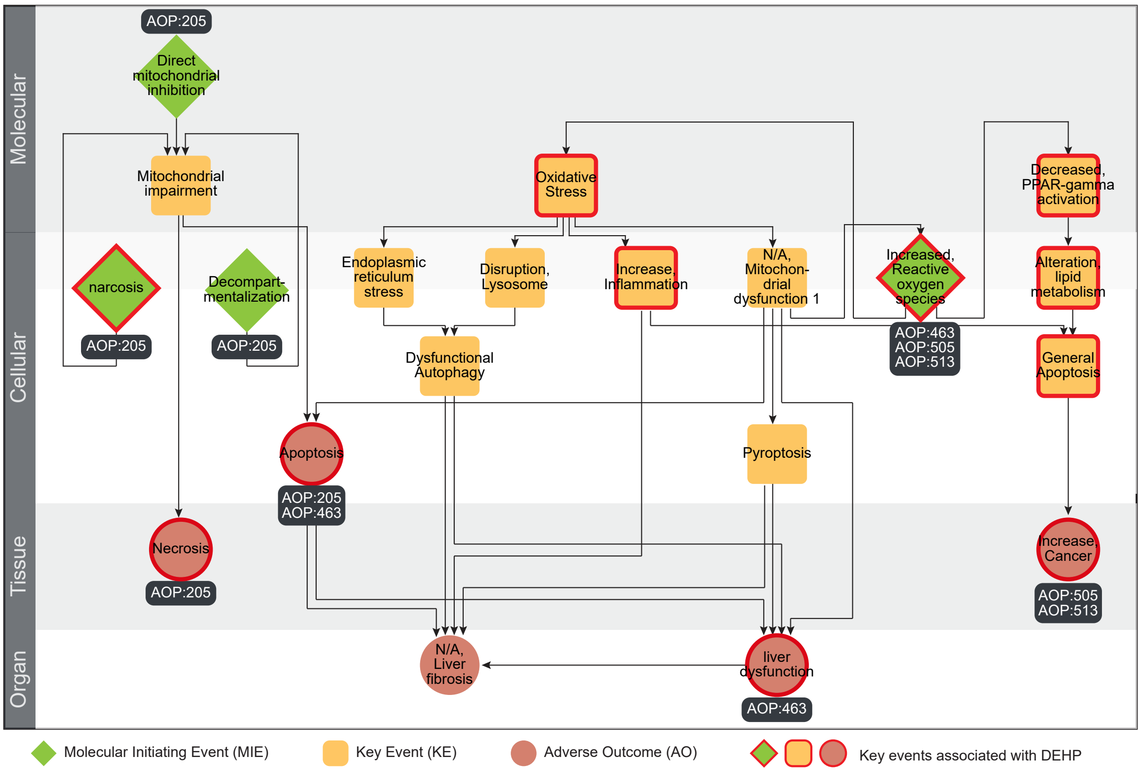


Figure 5.18: Directed network corresponding to the connected component (C2) in the undirected DEHP-AOP network comprising 20 KEs and 29 KERs. Among the 20 KEs, 4 are categorized as MIEs (denoted as diamond), 5 are categorized as AOs (denoted as circle), and the remaining 11 are categorized as KEs (denoted as rounded square). The 11 KEs (including MIEs and AOs) associated with DEHP are marked in 'red'. In this figure, the 20 KEs are arranged vertically according to their level of biological organization.

and 'Oxidative Stress' (KE:1392) have the highest out-degree of 4 (Supplementary Table S5.18). The KE 'Oxidative Stress' (KE:1392) has the highest betweenness centrality value, denoting that several toxicity pathways are passing through it in this network (Figure 5.19; Supplementary Table S5.18). The 'N/A, Mitochondrial dysfunction 1' (KE:177) has the highest eccentricity, denoting that it is the most remotely placed KE in this network (Figure 5.20; Supplementary Table S5.18). Finally, following a similar approach taken in the B[a]P-AOP directed network analysis, we compiled auxiliary evidence from published literature for all these 20 KEs (Supplementary Table S5.19), and performed a case study to explore human-relevant DEHP-induced toxicity pathway from this directed DEHP-AOP network.

We further constructed and visualized the directed AOP network of the connected

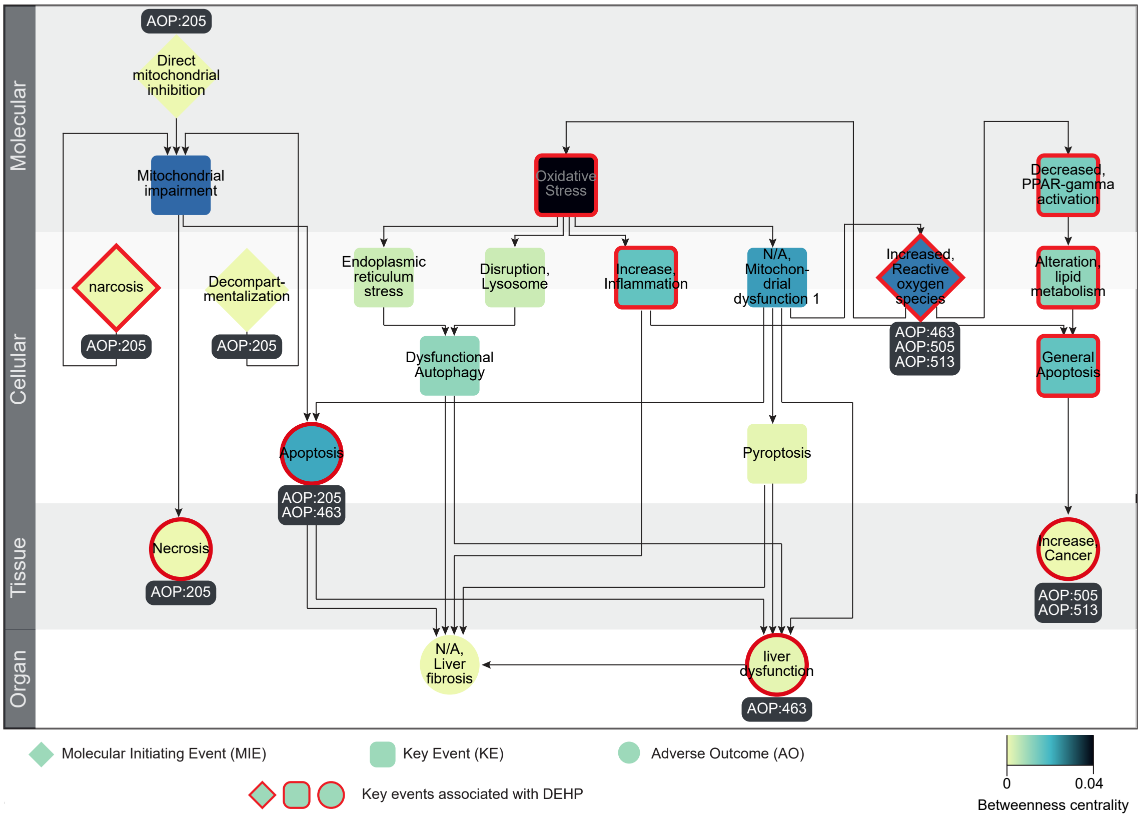


Figure 5.19: Directed network corresponding to the connected component (C2) in the DEHP-AOP network, where the KEs (including MIEs and AOs) are colored based on their betweenness centrality values. The 11 KEs (including MIEs and AOs) associated with DEHP are marked in 'red'. In this figure, the 20 KEs are arranged vertically according to their level of biological organization.

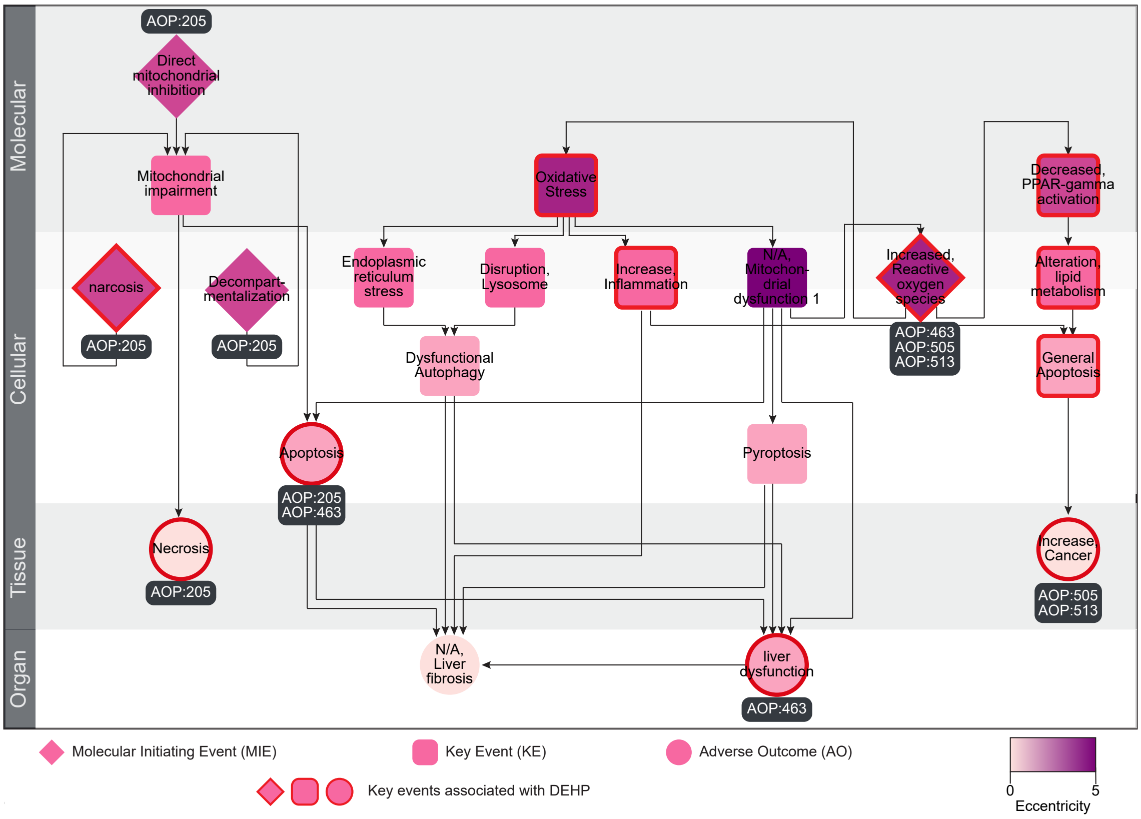


Figure 5.20: Directed network corresponding to the connected component (C2) in the DEHP-AOP network, where the KEs (including MIEs and AOs) are colored based on their eccentricity values. The 11 KEs (including MIEs and AOs) associated with DEHP are marked in ‘red’. In this figure, the 20 KEs are arranged vertically according to their level of biological organization.

component C3 and observed that it comprised 20 unique KEs (including 2 MIEs and 5 AOs) and 23 unique KERs (Figure 5.21; Supplementary Table S5.20). Among the 20 KEs, 14 KEs were associated with DEHP-induced toxicity through our systematic data integrative approach, of which 2 are MIEs and 3 are AOs (Figure 5.21). Then we computed several node-centric network measures for this directed network. We observed that the KE 'Increase, hepatocellular adenomas and carcinomas' (KE:719) has the highest in-degree of 4, while the MIE 'Activation, PPARα' (KE:227) has the highest out-degree of 6 (Supplementary Table S5.20). The KEs 'Reduction, Cholesterol transport in mitochondria' (KE:447) and 'Reduction, Testosterone synthesis in Leydig cells' (KE:413) have the highest betweenness centrality value, denoting that several toxicity pathways are passing through them in this network (Figure 5.22; Supplementary Table S5.20). The MIE 'Activation, PPARα' (KE:227) has the highest eccentricity, denoting that it is the most remotely placed KE in this network (Figure 5.22; Supplementary Table S5.20). Finally, following a similar approach taken in the B[a]P-AOP directed network analysis, we compiled auxiliary evidence from published literature for all these 20 KEs (Supplementary Table S5.21), and performed a case study to explore ecotoxicology-relevant DEHP-induced toxicity pathway from this directed DEHP-AOP network.

Toxicity pathway linking DEHP exposure to liver dysfunction in humans

DEHP exposure has been associated with hepatotoxicity in humans, but the underlying toxicity pathway is not well understood [288, 289]. Thus, we focused on the AOP titled 'The AOP framework on silica nanoparticles induced hepatoxicity' (AOP:463) to understand the toxicity pathway associated with DEHP-induced hepatotoxicity. Previous studies have shown that DEHP induces oxidative stress in rat hepatocytes due to excessive production of ROS [288–290]. This increased oxidative stress in hepatocytes induces pathways underlying mitochondrial dysfunction and inflammation [288,289]. Such pathways damage the hepatocytes, eventually leading to liver dysfunction in the DEHP-exposed rats [288,289]. In conclusion, by leveraging various published evidence, we were

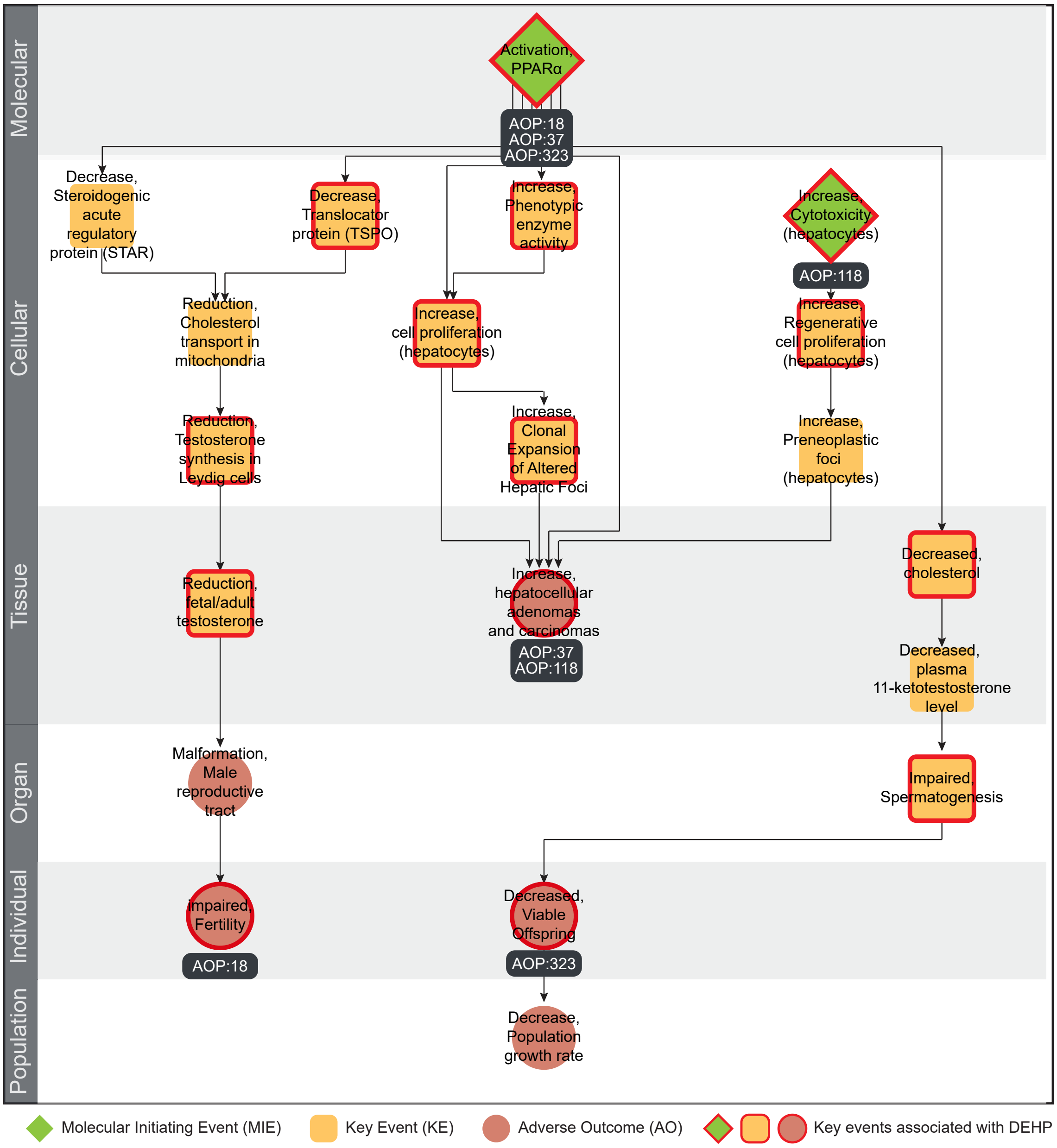

Molecular
Cellular
Tissue
Organ
Individual
Population
Activation, PPARα
AOP:18
AOP:37
AOP:323
Decrease, Steroidogenic acute regulatory protein (STAR)
Decrease, Translocator protein (TSPO)
Increase, Phenotypic enzyme activity
Increase, Cytotoxicity (hepatocytes)
AOP:118
Reduction, Cholesterol transport in mitochondria
Increase, cell proliferation (hepatocytes)
Increase, Regenerative cell proliferation (hepatocytes)
Reduction, Testosterone synthesis in Leydig cells
Increase, Clonal Expansion of Altered Hepatic Foci
Increase, Preneoplastic foci (hepatocytes)
Reduction, fetal/adult testosterone
Increase, hepatocellular adenomas and carcinomas
AOP:37
AOP:118
Decreased, cholesterol
Decreased, plasma 11-ketotestosterone level
Malformation, Male reproductive tract
Impaired, Spermatogenesis
Impaired, Fertility
AOP:18
Decreased, Viable Offspring
AOP:323
Decrease, Population growth rate
Molecular Initiating Event (MIE)
Key Event (KE)
Adverse Outcome (AO)
Key events associated with DEHP

Figure 5.21: Directed network corresponding to the connected component (C3) in the undirected DEHP-AOP network comprising 20 KEs and 23 KERs. Among the 20 KEs, 2 are categorized as MIEs (denoted as diamond), 5 are categorized as AOs (denoted as circle), and the remaining 13 are categorized as KEs (denoted as rounded square). The 14 KEs (including MIEs and AOs) associated with DEHP are marked in 'red'. In this figure, the 20 KEs are arranged vertically according to their level of biological organization.

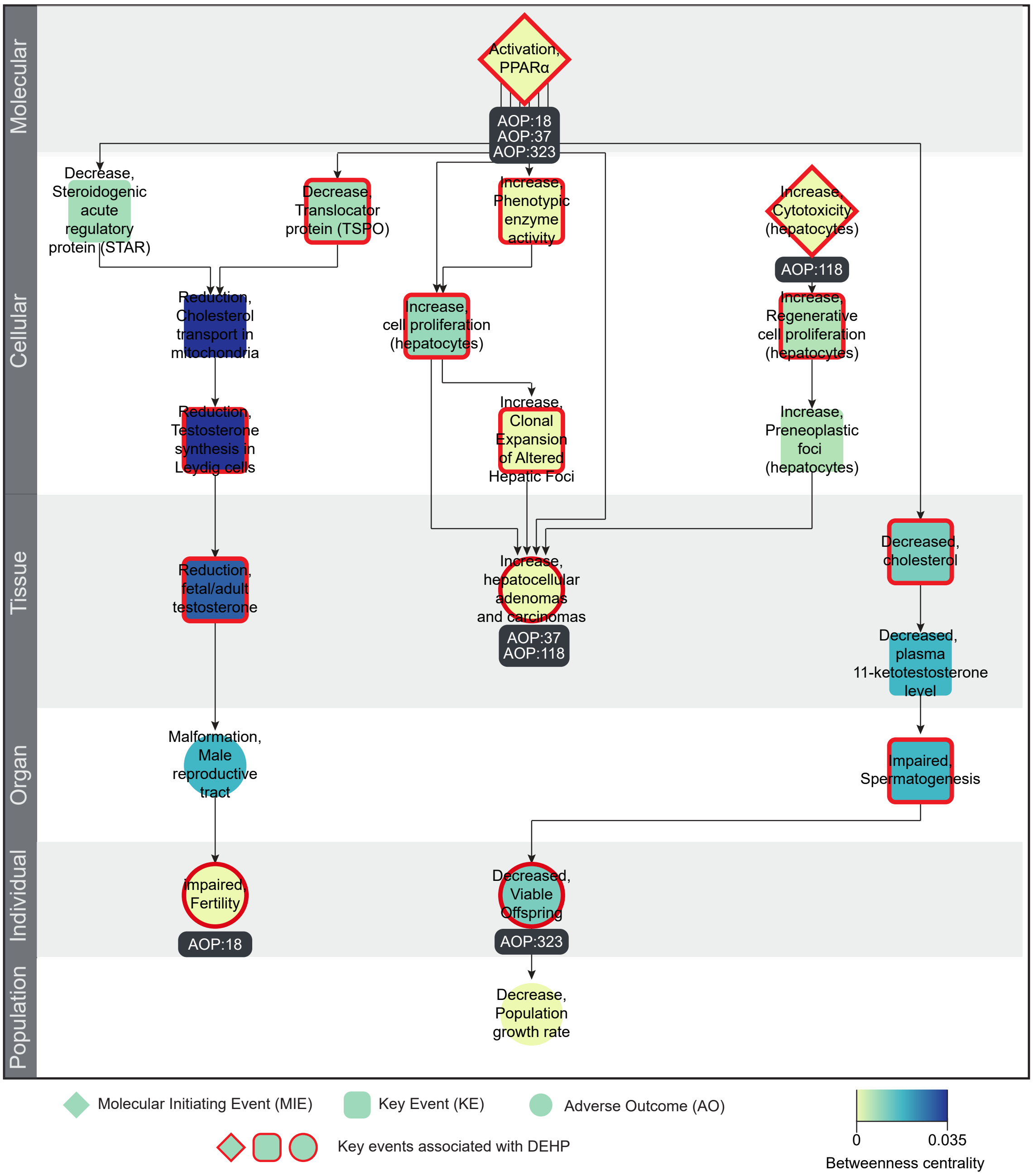

Molecular
Cellular
Tissue
Organ
Individual
Population
Activation, PPARα
AOP:18
AOP:37
AOP:323
Decrease, Steroidogenic acute regulatory protein (STAR)
Decrease, Translocator protein (TSPO)
Increase, Phenotypic enzyme activity
Increase, Cytotoxicity (hepatocytes)
AOP:118
Reduction, Cholesterol transport in mitochondria
Increase, cell proliferation (hepatocytes)
Increase, Regenerative cell proliferation (hepatocytes)
Reduction, Testosterone synthesis in Leydig cells
Increase, Clonal Expansion of Altered Hepatic Foci
Increase, Preneoplastic foci (hepatocytes)
Reduction, fetal/adult testosterone
Increase, hepatocellular adenomas and carcinomas
AOP:37
AOP:118
Decreased, cholesterol
Decreased, plasma 11-ketotestosterone level
Malformation, Male reproductive tract
Impaired, Spermatogenesis
Impaired Fertility
AOP:18
Decreased, Viable Offspring
AOP:323
Decrease, Population growth rate
Molecular Initiating Event (MIE)
Key Event (KE)
Adverse Outcome (AO)
Key events associated with DEHP
0
0.035
Betweenness centrality

Figure 5.22: Directed network corresponding to the connected component (C3) in the DEHP-AOP network, where the KEs (including MIEs and AOs) are colored based on their betweenness centrality values. The 14 KEs (including MIEs and AOs) associated with DEHP are marked in 'red'. In this figure, the 20 KEs are arranged vertically according to their level of biological organization.

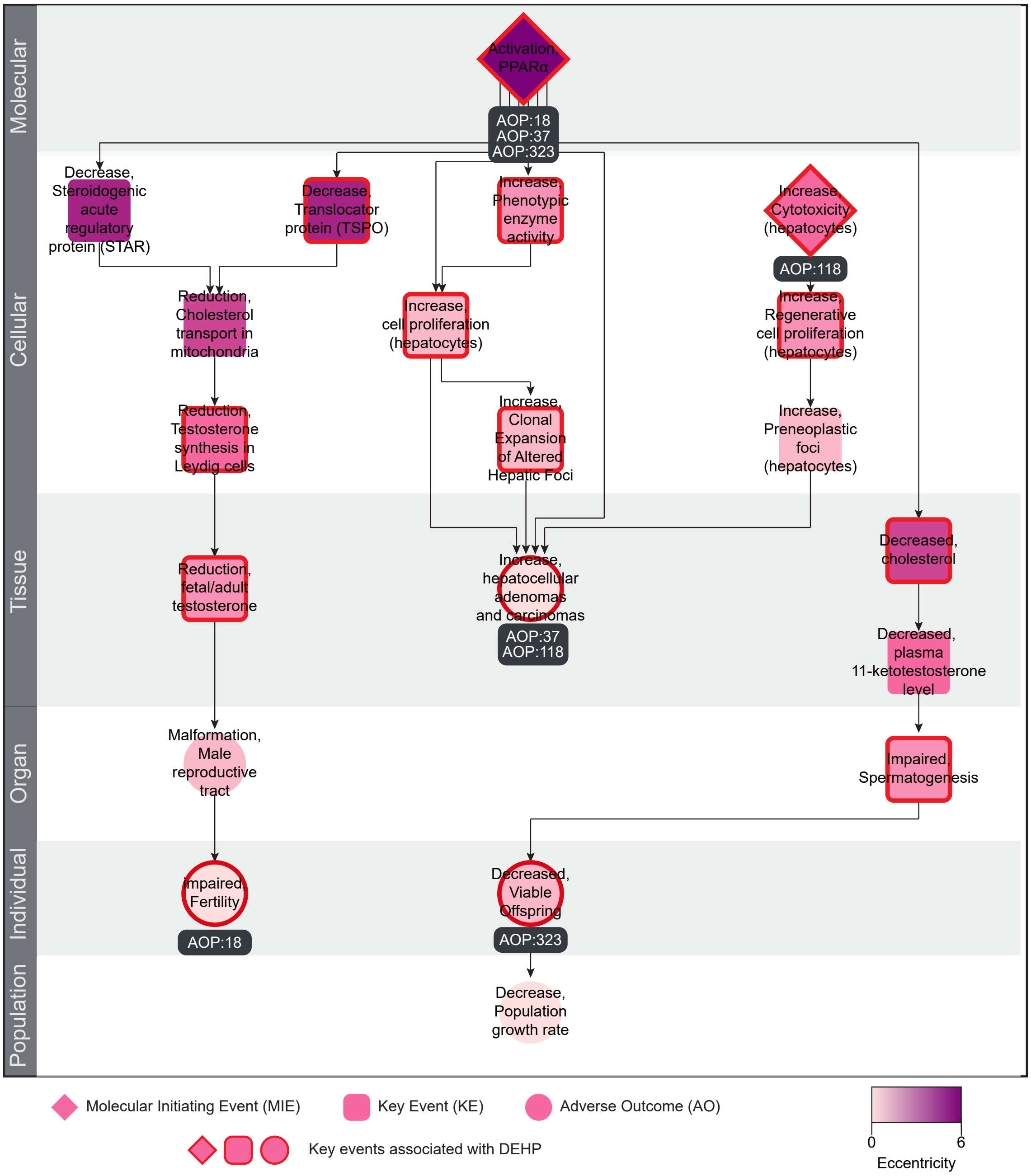

Molecular
Cellular
Tissue
Organ
Individual
Population
Activation, PPARα
AOP:18
AOP:37
AOP:323
Decrease, Steroidogenic acute regulatory protein (STAR)
Decrease, Translocator protein (TSPO)
Increase, Phenotypic enzyme activity
Increase, Cytotoxicity (hepatocytes)
AOP:118
Reduction, Cholesterol transport in mitochondria
Increase, cell proliferation (hepatocytes)
Increase, Regenerative cell proliferation (hepatocytes)
Reduction, Testosterone synthesis in Leydig cells
Increase, Clonal Expansion of Altered Hepatic Foci
Increase, Preneoplastic foci (hepatocytes)
Reduction, fetal/adult testosterone
Increase, hepatocellular adenomas and carcinomas
AOP:37
AOP:118
Decreased, cholesterol
Decreased, plasma 11-ketotestosterone level
Malformation, Male reproductive tract
Impaired, Spermatogenesis
Impaired, Fertility
AOP:18
Decreased, Viable Offspring
AOP:323
Decrease, Population growth rate
Molecular Initiating Event (MIE)
Key Event (KE)
Adverse Outcome (AO)
Key events associated with DEHP
0
6
Eccentricity

Figure 5.23: Directed network corresponding to the connected component (C3) in the DEHP-AOP network, where the KEs (including MIEs and AOs) are colored based on their eccentricity values. The 14 KEs (including MIEs and AOs) associated with DEHP are marked in 'red'. In this figure, the 20 KEs are arranged vertically according to their level of biological organization.

able to explore a potential toxicity pathway that links DEHP-induced toxicity with liver dysfunction in humans.

Toxicity pathway linking DEHP exposure to decreased population growth in aquatic species

To understand the ecotoxicological effects of DEHP exposure, we investigated the toxicity pathway outlined in the DEHP-AOP titled ‘PPARalpha Agonism Leading to Decreased Viable Offspring via Decreased 11-Ketotestosterone’ (AOP:323) having a cumulative WoE of ‘High’ (Supplementary Table S5.17) and taxonomical relevance to teleost fish (Supplementary Table S5.17). In DEHP-exposed fish, DEHP is metabolized into monoethylhexyl phthalate (MEHP), which preferentially binds to and activates the PPARα receptor [291–293]. The activation of PPARα promotes lipid catabolism in the heart and cholesterol uptake in the liver, leading to an overall reduction in cholesterol levels [291]. Golshan *et al.* [294] observed a significant reduction in 11-ketotestosterone (11-KT) levels in DEHP-treated goldfish. Similar reductions in 11-KT levels have been shown to impair spermatogenesis in male zebrafish, significantly affecting the viability of offspring and consequently hindering population growth [295]. In conclusion, by leveraging various published evidence, we were able to explore a potential toxicity pathway that links DEHP-induced toxicity with decreased population growth in aquatic species.

5.3 Discussion

Plastic additives are chemicals that are potentially released into the environment from various plastic products. The lack of information on their presence in various plastic products pose a challenge in evaluating their risks, thereby hindering proper regulatory measures. Towards this, the UNEP has published a detailed report [20] on chemicals found in plastic and their functions, compiled from two independent published studies [226,227]. Though, the reported functional annotations of the chemicals may be inaccurate [226], the UNEP report stands as one of the most comprehensive and largest resources on chemicals doc-

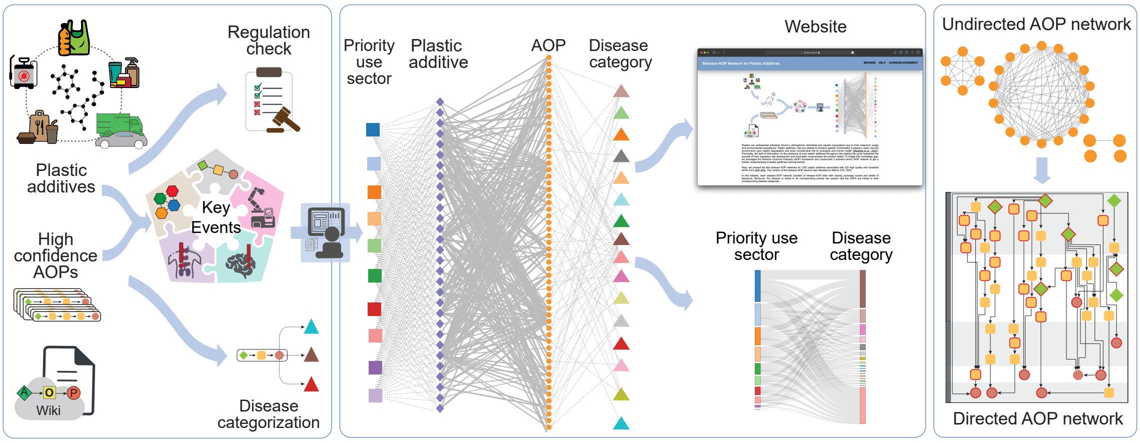


Figure 5.24: Schematic summary of our construction and analysis of stressor-AOP network for plastic additives.

umented to be found in plastic. We curated the list of plastic additives by leveraging the reported functional annotations and observed that many of these plastic additives are produced in high volumes globally and are documented to cause endocrine disruptions. Notably, we observed that these additives can accumulate in various human tissues as xenobiotics, suggesting that prolonged exposure to plastic additives can lead to highly deleterious effects in different organ systems [21, 296].

We utilized toxicogenomics and biological endpoints data from various exposome-relevant resources and identified novel associations between the plastic additives and AOPs which were not otherwise documented in AOP-Wiki (Figure 5.24). Additionally, we introduced two criteria namely, level of relevance and coverage score to characterize the plastic additive-AOP associations in the constructed stressor-AOP network which enabled identification of highly relevant AOPs associated with each plastic additive. From the stressor-AOP network, we noted that plastic additives used in various sectors can potentially induce a wide range of diseases with cancer being the most represented category, followed by gastrointestinal system disease. To reiterate, this chapter introduces stressor-AOP network for plastic additives (Figure 5.24) in addition to the undirected and directed AOP networks, similar to those reported in Chapter 4 for cadmium, providing a holistic understanding of the toxicities induced by plastic additives. The stressor-AOP network for each plastic additive can be visualized on the dedicated website: https://cb.imsc.res.in/saopadditives/.

However, the toxicogenomics based data integrative approach reported in this chapter primarily relies on the mammalian-centric biological data, thus limiting its scope in exploring various ecotoxicological events associated with plastic additives. Nonetheless, we present the first and most comprehensive stressor-AOP network for plastic additives which facilitates their risk assessment, thereby contributing towards a toxic-free circular economy for plastics.

Supplementary Information

Supplementary Tables S5.1-S5.21 associated with this chapter are available for download from the GitHub repository: https://github.com/asamallab/PhDThesis-Ajaya_Kumar_Sahoo/blob/main/SI/ST_Chapter5.xlsx.

Code Availability

The computer programs used to perform the computations reported in this chapter are available in the following GitHub repository:
https://github.com/asamallab/PhDThesis-Ajaya_Kumar_Sahoo/tree/main/Codes.

Plastic additive	**Functions**	**Number of highly relevant stressor-AOP links**	**Number of stressor-AOP links not present in AOP-Wiki**
Benzo[a]pyrene (CAS:50-32-8)	Plasticizers, Cross-linkers, Lubricants, Fillers	28	28
Bisphenol A (CAS:80-05-7)	Catalysts, Cross-linkers, Fillers, Antioxidants, Light stabilizers, Lubricants, Blowing agents, Plasticizers, Colorants, Flame retardants, Antistatic agents	27	26
Bis(2-ethylhexyl) phthalate (CAS:117-81-7)	Cross-linkers, Fillers, Light stabilizers, Fragrances, Plasticizers, Colorants	19	17
Arsenic (CAS:7440-38-2)	Biocides, Cross-linkers, Fillers, Colorants	16	13
Ethanol (CAS:64-17-5)	Biocides, Catalysts, Cross-linkers, Fillers, Light stabilizers, Lubricants, Fragrances, Colorants, Antistatic agents	15	13
Perfluorooctanoic Acid (CAS:335-67-1)	Biocides, Colorants	14	14
Triclosan (CAS:3380-34-5)	Biocides, Light stabilizers, Fragrances, Colorants	14	13
Cadmium (CAS:7440-43-9)	Biocides, Catalysts, Cross-linkers, Fillers, Heat stabilizers, Light stabilizers, Colorants, Pigments	14	11
Cadmium chloride (CAS:10108-64-2)	Biocides, Antioxidants, Heat stabilizers, Light stabilizers, Colorants, Flame retardants	13	12
Lead (CAS:7439-92-1)	Cross-linkers, Fillers, Antioxidants, Heat stabilizers, Light stabilizers, Lubricants, Other stabilizers, Colorants	13	10
Acrylamide (CAS:79-06-1)	Cross-linkers, Fillers, Colorants	13	13
Pentachlorophenol (CAS:87-86-5)	Biocides, Colorants	12	11
Manganese (CAS:7439-96-5)	Biocides, Catalysts, Fillers, Lubricants, Colorants	11	9
Oxygen (CAS:7782-44-7)	Blowing agents, Antioxidants	11	11
Lead acetate (CAS:301-04-2)	Cross-linkers, Flame retardants, Colorants	11	11

Table 5.1: The top 15 plastic additives based on the associated number of highly relevant AOPs in the plastic additives-AOP network with Level 5 stressor-AOP links. For each of the 15 plastic additives, the table also provides the functions cataloged in the annex of the UNEP report.

Chapter 6

Network-based investigation of petroleum hydrocarbons-induced ecotoxicological effects and their risk assessment

Petroleum hydrocarbons (PHs) are compounds consisting mostly of carbon and hydrogen besides other elements that originate from crude oil and its derivatives such as gasoline and diesel, among others [22]. PHs are released into the environment primarily through the diffusion of oils, stemming from anthropogenic activities such as transportation and offshore drilling, as well as accidental incidents like oil spills [297, 298]. Eventually, PHs are absorbed through various exposure routes, including ingestion, dermal contact, and inhalation, where they can bioaccumulate and lead to carcinogenic, developmental, and endocrine toxicities in humans and other species [23–26]. PHs can be broadly classified into aliphatic and aromatic hydrocarbons, with the aromatic PHs being widely studied due to their higher stability, water solubility and environmental persistence [299, 300]. Moreover, the United States Environmental Protection Agency (US EPA) has designated 16 of these polycyclic aromatic hydrocarbons (PAHs) as priority pollutants based on their environmental prevalence and persistence [301–303]. Therefore, identifying PHs and understanding their adverse effects will enable the formulation of effective mitigation and remediation strategies for PH contamination.

Accidental oil spillage can contribute to increased environmental PH concentrations causing detrimental impact to diverse ecological habitats. The Deepwater Horizon oil spill in the northern Gulf of Mexico, off the coast of Louisiana, USA [9], is one of the largest oil spills documented to have long-term ecological impacts on fishes, deep ocean corals and oysters, and reduction in the population of marine mammals, sea turtles and seabirds [304, 305]. Similarly, the oil spill in the coastal waters of Ennore near Chennai, India, has been documented to reduce water quality and impact marine biota including phytoplankton, zooplankton, benthic communities and vertebrates like fish and sea turtles [306]. Although the hazards associated with such oil contaminations are often assessed by measuring total petroleum hydrocarbon (TPH) concentrations in the affected environments [307, 308], there is a lack of studies focusing on the risks posed by individual PHs.

In Chapters 4 and 5, we investigated toxicities induced by inorganic cadmium and plastic additives, respectively, by leveraging the adverse outcome pathway (AOP) framework. Specifically, we utilized heterogeneous data from various exposome-relevant resources to construct different networks of AOPs which primarily facilitated the exploration of chemical-induced adverse health effects in humans. In other words, the scope of these two studies was limited in exploration of the ecotoxicological events as our data integrative approach primarily relied on mammalian-centric biological datasets. In this chapter, we undertake a network-based investigation of PH-induced toxicities among ecological species, and moreover, perform an ecological risk assessment. We integrate ecotoxicologically-relevant biological endpoint data corresponding to the PHs from various toxicological resources and construct different networks of AOPs. Notably, we construct stressor-species networks and derive hazard concentrations for a group of PHs in aquatic environment. **The work reported in this chapter is contained in the manuscript**

[100].

6.1 Methods

6.1.1 Compilation and curation of PHs

Organic compounds originating from crude oil, and consisting mostly of carbon and hydrogen atoms are termed as PHs [22]. TPHs refer to the total recoverable concentrations of PHs measured in an environmental sample [22]. The Total Petroleum Hydrocarbon Criteria Working Group (TPHCWG), consisting of representatives from industry, academia and government, had convened to provide extensive technical information relevant for risk assessment of PHs in petroleum contaminated sites [309]. TPHCWG compiled their findings and data into a series of reports dealing with analytical methods for quantifying TPH, composition of various fuel mixtures, fate and transport of TPH, fraction-based reference dose and reference concentration development, and framework for human health risk-based evaluation of PH contaminated sites.

In this study, we relied on TPHCWG report Volume 2 [98], titled ‘Composition of Petroleum Mixtures’ and TPHCWG report Volume 3 [97], titled ‘Selection of Representative TPH Fractions Based on Fate and Transport Considerations’, to retrieve PHs present in 11 petroleum mixtures namely, gasoline, diesel, kerosene, number 2 fuel oil, number 6 fuel oil, JP-4, JP-5, JP-7, JP-8, lubricating and motor oils, and crude oil. We mapped these retrieved chemicals to standardized chemical information from PubChem [145] and Chemical Abstracts Service (CAS) [228] and subsequently compiled a list of 320 unique PHs (Supplementary Table S6.1). The workflow for identifying these 320 PHs is described in Figure 6.1a. Finally, for each of these identified PHs, we obtained their corresponding chemical structures from PubChem, and employed RDKit [123] to classify them as aliphatic, monocyclic aromatic, or polycyclic aromatic hydrocarbons (PAHs) (Supplementary Table S6.1). Further, we employed ClassyFire [110] to obtain additional chemical categorizations, namely, chemical Kingdom, Superclass and Class (Supplementary Table S6.1).

6.1.2 Identification of ecotoxicologically-relevant ‘high confidence’ AOPs within AOP-Wiki

AOP-Wiki [53] is the largest publicly accessible global repository that systematically catalogs all the developed AOPs till date, including the scientific evidence supporting these AOPs at various levels. Therefore, we relied on AOP-Wiki to retrieve the AOPs. First, we downloaded the XML file (released on 1 January 2024) from ‘Project Downloads’ page in AOP-Wiki, and using an in-house python script, we extracted various information associated with the AOPs like their title, identifier, associated key events (KEs) (including molecular initiating events - MIEs and adverse outcomes - AOs), key event relationships (KERs), stressors, Organisation for Economic Co-operation and Development (OECD) and Society for the Advancement of Adverse Outcome Pathways (SAAOP) status, biological applicability information such as taxonomy, sex, life-stage of the organism, and the weight of evidence (WoE). The AOP documentation within AOP-Wiki is constantly updated based on novel understanding and experimental information as and when they are available, and thus are considered as living documents [171]. Therefore, to assess the quality and completeness of the AOP data available within AOP-Wiki, we followed a systematic workflow developed in our previous work (Figure 6.2) [96].

We first manually checked and removed ‘archived’ AOPs, AOPs that lacked any KEs or KERs, and AOPs comprising undefined KEs. Then, we employed NetworkX [173] to identify the directed paths between MIE(s) and AO(s) within each AOP, and removed AOPs if they were disconnected. Finally, through this extensive manual and computational effort, we identified 328 non-empty, connected, complete and high quality AOPs, which we designate as ‘high confidence’ AOPs.

Next, we followed the workflow proposed by Jagiello *et al.* [310] to identify AOPs relevant for ecotoxicology based on their associated taxonomic applicability. We first fil-

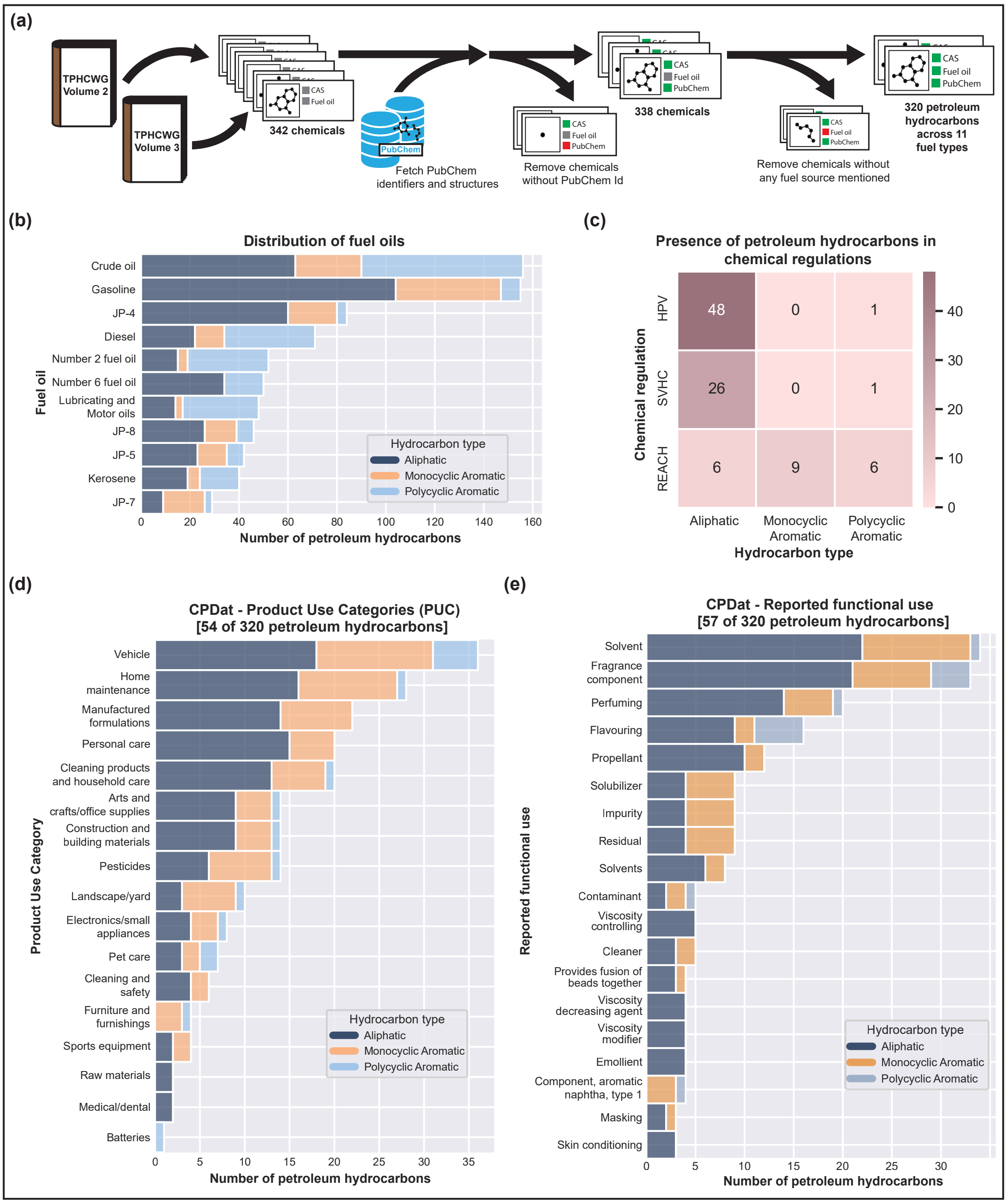


Figure 6.1: Curation and exploration of PHs. (a) Workflow followed to curate a list of 320 PHs from published literature. (b) Distribution of different types of PHs across various fuel oils. (c) Heatmap depicting the presence of PHs in chemical regulations. The number of different types of PHs in each of the chemical regulations is denoted in the heatmap. (d) Distribution of different types of PHs across various Product Use Categories (PUCs) from CPDat. (e) Distribution of different types of PHs across various functional uses reported in CPDat which are associated with at least three PHs.

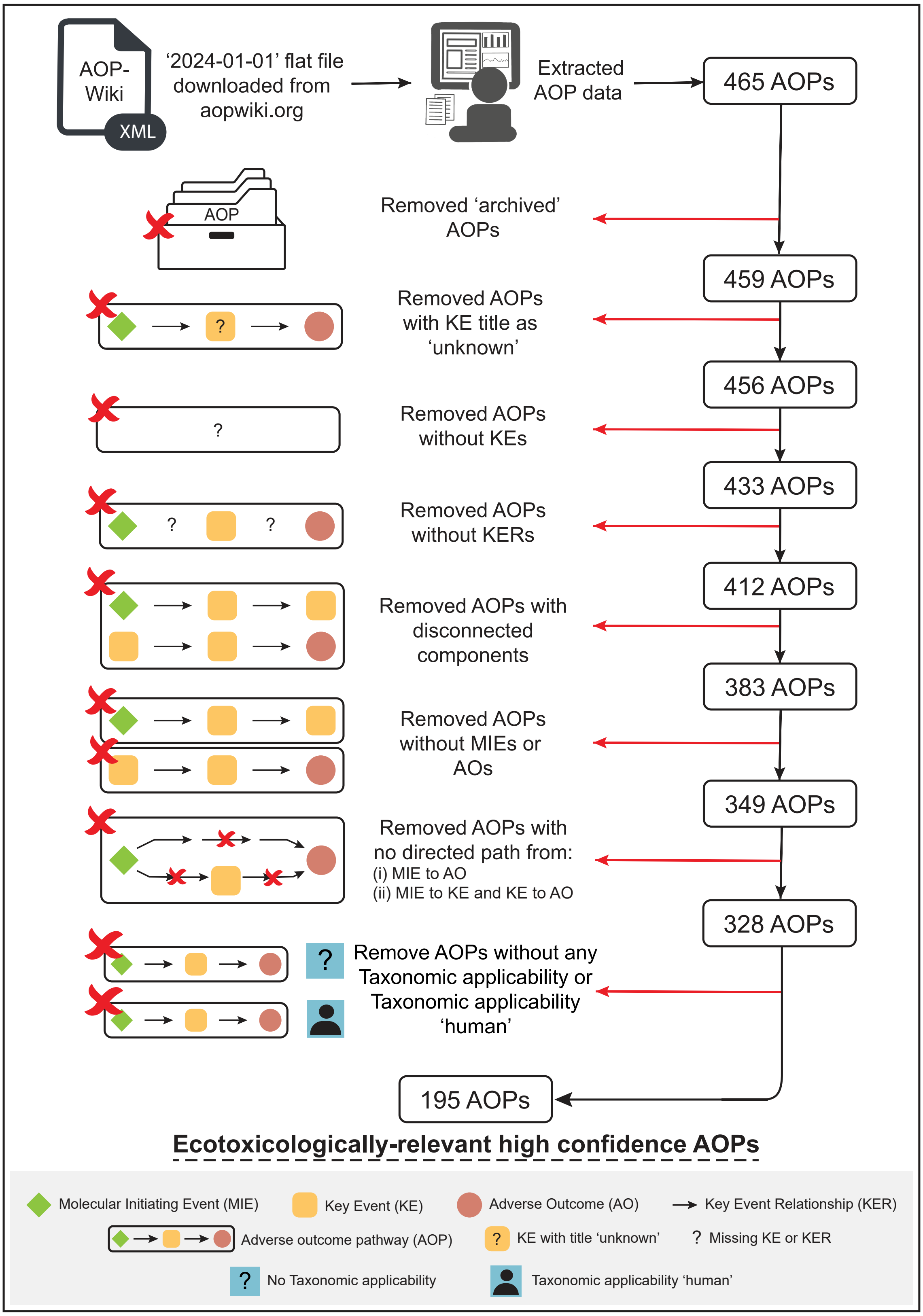

AOP-Wiki
XML
'2024-01-01' flat file downloaded from aopwiki.org
Extracted AOP data
465 AOPs
AOP
Removed 'archived' AOPs
459 AOPs
?
Removed AOPs with KE title as 'unknown'
456 AOPs
?
Removed AOPs without KEs
433 AOPs
? ?
Removed AOPs without KERs
412 AOPs
Removed AOPs with disconnected components
383 AOPs
Removed AOPs without MIEs or AOs
349 AOPs
Removed AOPs with no directed path from:
(i) MIE to AO
(ii) MIE to KE and KE to AO
328 AOPs
?
Remove AOPs without any Taxonomic applicability or Taxonomic applicability 'human'
195 AOPs
Ecotoxicologically-relevant high confidence AOPs
Molecular Initiating Event (MIE)
Key Event (KE)
Adverse Outcome (AO)
Key Event Relationship (KER)
Adverse outcome pathway (AOP)
? KE with title 'unknown'
? Missing KE or KER
? No Taxonomic applicability
Taxonomic applicability 'human'

Figure 6.2: Workflow to filter ecotoxicologically-relevant high confidence AOPs from AOP-Wiki by employing computation and manual curation in conjunction.

tered out the high confidence AOPs that lacked any taxonomic applicability information. Then, we filtered out the AOPs if their taxonomic applicability contained only terms related to humans. Through this systematic process, we identified 195 of the 328 high confidence AOPs to be relevant for ecotoxicology, which we designate as 'ecotoxicologically-relevant AOPs'.

Supplementary Table S6.2 contains the list of 195 ecotoxicologically-relevant AOPs obtained through our systematic workflow. These 195 ecotoxicologically-relevant AOPs comprise 727 unique KEs (Supplementary Table S6.3) and 1047 unique KERs (Supplementary Table S6.4).

6.1.3 Identification of ecotoxicologically-relevant KEs associated with PHs

In this study, we aimed to analyze the ecotoxicity of the PHs through the AOP framework. To achieve this, we first identified the ecotoxicologically-relevant KEs associated with PHs from three different sources namely, ToxCast [16], Comparative Toxicogenomics Database (CTD) [27] and ECOTOX [28]. Here, we note that AOP-Wiki did not catalog any of the PH as a prototypical stressor, and therefore we did not rely on AOP-Wiki to obtain KEs associated with PHs.

Using ToxCast

The US EPA's chemical prioritization program, ToxCast, provides several high-throughput screening assay data points to assess the toxicity of several thousand environmental chemicals [16]. Based on our previous works [95, 96], we followed a systematic pipeline to identify the KEs associated with the curated PHs by relying on their ToxCast assay endpoints.

Briefly, we accessed the ToxCast invitrodb version 4.1 dataset and identified active assay endpoints ('hitc' $\geq$ 0.9) [179] associated with PHs from the 'mc5-6_winning_model_fits-flags_invitrodb_v4_1_SEPT2023.csv' file. We additionally identified the ‘activatory’ or ‘inhibitory’ response of the PHs based on the ‘top’ value of the corresponding winning model from the ‘mc4_all_model_fits_invitrodb_v4_1_SEPT2023.csv’ file [179]. Next, we relied on the ‘cytotox_invitrodb_v4_1_SEPT2023.xlsx’ file to identify and discard the cytotoxicity-associated bursts associated with PHs, as they result from non-specific reporter gene activations associated with cell stress and cytotoxicity [96, 179, 236].

Finally, to identify the ecotoxicologically-relevant ToxCast assay endpoints, we relied on the ‘assay_annotations_invitrodb_v4_1_SEPT2023.xlsx’ file to obtain the organism information associated with the assay endpoints. We removed chemical-assay endpoint pairs related to humans. Through this effort, we identified 390 ecotoxicologically-relevant ToxCast assay endpoints which are associated with 56 PHs. Next, we relied on the assay annotations for these ToxCast endpoints to manually inspect and assess endpoints that can be potentially linked to KEs. Through this process, we identified 14 KEs from ecotoxicologically-relevant AOPs which are associated with 19 assay endpoints (Supplementary Table S6.5). Briefly, ToxCast provides the assay endpoints, associated genes and ‘activatory’ or ‘inhibitory’ information for every tested chemical. We leveraged this information and manually mapped it to the KEs within AOP-Wiki based on their title, object identifier and object name [96].

Using CTD

CTD [27] compiles data on chemical-gene/protein, chemical-phenotype, chemical-disease and gene-disease associations (or links) from published literature. Based on our previous work [95,96], we utilized chemical (C), gene (G), phenotype (P) and disease (D) tetramers (CGPD-tetramers) within CTD data to identify the ecotoxicologically-relevant KEs associated with the curated PHs. To identify the gene-phenotype associations, we relied on the GO term annotations from the NCBI Gene resource [175] (last accessed on 22 April 2024).

We downloaded the March 2024 release of the CTD data and retrieved CGPD-tetramers associated with the curated PHs. Here, we observed that CTD additionally provides information on the organisms associated with every chemical-gene and chemical-phenotype link. Therefore, to identify ecotoxicologically-relevant CGPD-tetramers, we removed CGPD-tetramers where both the chemical-gene and chemical-phenotype links were related to human. Through this process, we identified 8454 ecotoxicologically-relevant CGPD-tetramers comprising 18 PHs, 904 genes, 210 phenotypes and 148 diseases (Supplementary Table S6.6). Additionally, we employed the GOSim package [177] in R programming language to identify the neighboring GO terms of phenotypes. Finally, we manually inspected phenotypes (including their neighbor terms) and diseases, and identified that 90 KEs are associated with 49 phenotypes and 35 KEs are associated with 54 diseases (Supplementary Table S6.5). The mapped phenotypes were associated with 18 PHs while the mapped diseases were associated with 15 PHs.

Using ECOTOX

US EPA's ECOTOX [28] is one of the largest ecotoxicological knowledgebases that has compiled manually curated ecotoxicology information for more than 12000 chemicals across more than 13000 terrestrial and aquatic species. Importantly, ECOTOX relies on a standardized pipeline to extract various toxicity data from published literature, including the biological effects observed in organisms after exposure to environmental chemicals. Therefore, we utilized the ECOTOX dataset to identify the KEs associated with the curated PHs.

First, we downloaded the latest ECOTOX dataset (released in March 2024) by selecting the 'Download ASCII Data' option on the ECOTOX website. Then, we used an in-house python script to parse and extract data associated with the curated PHs from 'tests.txt', 'results.txt', 'chemicals.txt' and 'species.txt' files. We observed that ECOTOX provides curated data points such as biological effects associated with the chemical (Effect), the parameter that measures the corresponding biological effect (Measurement) and the trend (Trend) of this parameter with respect to a control. Here, we removed data points where the Measurement value is ‘Not Reported’ or is empty, as they lack information on the adverse effects caused by such chemicals. Finally, we manually inspected the Effect, Measurement and Trend parameters, and identified that 101 KEs are associated with 74 PHs (Supplementary Table S6.5).

Overall, we identified 206 ecotoxicologically-relevant KEs associated with 75 out of the 320 PHs in our curated list through biological endpoints data from three sources: ToxCast, CTD and ECOTOX.

6.1.4 Construction and visualization of stressor-AOP network for PHs

Stressor-AOP network can capture a diverse array of AOPs linked to stressors, and thus, provide a comprehensive understanding of the adverse effects induced by a chemical. To understand the adverse effects induced by the PHs, we constructed a bipartite graph between the PHs and the ecotoxicologically-relevant AOPs, where PHs are linked to an AOP through an associated KE. Further, we characterized the stressor-AOP links between PHs and AOPs by computing the coverage score and level of relevance. The coverage score of an AOP is obtained by taking the ratio of the number of KEs which are associated with PHs to the total number of KEs in that AOP. Further, the level of relevance is assessed based on the following five-level criterion:

- *Level 1*: Stressor is associated with the KE of the AOP but not associated with any MIE or AO of the AOP.
- *Level 2*: Stressor is associated with the AO of the AOP but not associated with any MIE of the AOP.
- *Level 3*: Stressor is associated with the MIE of the AOP but not associated with any AO of the AOP.
- *Level 4*: Stressor is associated with both MIE and AO of the AOP.
- *Level 5*: Stressor is associated with both MIE and AO of the AOP and there exist a directed path between the associated MIE and AO.

We visualized the stressor-AOP network using Cytoscape [180], where the stressor-AOP link is annotated by the coverage score and the level of relevance. Supplementary Table S6.7 provides the complete stressor-AOP network for the PHs.

6.1.5 Construction and visualization of stressor-species network using ECOTOX data

For a chemical, ECOTOX provides toxicity concentration value at which the biological effect is observed and the bioconcentration factor, across different species. To construct a stressor-species network using the toxicity concentration value, we first retrieved the chemical concentrations for acute toxicity endpoints (LC_{50} and EC_{50}) for the curated list of 320 PHs across different species. Note that, ECOTOX provides the toxicity concentration values in different units of measurement, therefore we standardized these values into their ppm equivalent (mg/L or mg/kg) (Supplementary Table S6.8). In case multiple toxicity concentration values were reported for a given stressor-species pair in ECOTOX, we selected the minimum concentration value for analysis. This process yielded toxicity concentration values for 80 PHs across 221 species (Supplementary Table S6.8). Thereafter we constructed a stressor-species network for these 80 PHs, where the edge between a stressor and a species is represented by the logarithm of the standardized toxicity concentration value. Further, we visualized a subnetwork of this stressor-species network for PAHs from the EPA priority PAHs list [301,311] in Cytoscape [180].

Similarly, to construct the stressor-species network using the bioconcentration factor value, we first retrieved the bioconcentration factor values for the curated list of 320 PHs across different species. Note that bioconcentration factor values for stressor-species pairs are provided in the L/kg unit in ECOTOX. In case multiple bioconcentration factor values were reported for a given stressor-species pair in ECOTOX, we selected the maximum bioconcentration factor value for analysis. This process yielded bioconcentration factor values for 28 PHs across 59 species (Supplementary Table S6.9) for which we constructed a stressor-species network, where the edge between a stressor and a species is represented by the logarithm of the bioconcentration factor value. Finally, we visualized a subnetwork of this stressor-species network for PAHs from the EPA priority PAHs list [301, 311] in Cytoscape [180].

6.1.6 Construction of Species Sensitivity Distributions for PHs

Species Sensitivity Distribution (SSD) is a commonly used tool for performing ecotoxicological risk assessment [312–314]. SSD utilizes statistical distribution of responses to chemical exposure by various species [99] and provides a threshold chemical concentration (HC05) which is hazardous to 5% of species but is not harmful to 95% of species in a particular environment [315, 316]. In order to construct the SSDs for PHs, we followed the steps outlined in the SSD Toolbox technical manual [317] to curate data from ECOTOX (Figure 6.3).

First, we selected toxicity data only for aquatic species by choosing the organism habitat as 'Water' (Figure 6.3). We then retrieved chemical concentration values for acute toxicity endpoints (LC_{50} and EC_{50}) of PHs from ECOTOX along with information on their associated species (Figure 6.3). Next, we standardized these concentration values by converting their corresponding units to ppm equivalents (mg/L or mg/kg). In the case of multiple toxicity values for a given chemical-species pair, we considered the geometric mean [318, 319] of the standardized concentration values and computed the logarithm of this mean (Figure 6.3). To ensure the relevance of the data, we excluded any data points with a mean observation time less than 24 hours or exceeding 96 hours [320], retaining only toxicity concentrations from acute toxicity studies. Finally, we considered PHs with toxicity data reported across species belonging to at least five different ECOTOX species groups to construct SSDs (Figure 6.3) [321].

We relied on two different tools namely, SSD Toolbox [322] developed by the US EPA and R-based ssdtools [318] package developed by the Ministry of Environment and

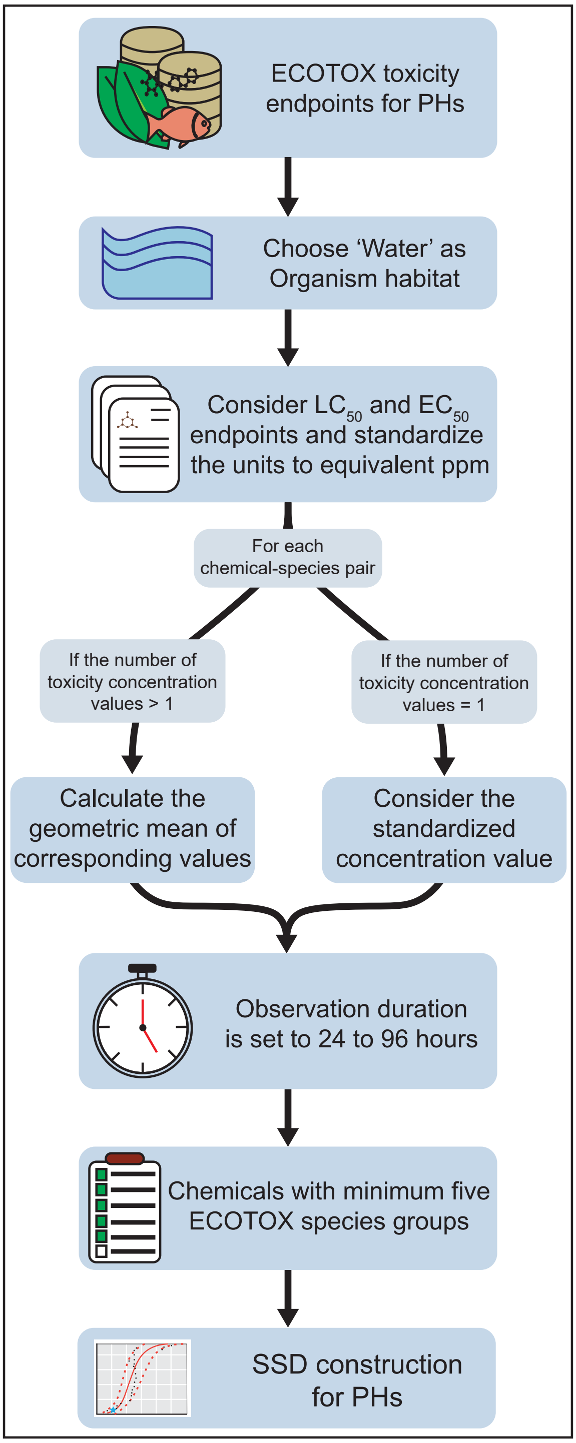

ECOTOX toxicity endpoints for PHs
Choose 'Water' as Organism habitat
Consider LC_{50} and EC_{50} endpoints and standardize the units to equivalent ppm
For each chemical-species pair
If the number of toxicity concentration values > 1
If the number of toxicity concentration values = 1
Calculate the geometric mean of corresponding values
Consider the standardized concentration value
Observation duration is set to 24 to 96 hours
Chemicals with minimum five ECOTOX species groups
SSD construction for PHs

Figure 6.3: Workflow to curate chemical acute toxicity endpoint data from ECOTOX database to construct SSDs for PHs.

Climate Change Strategy of British Columbia to construct SSDs for PHs. In both cases, we employed the maximum likelihood method to fit five statistical distributions, namely, Log-Normal, Log-Logistic, Log-Gumbel, Weibull and Burr Type III to the toxicity concentration data. The SSD Toolbox and ssdtools utilize bootstrap resampling to quantify uncertainty in fitted parameters and estimate confidence intervals [317,318]. In this study, we set the number of bootstrap resampling iterations to 10000.

6.2 Results

6.2.1 Exploration of the curated list of PHs

PHs are a class of organic compounds that are composed mainly of carbon and hydrogen, and originate from crude oils [22]. In this study, we curated a list of 320 PHs that are experimentally detected in 11 different fuel oils including crude oil (Figure 6.1a; Supplementary Table S6.1). For these PHs, we first obtained their chemical structure from PubChem, and then employed RDKit [123] to classify them into aliphatic and aromatic compounds. Among the 320 PHs, we identified 177 as aliphatic hydrocarbons (no aromatic ring), 60 as monocyclic aromatic hydrocarbons (one aromatic ring), and 83 as polycyclic aromatic hydrocarbons (PAHs - more than one aromatic ring) (Supplementary Table S6.1). Further, we explored the distribution of these 320 PHs across the 11 fuel oils and observed that more than 100 PHs are found in two fuel oils namely, crude oil and gasoline (Figure 6.1b).

Next, we checked the presence of these 320 PHs in various chemical regulation lists, namely, the United States High Production Volume (USHPV) [147] chemical list, Organisation for Economic Co-operation and Development High Production Volume (OECD HPV) [146] chemical list, substances of very high concern (SVHC) [126] and REACH prohibited chemicals [125] list (Figure 6.1c). We observed that 49 PHs are documented to be produced in high volumes globally, 27 PHs are substances of very high concern and 21 PHs are prohibited for use under REACH regulation (Figure 6.1c). Moreover, we noted that the majority of HPV and SVHC chemicals among the PHs are classified as aliphatic PHs (Figure 6.1c).

The Chemical and Products Database (CPDat) [148] is a US EPA project that has compiled information on 75000 chemicals and their presence in 15000 consumer products. CPDat provides Product Use Categories (PUCs) and functional use information for these chemicals across products. We leveraged data within CPDat to check the presence of the 320 PHs across various PUCs and their reported functional use information (Figure 6.1d,e). We retrieved PUC data for 54 PHs across 17 categories, with 20 PHs classified under the categories 'Vehicle' and 'Home maintenance' (Figure 6.1d). We also retrieved functional use data for 61 chemicals across 76 different uses, with more than 20 PHs reported to be used as 'solvent' and 'fragrance component' (Figure 6.1e).

The US EPA has identified 16 PAHs as priority pollutants due to their frequent occurrence in environmental samples such as air, water, soil and food, and their potential carcinogenic and mutagenic properties [311, 323–328]. These priority PAHs include Naphthalene, Acenaphthene, 9H-Fluorene, Phenanthrene, Anthracene, Fluoranthene, Pyrene, Benzo[a]anthracene, Chrysene, Benzo[b]fluoranthene, Benzo[k]fluoranthene, Benzo[a]pyrene, Benzo[ghi]perylene, Indeno[1,2,3-cd]pyrene and Dibenzo[a,h]anthracene [311]. We observed that 15 of these 16 priority PAHs are present in the curated list of 320 PHs (Supplementary Table S6.1). Dibenzo[a,h]anthracene was not included in the curated PH list as it did not have an associated fuel oil source.

6.2.2 Stressor-AOP network for PHs

A stressor-AOP network can elucidate different AOPs associated with a stressor of interest, thereby enhancing our understanding of the various adverse biological effects induced by that stressor [93, 96]. In this study, we leveraged the ecotoxicologically-relevant biological endpoints from three different sources namely, CTD [27], ToxCast [16] and ECOTOX [28], and identified 206 KEs from 177 ecotoxicologically-relevant AOPs to be associated with 75 of the 320 PHs (Supplementary Table S6.5). Thereafter we mapped a PH to an ecotoxicologically-relevant AOP if at least one KE within that AOP is associated with the PH. Following this procedure, we identified 3265 PH-AOP associations for 75 PHs and 177 ecotoxicologically-relevant AOPs, and constructed a bipartite stressor-AOP network, which we designate as 'PH-AOP' network (Supplementary Table S6.7). Notably, all the PH-AOP associations in the constructed stressor-AOP network are identified through the systematic data integrative approach followed in this study and none were documented in AOP-Wiki.

Next, we computed the coverage score for all PH-AOP links in the constructed stressor-AOP network and observed that Benzo[a]pyrene (B[a]P or CAS:50-32-8) is associated with all the KEs (coverage score = 1) in two AOPs namely, AOP:30 and AOP:263. Further, we computed the levels of relevance for all the PH-AOP associations and observed that 548 links between 31 PHs and 122 AOPs are classified as Level 1, 2578 links between 75 PHs and 110 AOPs are classified as Level 2, 77 links between 19 PHs and 34 AOPs are classified as Level 3, and 62 links between 10 PHs and 33 AOPs are classified as Level 5 (Supplementary Table S6.7). Note, all the Level 4 links between PHs and AOPs also satisfy Level 5 criterion, and therefore we had no PH-AOP link with Level 4 relevance.

Notably, the constructed PH-AOP network provides 975 stressor-AOP links for 14 priority PAHs and 171 ecotoxicologically-relevant AOPs with varying coverage scores and levels of relevance (Supplementary Table S6.7). Here we observed that 305 links between 14 PAHs and 92 AOPs are classified as Level 1, 591 links between 14 PAHs and 99 AOPs are classified as Level 2, 41 links between 8 PAHs and 24 AOPs are classified as Level 3, and 38 links between 4 PAHs and 31 AOPs are classified as Level 5 (Supplementary Table S6.7). We noted that, B[a]P is associated with the maximum number of ecotoxicologically-relevant AOPs (169), with 29 AOPs identified to have Level 5 stressor-AOP links. Figure 6.4 shows a portion of the PH-AOP network, comprising Level 5 stressor-AOP links for 10 PHs and 33 AOPs, wherein the 4 EPA priority PAHs are marked in red border.

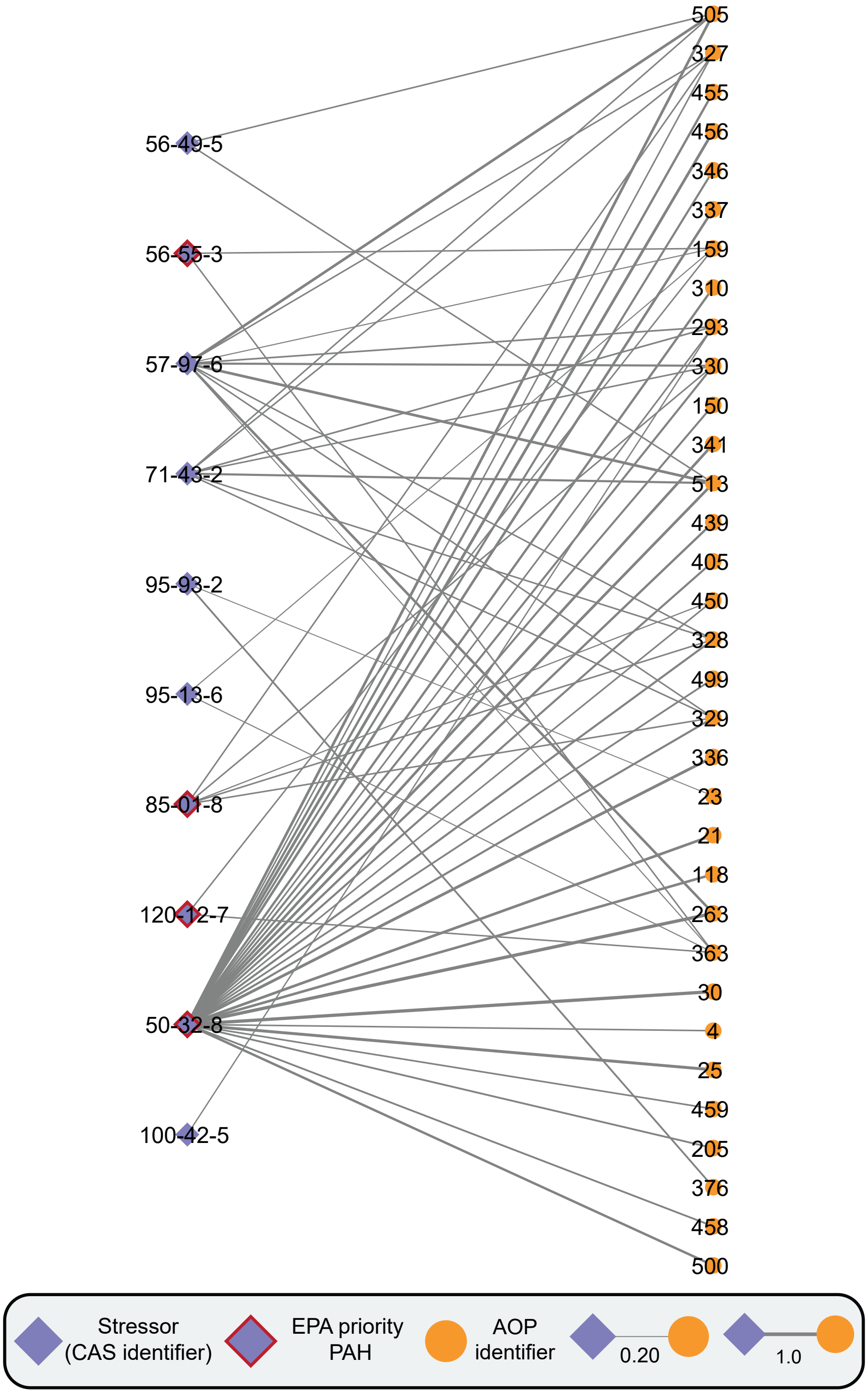


Figure 6.4: Visualization of a stressor-centric AOP network for PHs. In the stressor-AOP network, only edges or stressor-AOP links with Level 5 relevance are shown. The edges in the stressor-AOP network are weighted based on their coverage score, i.e., the fraction of KEs within AOP that are linked with the PHs. Further, the nodes in the stressor-AOP network that correspond to the EPA priority PAHs are highlighted with red borders.

6.2.3 Exploration of ecotoxicologically-relevant pathways in AOP network associated with B[a]P

PHs are known contaminants in soil and aquatic ecosystems. They can persist in the environment and negatively impact various ecological species [329]. Oil spills are frequently documented events that result in the accumulation of these toxic PHs in aquatic environments [330]. B[a]P is a well-documented terrestrial and aquatic pollutant which is found in different fuel oils namely, crude oil, diesel, gasoline, lubricating and motor oils, number 2 fuel oil, and number 6 fuel oil (Supplementary Table S6.1). Therefore, we studied the ecotoxicity induced by B[a]P by identifying the highly relevant AOPs associated with B[a]P (designated as B[a]P-AOPs) in the constructed PH-AOP network. We filtered 29 B[a]P-AOPs by selecting stressor-AOP links with Level 5 relevance and coverage score threshold of 0.4 (i.e., ≥ 0.4) in the constructed PH-AOP network.

Further, we computed cumulative WoE [77, 95] for each of the 29 B[a]P-AOPs and observed that 11 B[a]P-AOPs have 'High' cumulative WoE, 15 B[a]P-AOPs have 'Moderate' cumulative WoE (Supplementary Table S6.10). Thereafter, we analyzed the ecotoxicity induced by B[a]P by constructing and analyzing an undirected AOP network of the 29 B[a]P-AOPs (Figure 6.5). We observed that the B[a]P-AOPs form two connected components (wherein at least two B[a]P-AOPs are connected) namely, C1 and C2, and one isolated node, with the largest connected component (LCC) C1 consisting of 22 B[a]P-AOPs (Figure 6.5).

Next, we constructed a directed network of the 22 B[a]P-AOPs to explore the interaction among these AOPs (Figure 6.6). We observed that the directed network comprises 92 KEs and 125 KERs, wherein 51 KEs (including 9 MIEs and 13 AOs) are associated with B[a]P through integration of biological endpoint data from various sources. Thereafter, we analyzed node-centric properties of the directed network by employing various network measures (Supplementary Table S6.11). The KE 'Altered, Cardiovascular development/function' (KE:317) and the AO 'N/A, Breast Cancer' (KE:1193) have the highest

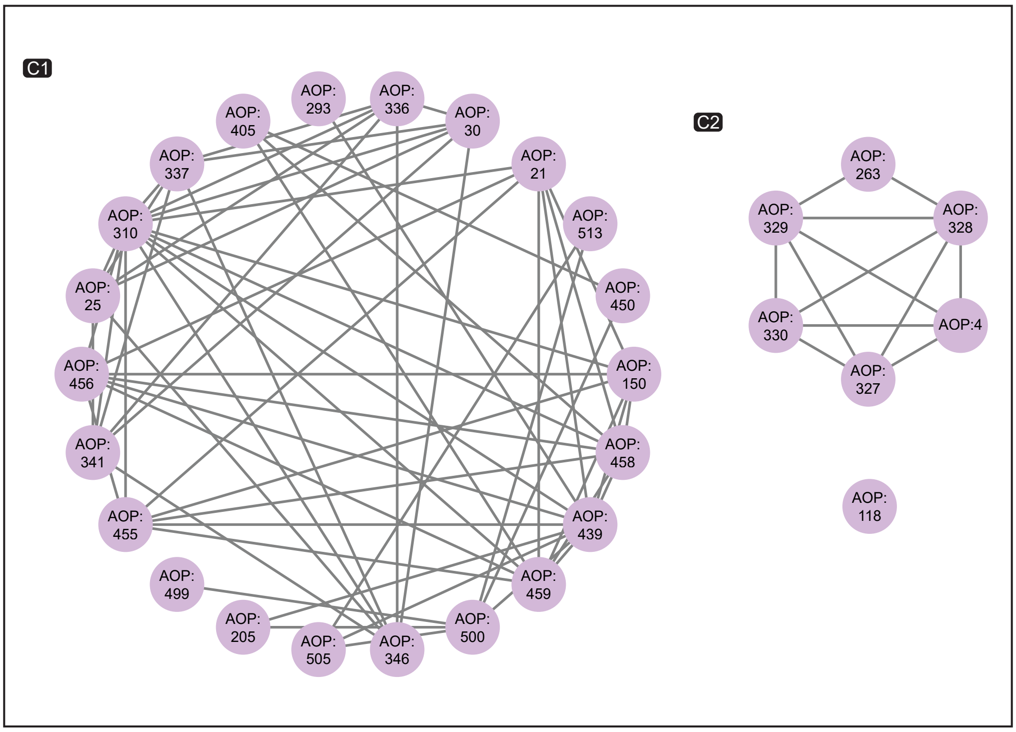


Figure 6.5: Undirected network of B[a]P-AOPs. Each node corresponds to a B[a]P-AOP and an edge between two nodes denotes that the two AOPs share at least one KE. This undirected network has 2 connected components (wherein at least two nodes are connected) which are labeled as C1 and C2, and one isolated node.

in-degree of 5. The MIE 'Activation, AhR' (KE:18) has the highest out-degree of 12. The AO 'Apoptosis' (KE:1262) has the highest betweenness centrality value suggesting that this node is centrally located in the network (Figure 6.7) [59]. The MIE 'Activation, AhR' (KE:18) has the highest eccentricity value suggesting that this node is the farthest node in the network (Figure 6.8) [181]. Further, we utilized Abstract Sifter [185] and AOP-helpFinder, an artificial intelligence (AI) based tool [183, 184], to find associations between B[a]P-induced toxicities and the 92 KEs in the B[a]P-AOP directed network (Supplementary Table S6.11).

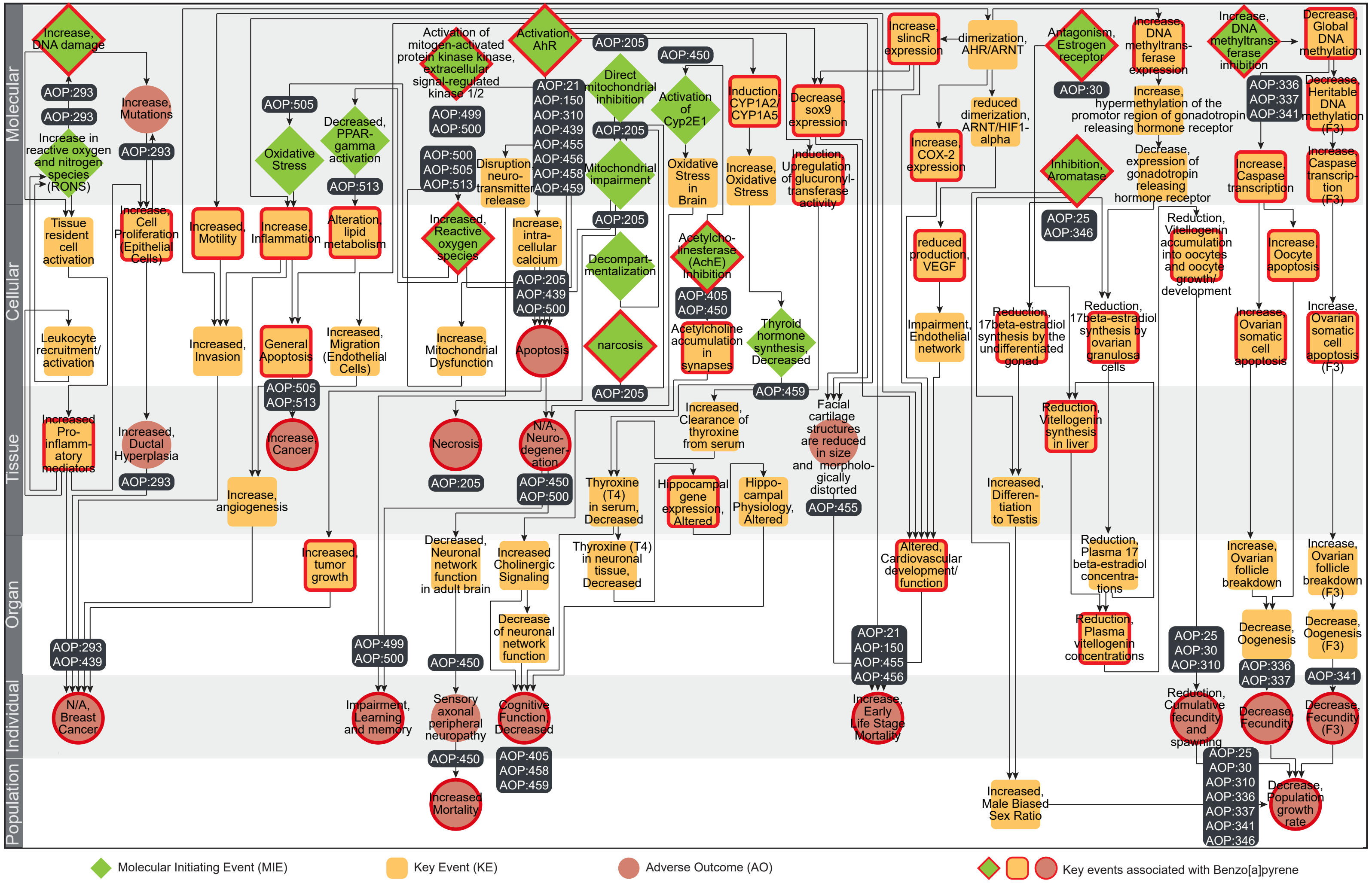

Molecular
Cellular
Tissue
Organ
Individual
Population
Increase, DNA damage
AOP:293
AOP:293
Increase, Mutations
Increase in reactive oxygen and nitrogen species (RONS)
AOP:293
Tissue resident cell activation
Increase, Cell Proliferation (Epithelial Cells)
Increased, Motility
Increase, Inflammation
Alteration, lipid metabolism
AOP:505
Oxidative Stress
Decreased, PPAR-gamma activation
AOP:513
Activation of mitogen-activated protein kinase kinase, extracellular signal-regulated kinase 1/2
AOP:499
AOP:500
AOP:500
AOP:505
AOP:513
Disruption neuro-transmitter release
Activation, AhR
AOP:21
AOP:150
AOP:310
AOP:439
AOP:455
AOP:456
AOP:458
AOP:459
AOP:205
Direct mitochondrial inhibition
AOP:205
Mitochondrial impairment
AOP:450
Activation of Cyp2E1
Oxidative Stress in Brain
Induction CYP1A2/ CYP1A5
Increase, Oxidative Stress
Decrease sox9 expression
Induction Upregulation of glucuronyl-transferase activity
Increase slincR expression
dimerization, AHR/ARNT
reduced dimerization, ARNT/HIF1-alpha
Increase COX-2 expression
Antagonism, Estrogen receptor
AOP:30
Increase, DNA methyltrans-ferase expression
Increase, hypermethylation of the promotor region of gonadotropin releasing hormone receptor
Decrease, expression of gonadotropin releasing hormone receptor
Increase, DNA methyltrans-ferase inhibition
AOP:336
AOP:337
AOP:341
Decrease, Global DNA methylation
Decrease, Heritable DNA methylation (F3)
Increase, Caspase transcrip-tion (F3)
Increase, Caspase transcription
Inhibition, Aromatase
AOP:25
AOP:346
Increased, Reactive oxygen species
Increase, intra-cellular calcium
AOP:205
AOP:439
AOP:500
AOP:205
Decompart-mentalization
Acetylcho-linesterase (AchE) Inhibition
AOP:405
AOP:450
reduced production, VEGF
Reduction, Vitellogenin accumulation into oocytes and oocyte growth/ development
Increase, Oocyte apoptosis
Leukocyte recruitment/ activation
Increased, Invasion
General Apoptosis
Increased, Migration (Endothelial Cells)
Increase, Mitochondrial Dysfunction
Apoptosis
narcosis
Acetylcholine accumulation in synapses
Thyroid hormone synthesis, Decreased
AOP:459
Impairment, Endothelial network
Reduction, 17beta-estradiol synthesis by the undifferentiated gonad
Reduction, 17beta-estradiol synthesis by ovarian granulosa cells
Increase, Ovarian somatic cell apoptosis
Increase, Ovarian somatic cell apoptosis (F3)
AOP:505
AOP:513
AOP:205
Increased Pro-inflamm-atory mediators
Increased, Ductal Hyperplasia
AOP:293
Increase Cancer
Increase, angiogenesis
Necrosis
AOP:205
N/A, Neuro-degener-ation
AOP:450
AOP:500
Increased, Clearance of thyroxine from serum
Facial cartilage structures are reduced in size and morphologically distorted
AOP:455
Thyroxine (T4) in serum, Decreased
Hippocampal gene expression, Altered
Hippo-campal Physiology, Altered
Reduction, Vitellogenin synthesis in liver
Increased, Differen-tiation to Testis
Increased, tumor growth
Decreased, Neuronal network function in adult brain
Increased Cholinergic Signaling
Thyroxine (T4) in neuronal tissue, Decreased
Decrease of neuronal network function
Altered, Cardiovascular development/ function
Reduction, Plasma 17 beta-estradiol concentra-tions
Reduction, Plasma vitellogenin concentrations
Increase, Ovarian follicle breakdown
Increase, Ovarian follicle breakdown (F3)
Decrease, Oogenesis
Decrease, Oogenesis (F3)
AOP:336
AOP:337
AOP:341
AOP:25
AOP:30
AOP:310
AOP:21
AOP:150
AOP:455
AOP:456
AOP:293
AOP:439
AOP:499
AOP:500
AOP:450
N/A, Breast Cancer
Impairment, Learning and memory
Sensory axonal peripheral neuropathy
Cognitive Function, Decreased
Increase, Early Life Stage Mortality
Reduction, Cumulative fecundity and spawning
Decrease, Fecundity
Decrease, Fecundity (F3)
AOP:450
AOP:405
AOP:458
AOP:459
Increased Mortality
AOP:25
AOP:30
AOP:310
AOP:336
AOP:337
AOP:341
AOP:346
Increased, Male Biased Sex Ratio
Decrease, Population growth rate
Molecular Initiating Event (MIE)
Key Event (KE)
Adverse Outcome (AO)
Key events associated with Benzo[a]pyrene

Figure 6.6: Directed network corresponding to the LCC in the undirected B[a]P-AOP network, comprising 92 KEs and 125 KERs. Among the 92 KEs, 17 are categorized as MIEs (denoted as diamond), 17 are categorized as AOs (denoted as circle), and the remaining 58 are categorized as KEs (denoted as rounded square). The 51 KEs (including MIEs and AOs) associated with B[a]P are marked in ‘red’. In this figure, the 92 KEs are arranged vertically according to their level of biological organization.

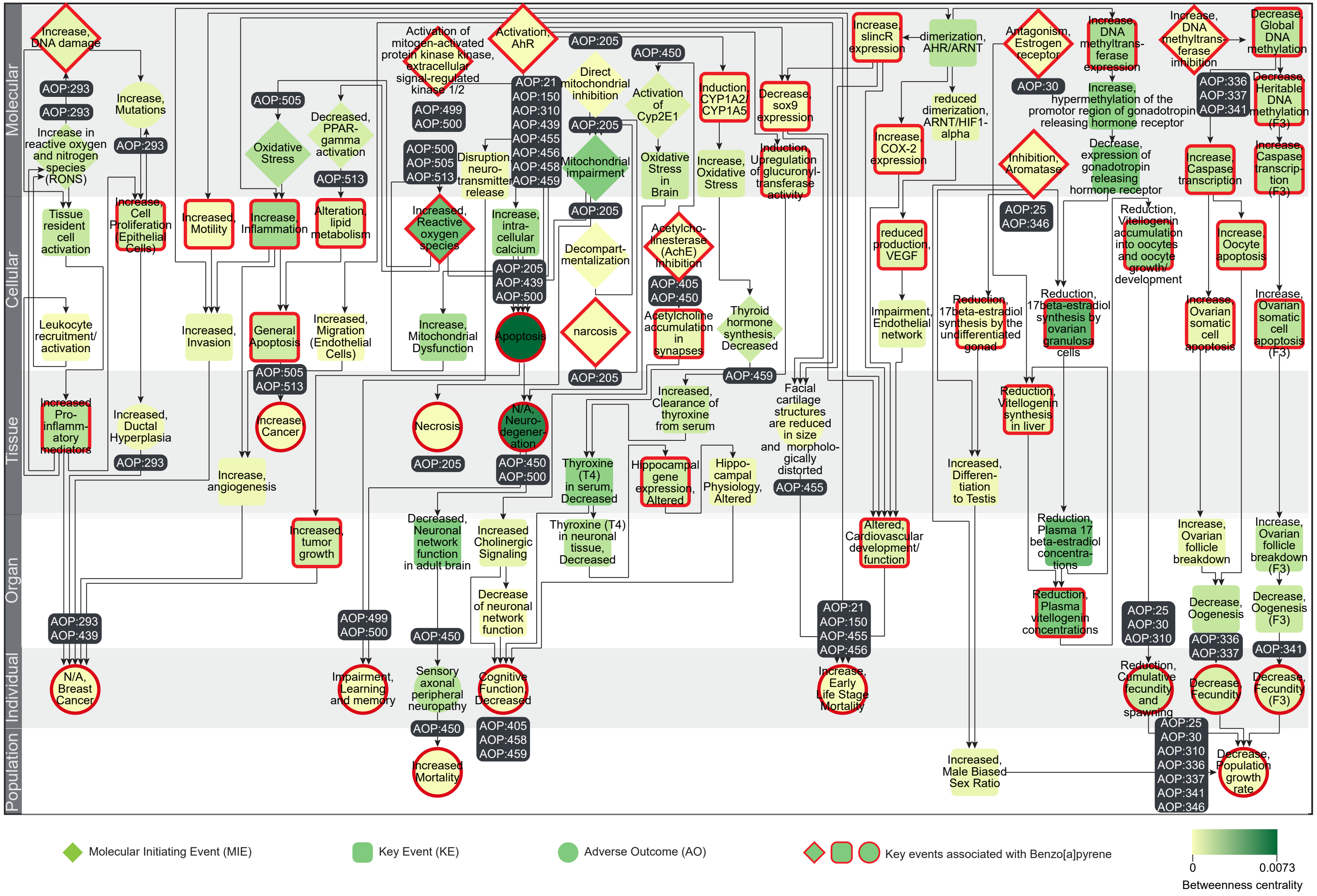

Molecular
Cellular
Tissue
Organ
Individual
Population
Increase, DNA damage
AOP:293
AOP:293
Increase, Mutations
Increase in reactive oxygen and nitrogen species (RONS)
AOP:293
Tissue resident cell activation
Increase, Cell Proliferation (Epithelial Cells)
Leukocyte recruitment/ activation
Increased Pro-inflammatory mediators
Increased, Ductal Hyperplasia
AOP:293
AOP:293
AOP:439
N/A, Breast Cancer
Increased, Motility
Increased, Invasion
AOP:505
Oxidative Stress
Increase, Inflammation
Decreased, PPAR-gamma activation
AOP:513
Alteration, lipid metabolism
General Apoptosis
Increased, Migration (Endothelial Cells)
AOP:505
AOP:513
Increase Cancer
Increase, angiogenesis
Increased, tumor growth
Activation of mitogen-activated protein kinase kinase, extracellular signal-regulated kinase 1/2
AOP:499
AOP:500
AOP:500
AOP:505
AOP:513
Disruption, neuro-transmitter release
Activation, AhR
AOP:21
AOP:150
AOP:310
AOP:439
AOP:455
AOP:456
AOP:458
AOP:459
Increased, Reactive oxygen species
Increase, intra-cellular calcium
AOP:205
AOP:439
AOP:500
Increase, Mitochondrial Dysfunction
Apoptosis
Necrosis
AOP:205
N/A, Neuro-degeneration
AOP:450
AOP:500
Decreased, Neuronal network function in adult brain
Increased Cholinergic Signaling
Decrease of neuronal network function
AOP:499
AOP:500
Impairment, Learning and memory
AOP:450
Sensory axonal peripheral neuropathy
AOP:450
Increased Mortality
Cognitive Function Decreased
AOP:405
AOP:458
AOP:459
AOP:205
Direct mitochondrial inhibition
AOP:205
Mitochondrial impairment
AOP:205
Decompart-mentalization
narcosis
AOP:205
AOP:450
Activation of Cyp2E1
Oxidative Stress in Brain
Acetylcholinesterase (AchE) Inhibition
AOP:405
AOP:450
Acetylcholine accumulation in synapses
Increased, Clearance of thyroxine from serum
Thyroxine (T4) in serum, Decreased
Thyroxine (T4) in neuronal tissue, Decreased
Hippocampal gene expression, Altered
Hippocampal Physiology, Altered
Induction, CYP1A2/ CYP1A5
Increase, Oxidative Stress
Thyroid hormone synthesis, Decreased
AOP:459
Decrease sox9 expression
Induction, Upregulation of glucuronyltransferase activity
Facial cartilage structures are reduced in size and morphologically distorted
AOP:455
Increase slincR expression
dimerization, AHR/ARNT
reduced dimerization, ARNT/HIF1-alpha
Increase COX-2 expression
reduced production, VEGF
Impairment, Endothelial network
Altered, Cardiovascular development/ function
AOP:21
AOP:150
AOP:455
AOP:456
Increase, Early Life Stage Mortality
Antagonism, Estrogen receptor
AOP:30
Inhibition, Aromatase
AOP:25
AOP:346
Reduction, 17beta-estradiol synthesis by the undifferentiated gonad
Increased, Differentiation to Testis
Increased, Male Biased Sex Ratio
Increase, DNA methyltransferase expression
Increase, hypermethylation of the promotor region of gonadotropin releasing hormone receptor
Decrease, expression of gonadotropin releasing hormone receptor
Reduction, Vitellogenin accumulation into oocytes and oocyte growth/ development
Reduction, 17beta-estradiol synthesis by ovarian granulosa cells
Reduction, Vitellogenin synthesis in liver
Reduction, Plasma 17 beta-estradiol concentrations
Reduction, Plasma vitellogenin concentrations
AOP:25
AOP:30
AOP:310
Reduction, Cumulative fecundity and spawning
AOP:25
AOP:30
AOP:310
AOP:336
AOP:337
AOP:341
AOP:346
Increase, DNA methyltransferase inhibition
AOP:336
AOP:337
AOP:341
Increase, Caspase transcription
Increase, Oocyte apoptosis
Increase, Ovarian somatic cell apoptosis
Increase, Ovarian follicle breakdown
Decrease, Oogenesis
AOP:336
AOP:337
Decrease, Fecundity
Decrease, Population growth rate
Decrease, Global DNA methylation
Decrease, Heritable DNA methylation (F3)
Increase, Caspase transcription (F3)
Increase, Ovarian somatic cell apoptosis (F3)
Increase, Ovarian follicle breakdown (F3)
Decrease, Oogenesis (F3)
AOP:341
Decrease, Fecundity (F3)
Molecular Initiating Event (MIE)
Key Event (KE)
Adverse Outcome (AO)
Key events associated with Benzo[a]pyrene
0
0.0073
Betweenness centrality

Figure 6.7: Directed network corresponding to the LCC (C1) in the B[a]P-AOP network, where the KEs (including MIEs and AOs) are colored based on their betweenness centrality values. The 51 KEs (including MIEs and AOs) associated with B[a]P are marked in 'red'. In this figure, the 92 KEs are arranged vertically according to their level of biological organization.

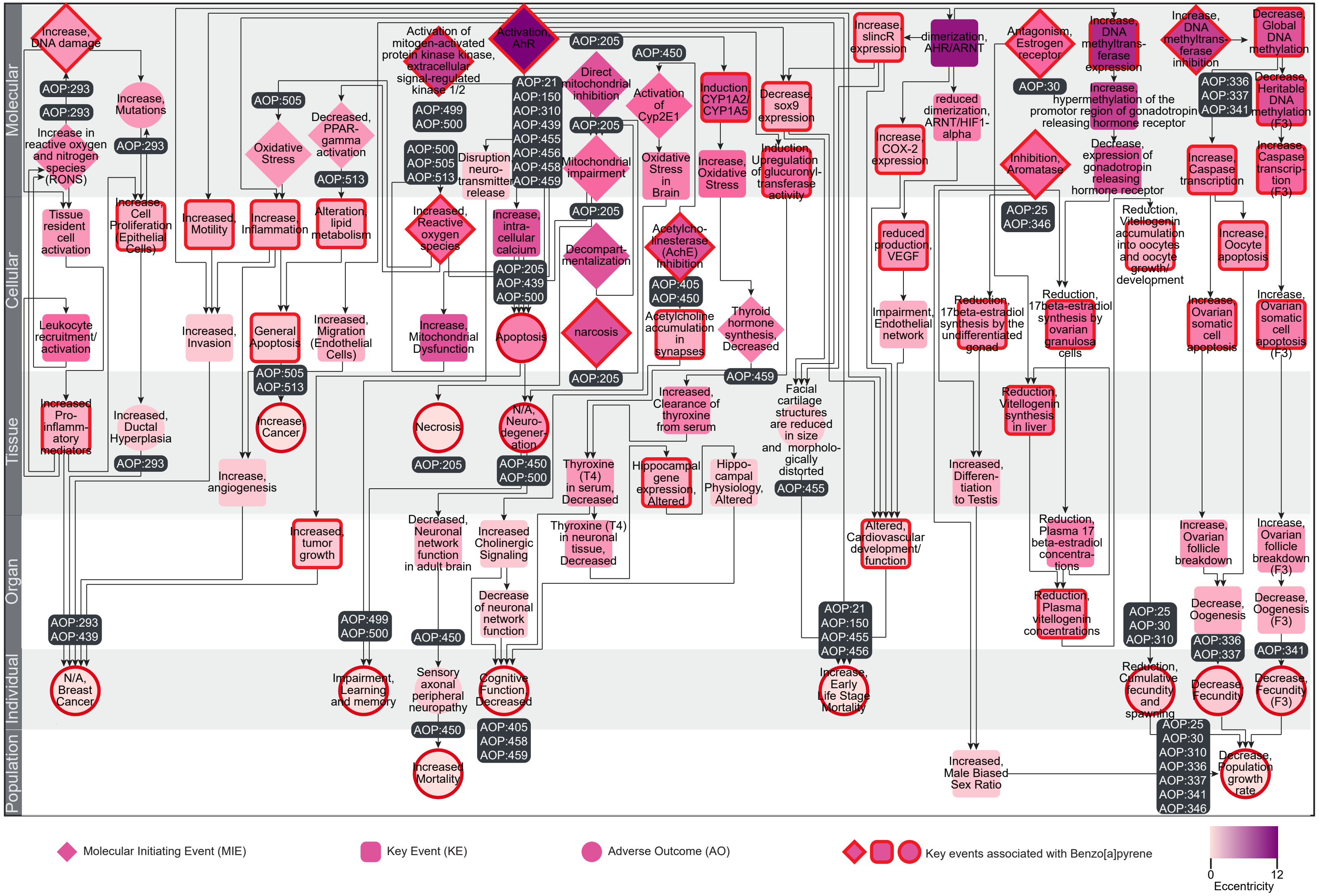

Molecular
Cellular
Tissue
Organ
Individual
Population
Increase, DNA damage
AOP:293
AOP:293
Increase in reactive oxygen and nitrogen species (RONS)
Increase, Mutations
AOP:293
Tissue resident cell activation
Increase, Cell Proliferation (Epithelial Cells)
Leukocyte recruitment/ activation
Increased Pro-inflammatory mediators
Increased, Ductal Hyperplasia
AOP:293
Increased, Motility
Increase, Inflammation
Increased, Invasion
General Apoptosis
AOP:505
Oxidative Stress
Decreased, PPAR-gamma activation
AOP:513
Alteration, lipid metabolism
Increased, Migration (Endothelial Cells)
AOP:505
AOP:513
Increase Cancer
Increase, angiogenesis
Increased, tumor growth
Activation of mitogen-activated protein kinase kinase, extracellular signal-regulated kinase 1/2
AOP:499
AOP:500
AOP:500
AOP:505
AOP:513
Disruption neuro-transmitter release
Increased, Reactive oxygen species
Increase, Mitochondrial Dysfunction
Necrosis
AOP:205
Activation, AhR
AOP:21
AOP:150
AOP:310
AOP:439
AOP:455
AOP:456
AOP:458
AOP:459
Increase, intra-cellular calcium
AOP:205
AOP:439
AOP:500
Apoptosis
N/A, Neuro-degeneration
AOP:450
AOP:500
AOP:205
Direct mitochondrial inhibition
AOP:205
Mitochondrial impairment
AOP:205
Decompart-mentalization
narcosis
AOP:205
AOP:450
Activation of Cyp2E1
Oxidative Stress in Brain
Acetylcholinesterase (AchE) Inhibition
AOP:405
AOP:450
Acetylcholine accumulation in synapses
Increased, Clearance of thyroxine from serum
Thyroxine (T4) in serum, Decreased
Hippocampal gene expression, Altered
Hippocampal Physiology, Altered
Induction, CYP1A2/ CYP1A5
Increase, Oxidative Stress
Thyroid hormone synthesis, Decreased
AOP:459
Decrease sox9 expression
Induction Upregulation of glucuronyltransferase activity
Facial cartilage structures are reduced in size and morphologically distorted
AOP:455
Decreased, Neuronal network function in adult brain
Increased Cholinergic Signaling
Thyroxine (T4) in neuronal tissue, Decreased
Decrease of neuronal network function
AOP:499
AOP:500
AOP:450
AOP:293
AOP:439
N/A, Breast Cancer
Impairment, Learning and memory
Sensory axonal peripheral neuropathy
AOP:450
Increased Mortality
Cognitive Function Decreased
AOP:405
AOP:458
AOP:459
Increase slincR expression
dimerization, AHR/ARNT
reduced dimerization, ARNT/HIF1-alpha
Increase COX-2 expression
reduced production, VEGF
Impairment, Endothelial network
Altered, Cardiovascular development/ function
AOP:21
AOP:150
AOP:455
AOP:456
Increase, Early Life Stage Mortality
Antagonism, Estrogen receptor
AOP:30
Inhibition, Aromatase
AOP:25
AOP:346
Reduction, 17beta-estradiol synthesis by the undifferentiated gonad
Increased, Differentiation to Testis
Increased, Male Biased Sex Ratio
Increase, DNA methyltransferase expression
Increase, hypermethylation of the promotor region of gonadotropin releasing hormone receptor
Decrease, expression of gonadotropin releasing hormone receptor
Reduction, 17beta-estradiol synthesis by ovarian granulosa cells
Reduction, Vitellogenin accumulation into oocytes and oocyte growth/ development
Reduction, Vitellogenin synthesis in liver
Reduction, Plasma 17 beta-estradiol concentrations
Reduction, Plasma vitellogenin concentrations
AOP:25
AOP:30
AOP:310
Reduction, Cumulative fecundity and spawning
Increase, DNA methyltransferase inhibition
AOP:336
AOP:337
AOP:341
Increase, Caspase transcription
Increase, Oocyte apoptosis
Increase, Ovarian somatic cell apoptosis
Increase, Ovarian follicle breakdown
Decrease, Oogenesis
AOP:336
AOP:337
Decrease Fecundity
Decrease, Global DNA methylation
Decrease, Heritable DNA methylation (F3)
Increase, Caspase transcription (F3)
Increase, Ovarian somatic cell apoptosis (F3)
Increase, Ovarian follicle breakdown (F3)
Decrease, Oogenesis (F3)
AOP:341
Decrease, Fecundity (F3)
AOP:25
AOP:30
AOP:310
AOP:336
AOP:337
AOP:341
AOP:346
Decrease, Population growth rate
Molecular Initiating Event (MIE)
Key Event (KE)
Adverse Outcome (AO)
Key events associated with Benzo[a]pyrene
0
12
Eccentricity

Figure 6.8: Directed network corresponding to the LCC (C1) in the B[a]P-AOP network, where the KEs (including MIEs and AOs) are colored based on their eccentricity values. The 51 KEs (including MIEs and AOs) associated with B[a]P are marked in 'red'. In this figure, the 92 KEs are arranged vertically according to their level of biological organization.

Toxicity pathway linking B[a]P exposure to transgenerational effects

B[a]P is a ubiquitous environmental pollutant that has been associated with transgenerational health consequences in humans and animals [331–333]. Here, we observed a B[a]P-AOP titled 'DNA methyltransferase inhibition leading to transgenerational effects (2)' (AOP:341) having a cumulative WoE of 'Moderate' (Supplementary Table S6.10) and taxonomical relevance to *Daphnia magna* (Supplementary Table S6.2). Therefore, we relied on this ecotoxicologically-relevant AOP to understand the rationale behind B[a]P-induced transgenerational effects.

Different *in vitro* experiments showed a reduction in methyltransferase reactions in embryonic fibroblasts upon exposure to B[a]P [334, 335]. Subsequently, Corrales *et al.* [336] observed a decrease in global DNA methylation following a parental and continued embryonic waterborne B[a]P exposure in zebrafish embryo and larvae. Wan *et al.* [333] showed that ancestral B[a]P exposure in medaka fish led to transgenerational skeletal deformities and changes in gene expression, primarily mediated by histone modifications and miRNAs rather than DNA methylation. Furthermore, Lin *et al.* showed that maternal exposure to B[a]P increased oxidative stress, leading to higher expression of cleaved caspase-3 in the neuroepithelium of mice embryos [337]. Malott *et al.* observed that the gestational exposure to B[a]P led to ovarian follicle depletion in the mice offspring ovaries and oocytes, with increased mitochondrial superoxide levels and induced apoptosis via the mitochondrial pathway [338]. Finally, Sui *et al.* observed that B[a]P exposure compromised oogenesis in mice offspring, leading to reduced oocyte maturation, increased meiotic abnormalities, and decreased embryo developmental competence due to mitochondrial dysfunction, oxidative stress and early apoptosis, all of which led to reduced population sizes in later generations [339]. Thus, we were able to explore a potential toxicity pathway underlying B[a]P-induced transgenerational effects by leveraging various published evidence.

6.2.4 Stressor-species networks for PHs

Using toxicity concentration as edge weight

A stressor-species network constructed using toxicity concentrations as edge weights can provide information on the variability of chemical toxicity across different species [94]. In this study, we leveraged acute toxicity concentration data (LC_{50} and EC_{50}) for PHs from the ECOTOX database and constructed a bipartite stressor-species network (Supplementary Table S6.8). The resulting network comprises 80 PHs and 221 species (spanning 12 ECOTOX species groups) with 815 stressor-species links (Supplementary Table S6.8). We observed that the 32 PAHs in the stressor-species network are documented to be toxic to the highest number of species (163), spanning 11 ECOTOX species groups (Figure 6.9). The species groups most tested by the PAHs are 'Crustaceans', 'Fish' and 'Algae' (Figure 6.9). Figure 6.10 shows the stressor-species network for 14 priority PAHs, which are linked to 160 species through 350 stressor-species connections. Among the 14 PAHs, Fluoranthene (CAS:206-44-0) has been documented to be toxic to the highest number of species (75), followed by Naphthalene (CAS:91-20-3) to 55 species, and Phenanthrene (CAS:85-01-8) to 50 species (Figure 6.10). The species *Daphnia magna* and *Oncorhynchus mykiss* are linked to 12 and 11 PAHs, respectively (Figure 6.10), making them the most tested species by the priority PAHs. Overall, we observed that 'Crustaceans' are the most tested ECOTOX species group by the priority PAHs.

Using bioconcentration factor as edge weight

A stressor-species network constructed using the bioconcentration factor as edge weight can elucidate the extent of chemical absorption by various species from their environment through respiration and dermal surfaces, excluding absorption through diet [340]. In this study, we leveraged the bioconcentration factors for PHs from the ECOTOX database and constructed a bipartite stressor-species network (Supplementary Table S6.9). The resulting network comprises 28 PHs and 59 species (spanning 9 ECOTOX species groups)

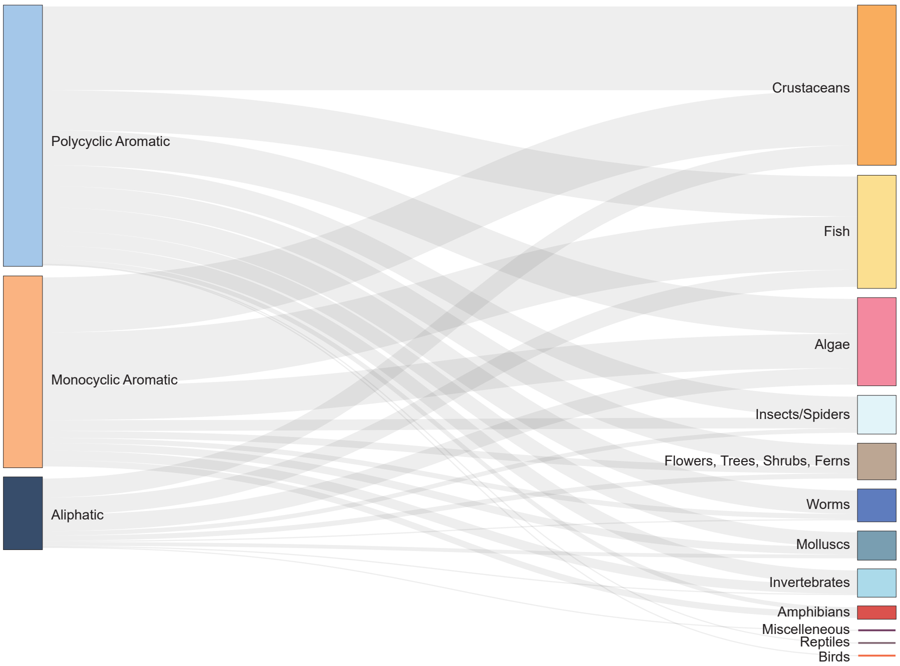
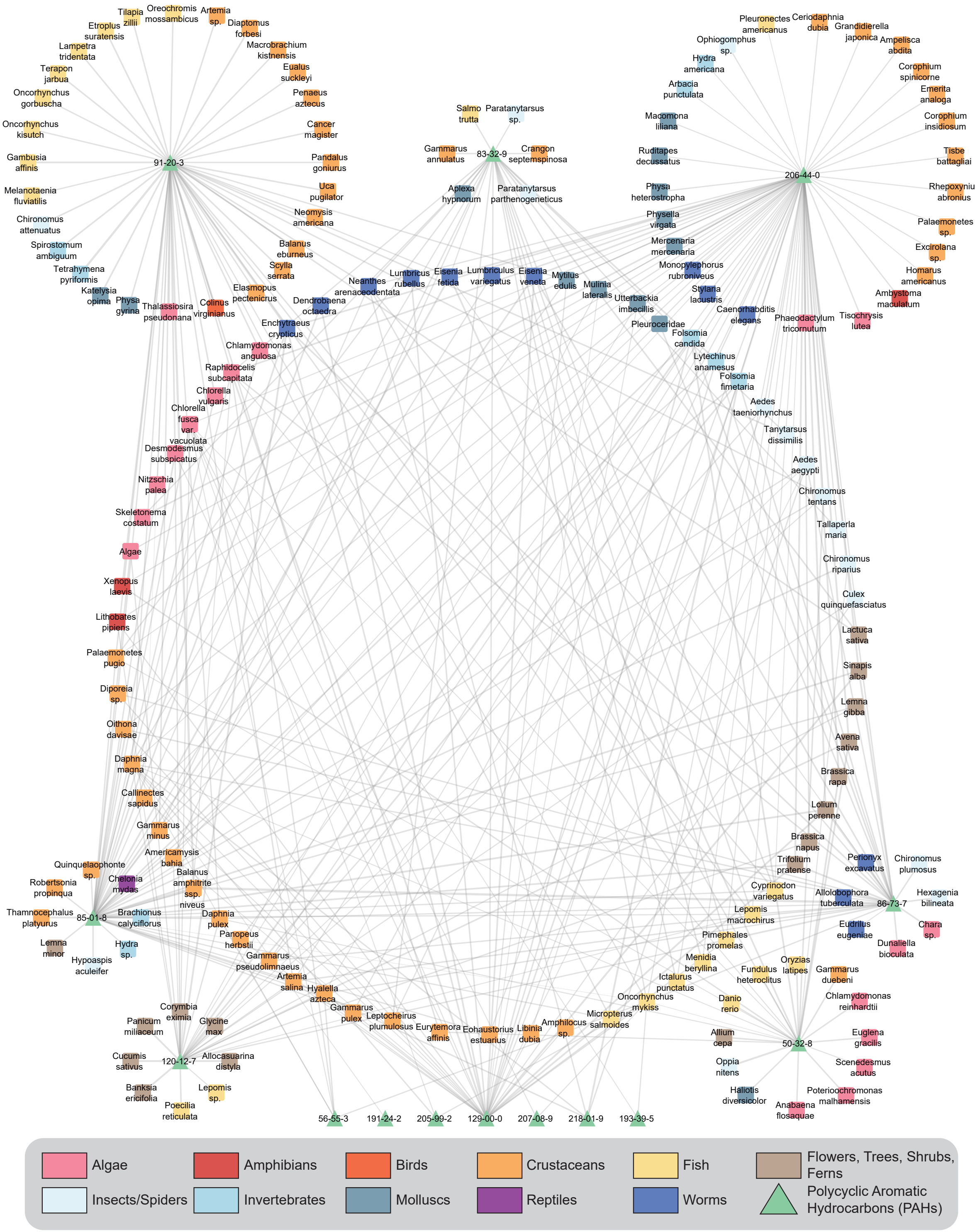


Figure 6.9: Sankey plot depicting associations between different types of PHs and ECOTOX species groups through their acute toxicity concentration data (LC_{50} and EC_{50}). The plot provides associations between 3 types of PHs (aliphatic, monocyclic aromatic and polycyclic aromatic) with 12 ECOTOX species groups.

with 159 stressor-species links (Supplementary Table S6.9). We observed that 22 PAHs in the stressor-species network are documented to be absorbed by majority of the ECOTOX species groups (Figure 6.11). Species groups namely, ‘Crustaceans’, ‘Fish’ and ‘Molluscs’ are reported to absorb 21, 14 and 12 PHs, respectively (Supplementary Table S6.9). Figure 6.12 shows the stressor-species network of 13 priority PAHs with 54 species comprising 117 stressor-species links. Among the 13 PAHs, B[a]P (CAS:50-32-8) is documented to be absorbed in highest number of species (20), followed by Phenanthrene (CAS:85-01-8) and Fluoranthene (CAS:206-44-0) in 17 species each. We observed that many of the PAHs are documented to be absorbed in species namely, *Daphnia magna*, *Daphnia pulex* and *Hyalella azteca* from ‘Crustaceans’ ECOTOX species group (Figure 6.12).



91-20-3
Tilapia zillii
Oreochromis mossambicus
Artemia sp.
Etroplus suratensis
Diaptomus forbesi
Lampetra tridentata
Macrobrachium kistnensis
Terapon jarbua
Eualus suckleyi
Oncorhynchus gorbuscha
Penaeus aztecus
Oncorhynchus kisutch
Cancer magister
Gambusia affinis
Pandalus goniurus
Melanotaenia fluviatilis
Uca pugilator
Chironomus attenuatus
Neomysis americana
Spirostomum ambiguum
Balanus eburneus
Tetrahymena pyriformis
Scylla serrata
Katelysia opima
Physa gyrina
Thalassiosira pseudonana
Colinus virginianus
Elasmopus pectenicrus
83-32-9
Salmo trutta
Paratanytarsus sp.
Gammarus annulatus
Crangon septemspinosa
Aplexa hypnorum
Paratanytarsus parthenogeneticus
Neanthes arenaceodentata
Lumbricus rubellus
Eisenia fetida
Lumbriculus variegatus
Eisenia veneta
Mytilus edulis
Mulinia lateralis
Dendrobaena octaedra
Enchytraeus crypticus
206-44-0
Pleuronectes americanus
Ceriodaphnia dubia
Grandidierella japonica
Ampelisca abdita
Ophiogomphus sp.
Hydra americana
Corophium spinicorne
Arbacia punctulata
Emerita analoga
Macomona liliana
Corophium insidiosum
Ruditapes decussatus
Tisbe battagliai
Physa heterostropha
Rhepoxynius abronius
Physella virgata
Palaemonetes sp.
Mercenaria mercenaria
Excirolana sp.
Monopylephorus rubroniveus
Homarus americanus
Stylaria lacustris
Ambystoma maculatum
Utterbackia imbecillis
Caenorhabditis elegans
Phaeodactylum tricornutum
Tisochrysis lutea
Pleuroceridae
Folsomia candida
Lytechinus anamesus
Folsomia fimetaria
Aedes taeniorhynchus
Tanytarsus dissimilis
Aedes aegypti
Chironomus tentans
Tallaperla maria
Chironomus riparius
Culex quinquefasciatus
Lactuca sativa
Sinapis alba
Lemna gibba
Avena sativa
Brassica rapa
Lolium perenne
Brassica napus
Trifolium pratense
Perionyx excavatus
Chironomus plumosus
Cyprinodon variegatus
Allolobophora tuberculata
86-73-7
Hexagenia bilineata
Lepomis macrochirus
Eudrilus eugeniae
Chara sp.
Pimephales promelas
Dunaliella bioculata
Menidia beryllina
Oryzias latipes
Gammarus duebeni
Ictalurus punctatus
Fundulus heteroclitus
Chlamydomonas reinhardtii
Oncorhynchus mykiss
Danio rerio
Micropterus salmoides
Euglena gracilis
Allium cepa
50-32-8
Oppia nitens
Scenedesmus acutus
Haliotis diversicolor
Poterioochromonas malhamensis
Anabaena flosaquae
Chlamydomonas angulosa
Raphidocelis subcapitata
Chlorella vulgaris
Chlorella fusca var. vacuolata
Desmodesmus subspicatus
Nitzschia palea
Skeletonema costatum
Algae
Xenopus laevis
Lithobates pipiens
Palaemonetes pugio
Diporeia sp.
Oithona davisae
Daphnia magna
Callinectes sapidus
Gammarus minus
Americamysis bahia
Quinquelaophonte sp.
Balanus amphitrite ssp. niveus
Robertsonia propinqua
Chelonia mydas
Thamnocephalus platyurus
85-01-8
Brachionus calyciflorus
Daphnia pulex
Panopeus herbstii
Lemna minor
Hydra sp.
Hypoaspis aculeifer
Gammarus pseudolimnaeus
Artemia salina
Hyalella azteca
Corymbia eximia
Gammarus pulex
Leptocheirus plumulosus
Eurytemora affinis
Eohaustorius estuarius
Libinia dubia
Amphilocus sp.
Panicum miliaceum
Glycine max
Cucumis sativus
120-12-7
Allocasuarina distyla
Banksia ericifolia
Lepomis sp.
Poecilia reticulata
56-55-3
191-24-2
205-99-2
129-00-0
207-08-9
218-01-9
193-39-5
Algae
Amphibians
Birds
Crustaceans
Fish
Flowers, Trees, Shrubs, Ferns
Insects/Spiders
Invertebrates
Molluscs
Reptiles
Worms
Polycyclic Aromatic Hydrocarbons (PAHs)

Figure 6.10 ***(previous page)*****:** Stressor-species network constructed for EPA priority PAHs using the acute toxicity concentration as edge weights. The network comprises 14 priority PAHs and 160 species with 350 stressor-species links. The edges in the stressor-AOP network are represented by the logarithm of standardized acute toxicity concentration data (LC_{50} and EC_{50}). The species in the stressor-species network are classified according to the ECOTOX species groups.

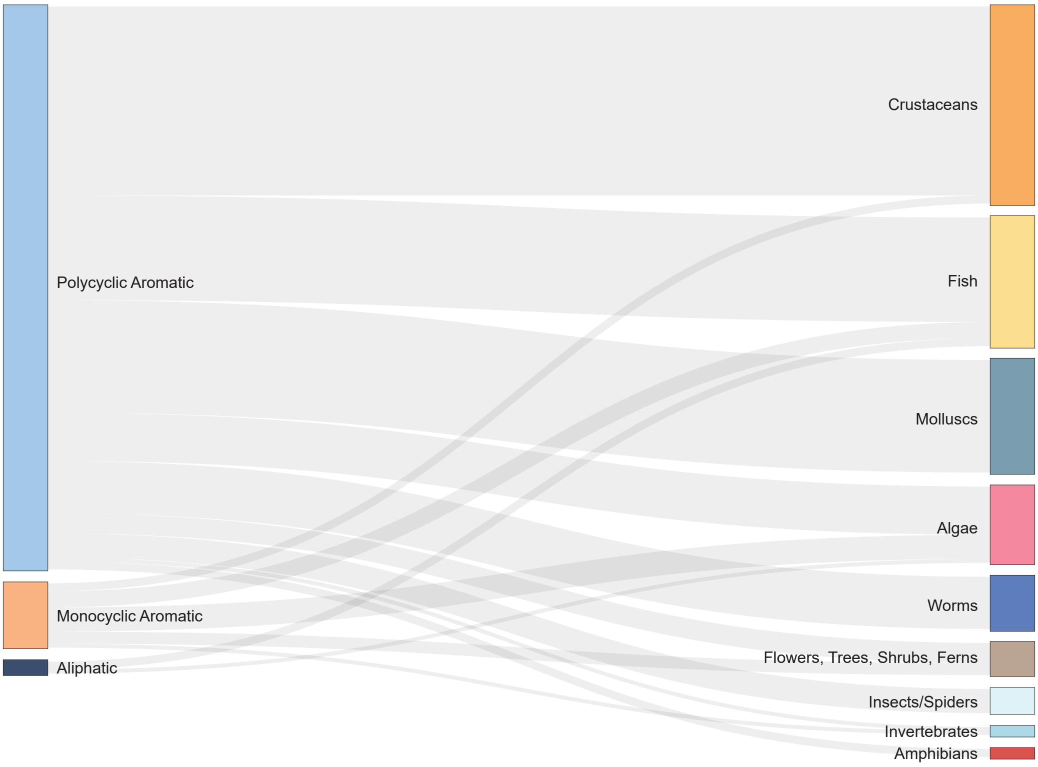


Figure 6.11: Sankey plot depicting associations between different types of PHs and ECOTOX species groups through their bioconcentration factors. The plot provides associations between 3 types of PHs (aliphatic, monocyclic aromatic and polycyclic aromatic) with 9 ECOTOX species groups.

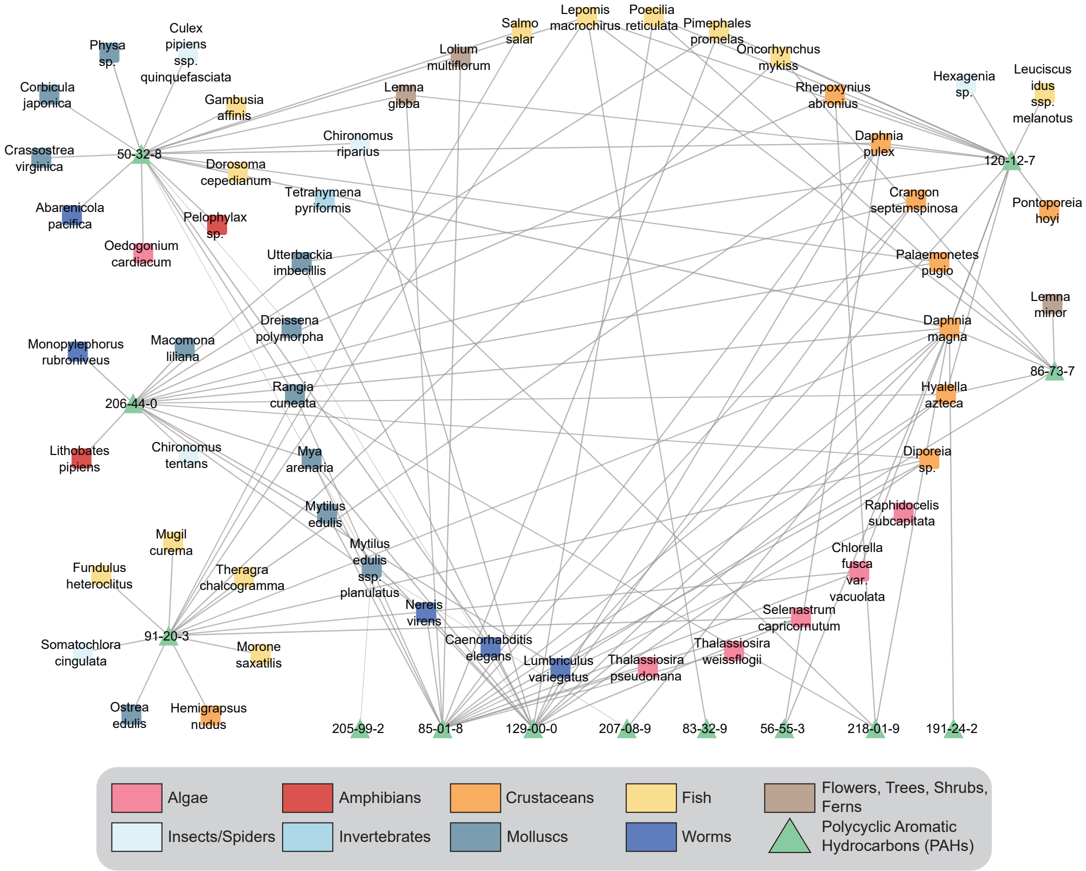


Figure 6.12: Stressor-species network constructed for EPA priority PAHs using the bioconcentration factors as edge weights. The network comprises 13 priority PAHs, 54 species and 117 stressor-species links. The edges in the stressor-AOP network are represented by the logarithm of bioconcentration factors. The species in the stressor-species network are classified according to the ECOTOX species groups.

6.2.5 SSD for EPA priority PAHs

SSD has been extensively used to derive environmental quality criterion or hazard concentration of chemicals in different environments [99,341,342]. In this study, we relied on the acute toxicity endpoints provided by ECOTOX to construct the SSDs for the priority PAHs [301, 311], and leveraged them to derive the corresponding hazard concentrations in aquatic environments (Figure 6.3; Supplementary Table S6.12). We observed that the acute toxicity endpoints associated with 8 of the 16 priority PAHs namely, Fluoranthene, Naphthalene, Phenanthrene, B[a]P, Acenaphthene, Pyrene, Anthracene, and 9H-Fluorene are reported in terms of LC_{50} or EC_{50} values, and observed within the duration of 24 to 96 hours in at least 5 ECOTOX species groups (Figure 6.3; Supplementary Table S6.12). Therefore, we accessed the acute toxicity endpoints from ECOTOX for each of these eight priority PAHs and employed both US EPA SSD Toolbox [317] and the R-based ssdtools [318] to construct SSDs and derive the corresponding hazard concentration values.

The SSD Toolbox and ssdtools fit the data using five distributions, namely Log-Normal, Log-Logistic, Log-Gumbel, Weibull and Burr Type III, and provide the HC05 value based on the best-fit model determined by the minimum corrected Akaike Information Criterion (AIC) value [317, 318, 343]. We observed that both tools identified the same best-fit model for each of the eight priority PAHs, and the derived HC05 values were also similar (Figure 6.13; Supplementary Table S6.13). For example, Figure 6.14 shows the plots of SSD for B[a]P computed using the best-fit Weibull model, as determined by both the SSD Toolbox and ssdtools. Figures 6.15-6.21 present the plots of SSD computed using the best-fit models for each of the other seven priority PAHs, as determined by both the SSD Toolbox and ssdtools.

Recently, the method of model averaging has been proposed to compute SSDs, wherein the averaged model is obtained by assigning model weight to each of the individual distributions [318, 344]. The HC05 values computed based on model averaging have been observed to be more reliable and stable compared to the values obtained from individ-

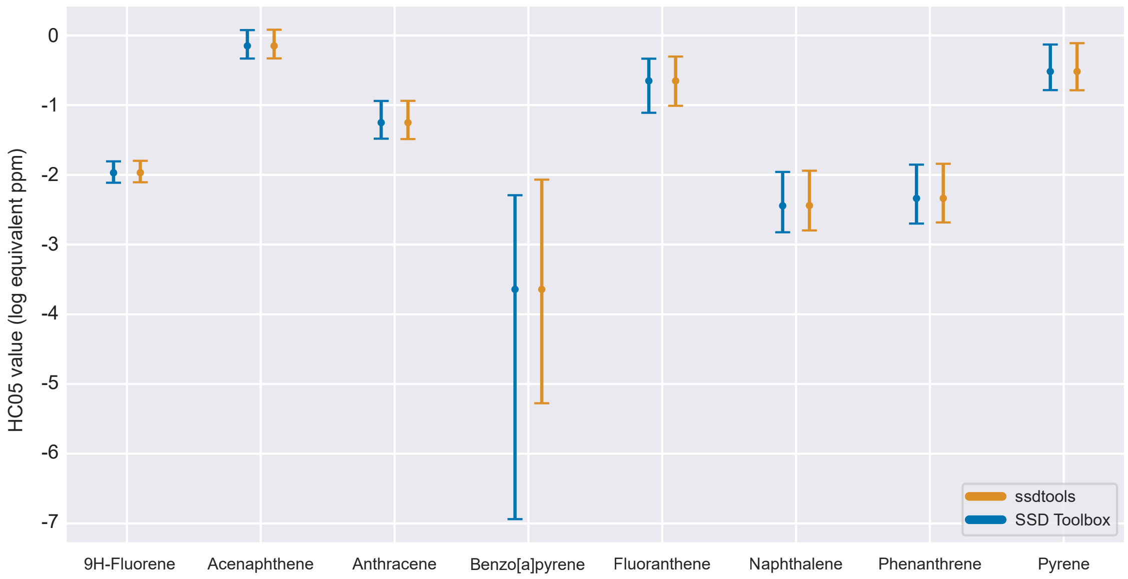


Figure 6.13: Comparison of the derived HC05 values for each of the eight priority PAHs using the corresponding best-fit model in US EPA SSD Toolbox and ssdtools.

ual models [318, 344]. Therefore, we computed model-averaged HC05 values for each of the eight priority PAHs and observed that both tools provided similar results (Table 6.1). We observed that B[a]P has the lowest model-averaged HC05 value, and naphthalene has the highest model-averaged HC05 value (Table 6.1). The HC05 values indicate the toxic effects of chemicals across species, with smaller HC05 values implying greater toxicity [345]. Based on the model-averaged HC05 values, we ordered the eight priority PAHs from most toxic to least toxic for aquatic organisms as follows: B[a]P > Pyrene > Anthracene > Fluoranthene > Phenanthrene > Acenaphthene > 9H-Fluorene > Naphthalene. Previously, Chen *et al.* leveraged chronic aquatic toxicity data and observed a similar order of the eight PAHs based on their computed HC05 values [346]. Moreover, we observed that this decreasing order coincides with the number of benzene rings in these compounds, potentially indicating a correlation between the number of rings in PAHs and their level of toxicity [94].

The species located in the region of lower toxicity in the SSD are identified as sensitive to that particular chemical [341]. We observed that the species belonging to the ECOTOX species groups ‘Crustaceans’, ‘Fish’ and ‘Molluscs’, are commonly found in

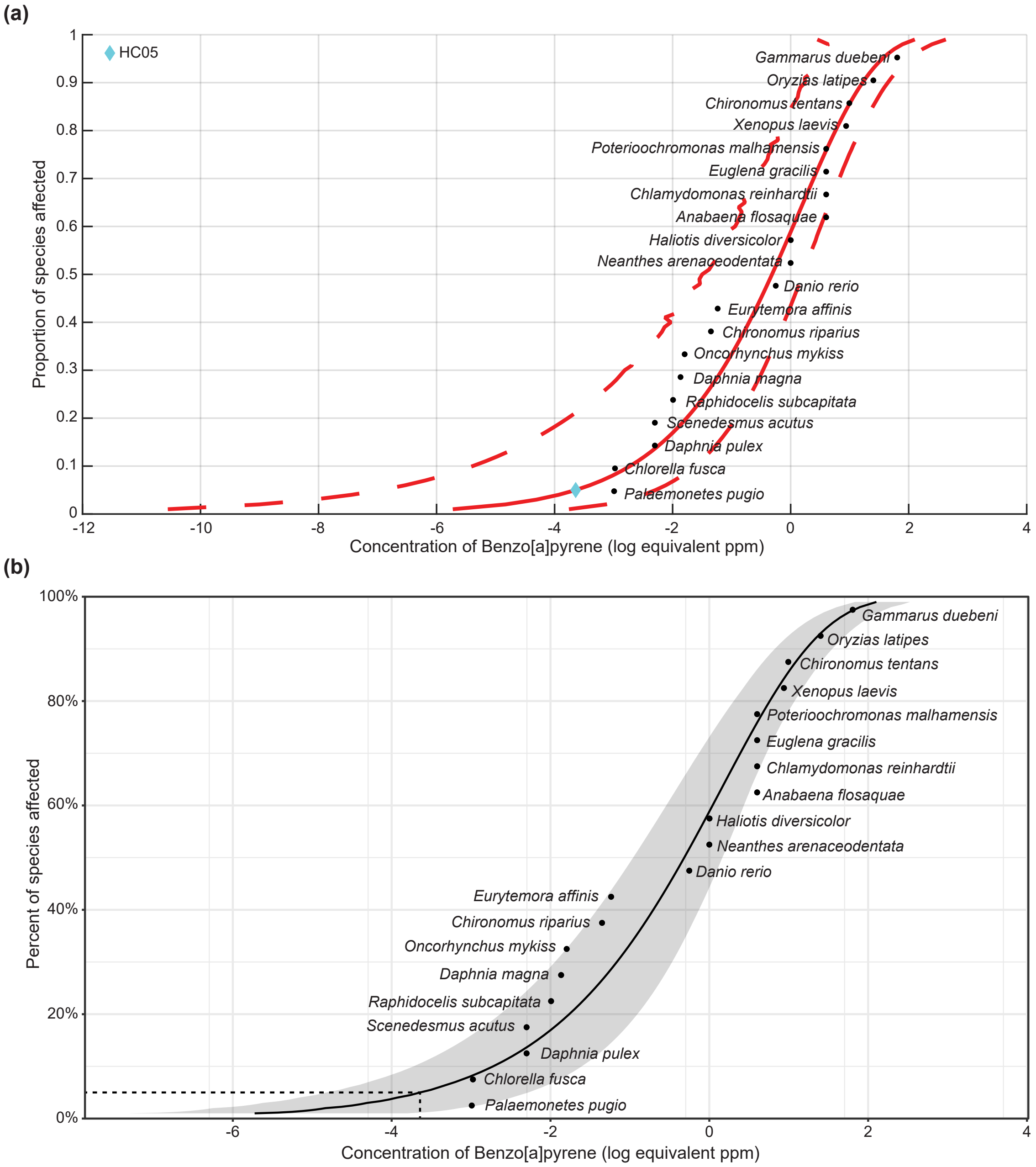


Figure 6.14: The plots of SSD for B[a]P computed using the best-fit Weibull model. (a) As determined by US EPA SSD Toolbox where the HC05 value is denoted by cyan colored diamond. (b) As determined by ssdtools where the HC05 value is denoted by a dotted line.

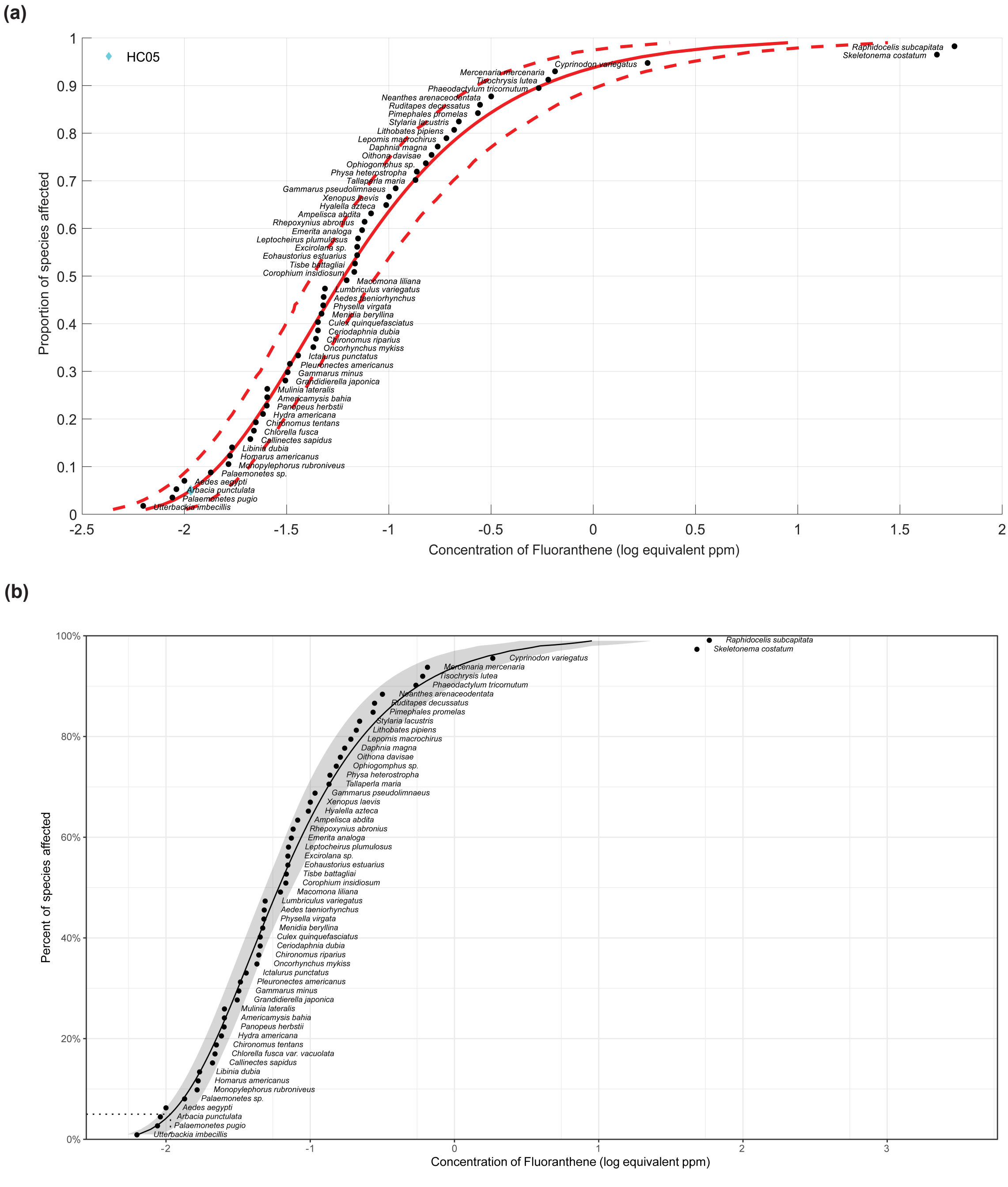


Figure 6.15: The plots of SSD for Fluoranthene computed using the best-fit Log-Gumbel model. (a) As determined by US EPA SSD Toolbox where the HC05 value is denoted by cyan colored diamond. (b) As determined by ssdtools where the HC05 value is denoted by a dotted line.

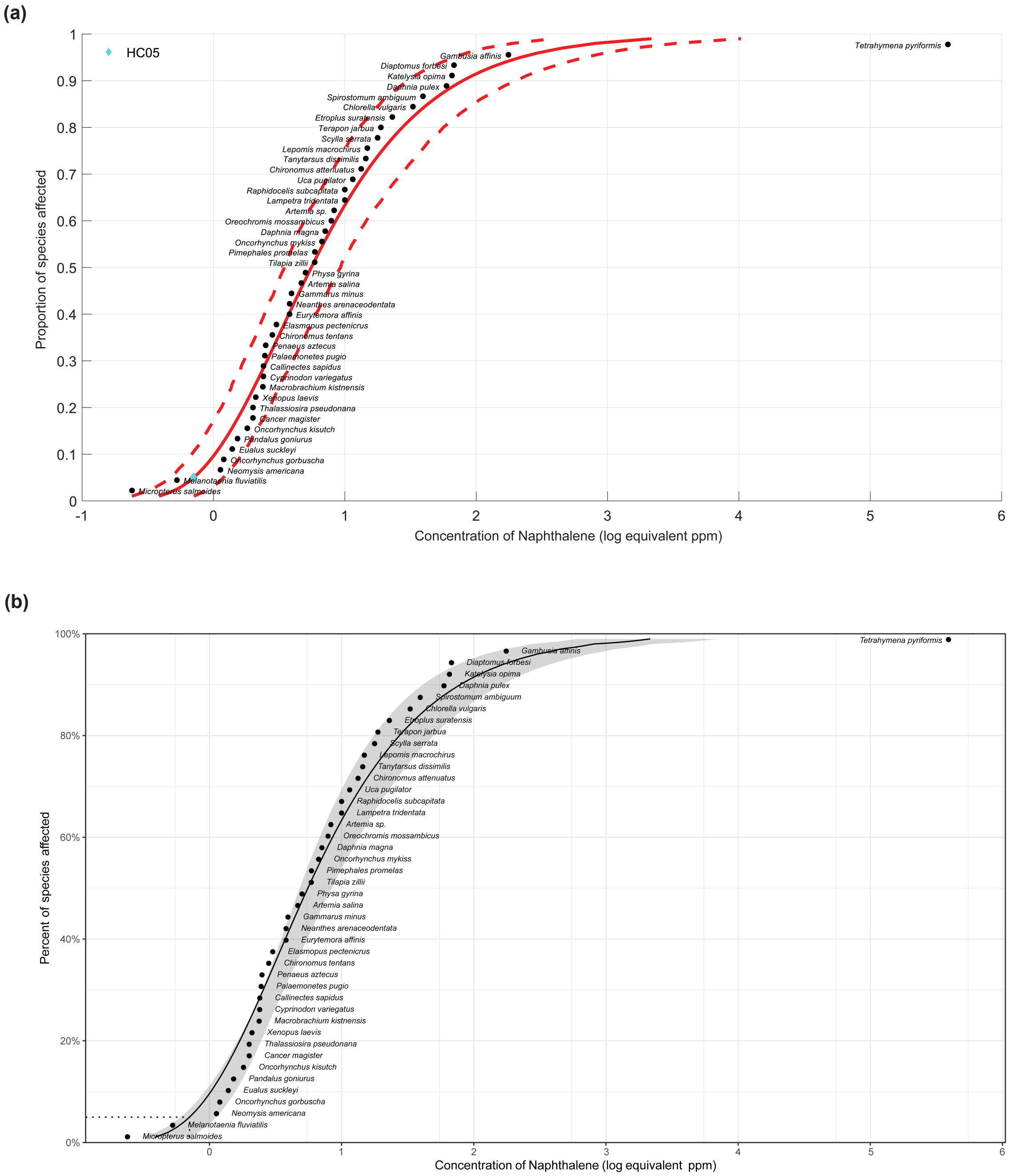


Figure 6.16: The plots of SSD for Naphthalene computed using the best-fit Log-Gumbel model. (a) As determined by US EPA SSD Toolbox where the HC05 value is denoted by cyan colored diamond. (b) As determined by ssdtools where the HC05 value is denoted by a dotted line.

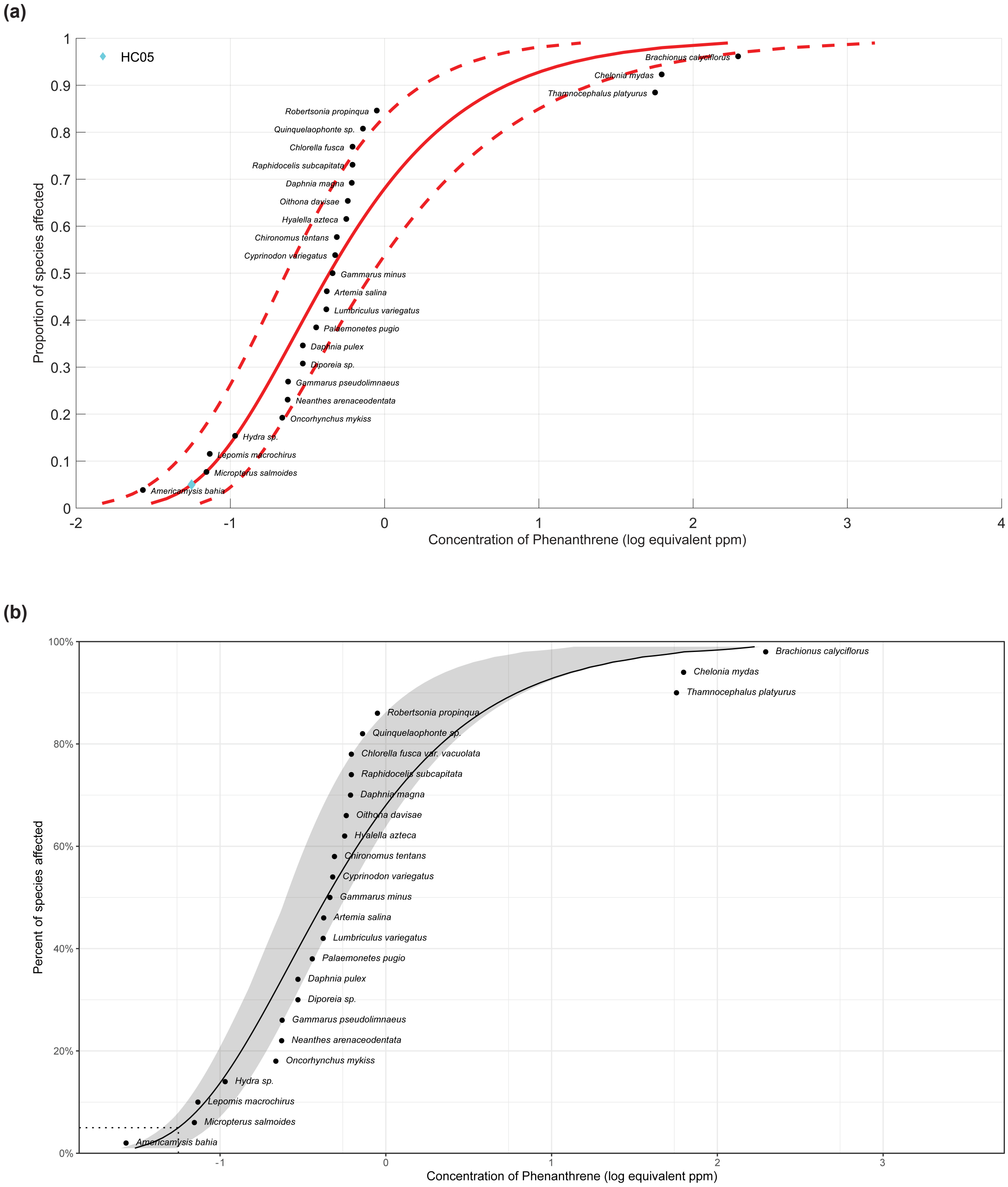


Figure 6.17: The plots of SSD for Phenanthrene computed using the best-fit Log-Gumbel model. (a) As determined by US EPA SSD Toolbox where the HC05 value is denoted by cyan colored diamond. (b) As determined by ssdtools where the HC05 value is denoted by a dotted line.

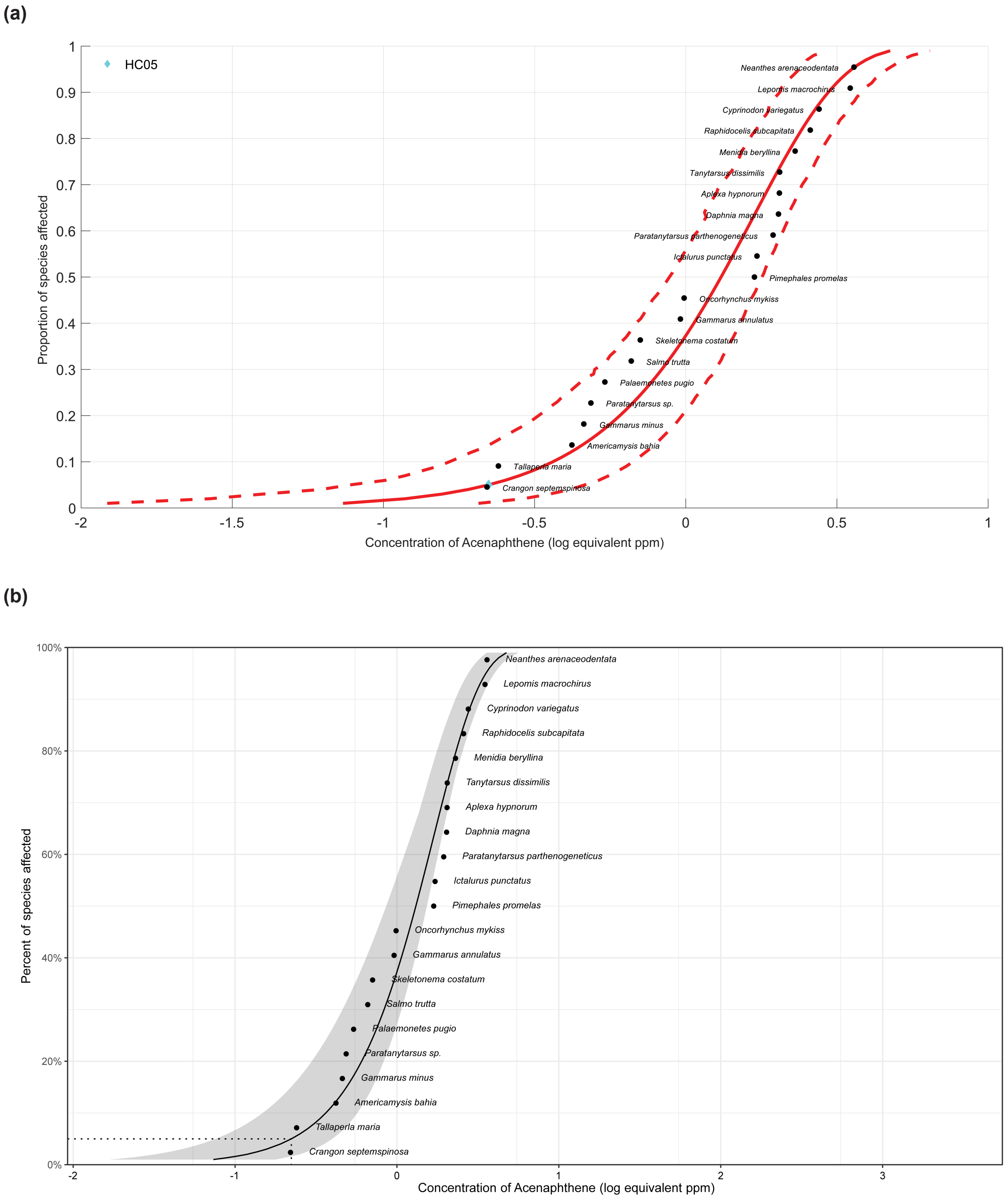


Figure 6.18: The plots of SSD for Acenaphthene computed using the best-fit Weibull model. (a) As determined by US EPA SSD Toolbox where the HC05 value is denoted by cyan colored diamond. (b) As determined by ssdtools where the HC05 value is denoted by a dotted line.

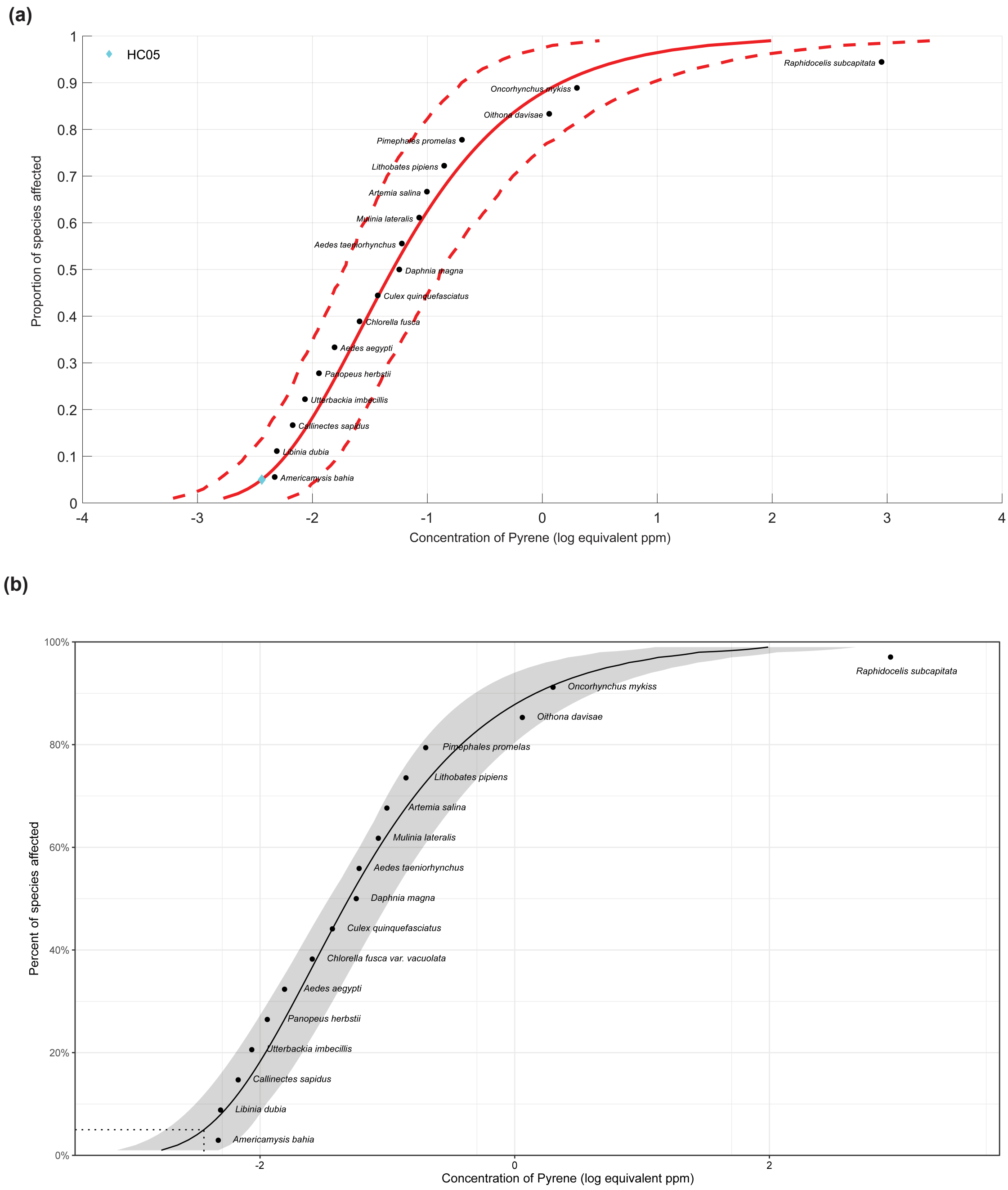


Figure 6.19: The plots of SSD for Pyrene computed using the best-fit Log-Gumbel model. (a) As determined by US EPA SSD Toolbox where the HC05 value is denoted by cyan colored diamond. (b) As determined by ssdtools where the HC05 value is denoted by a dotted line.

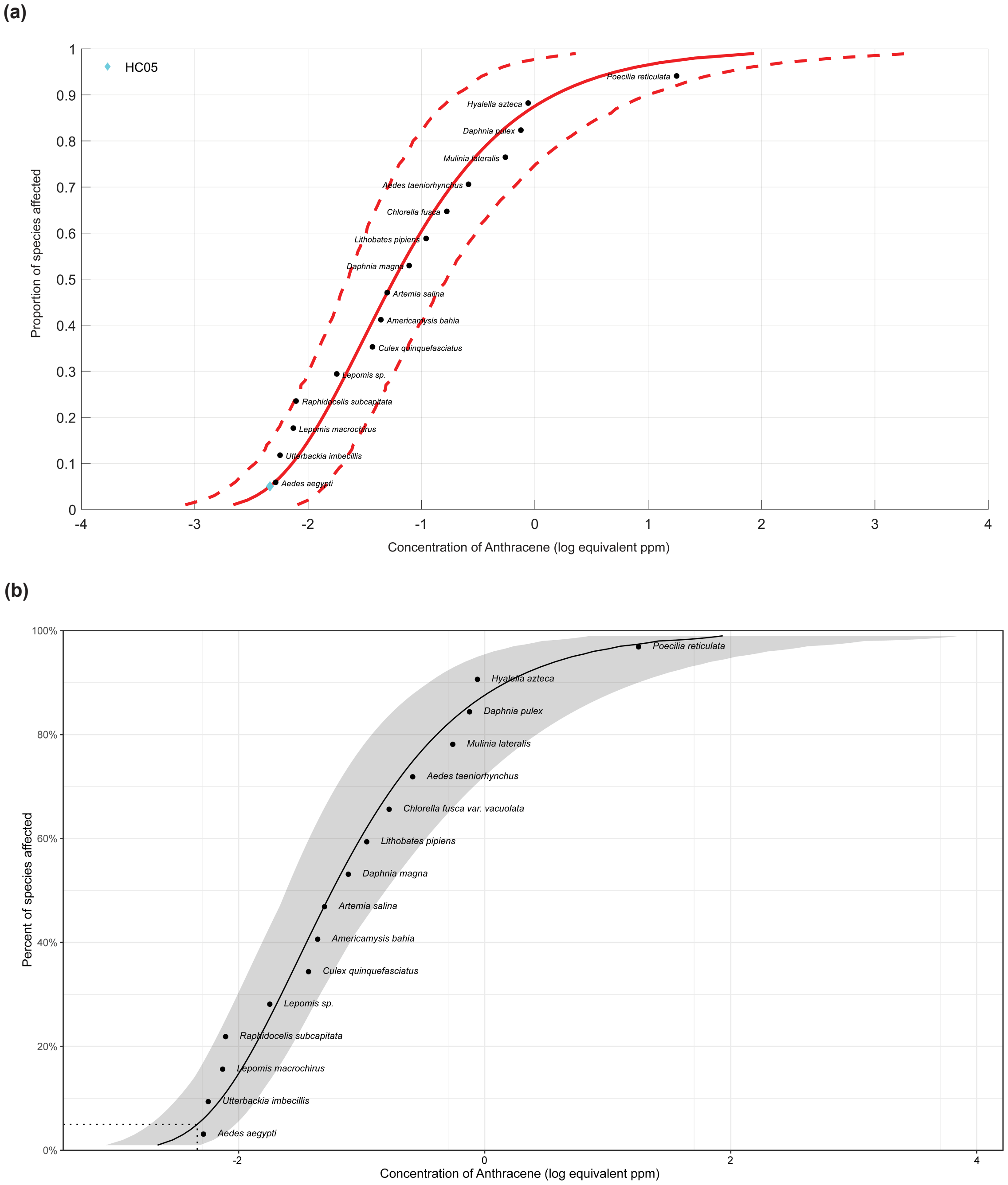


Figure 6.20: The plots of SSD for Anthracene computed using the best-fit Log-Gumbel model. (a) As determined by US EPA SSD Toolbox where the HC05 value is denoted by cyan colored diamond. (b) As determined by ssdtools where the HC05 value is denoted by a dotted line.

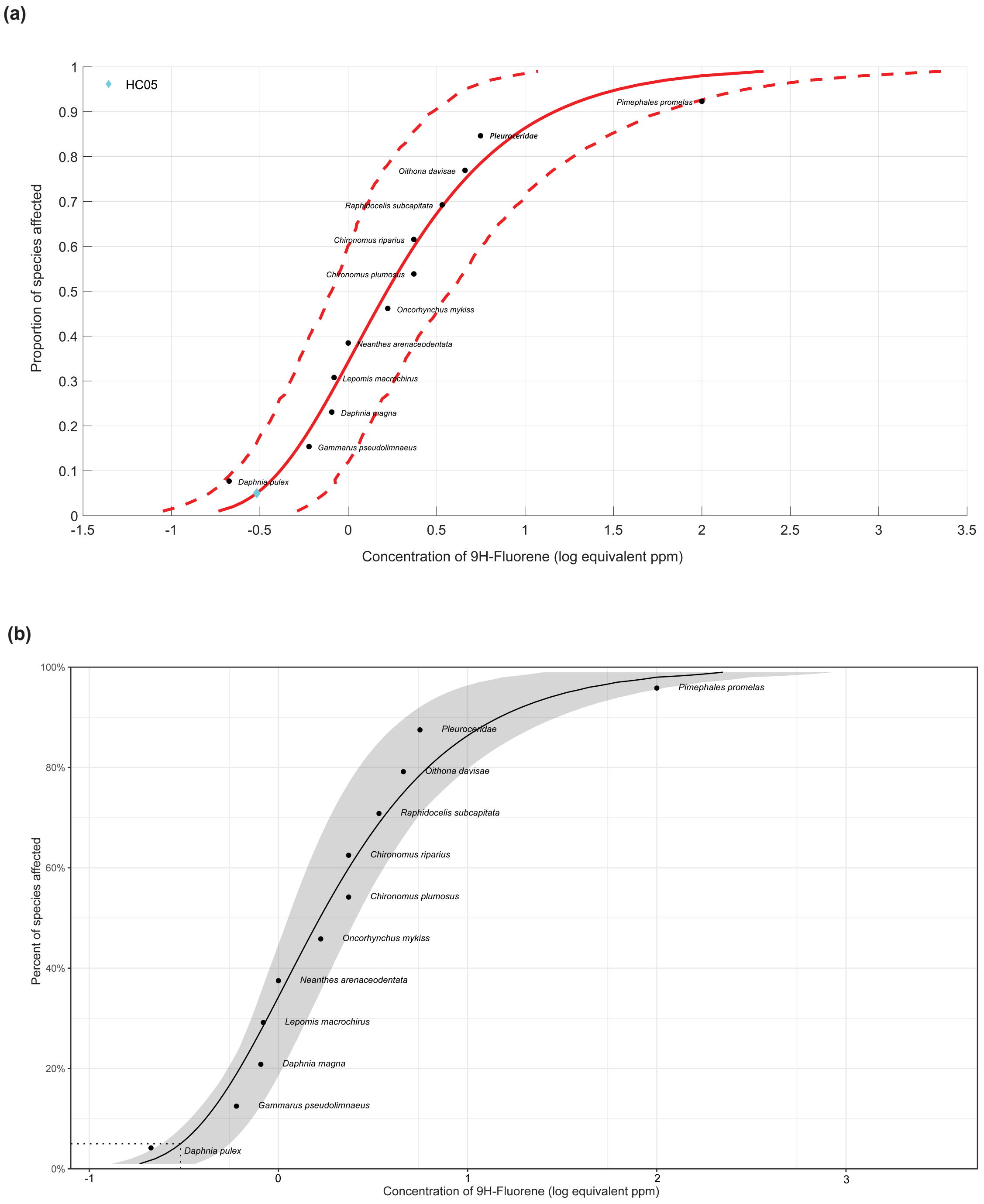


Figure 6.21: The plots of SSD for 9H-Fluorene computed using the best-fit Log-Gumbel model. (a) As determined by US EPA SSD Toolbox where the HC05 value is denoted by cyan colored diamond. (b) As determined by ssdtools where the HC05 value is denoted by a dotted line.

the region of lower toxicity in the SSDs computed for the eight PAHs. In particular, the species *Palaemonetes pugio* (crustacean), *Oncorhynchus mykiss* (fish), *Americamysis bahia* (crustacean), *Daphnia pulex* (crustacean) and *Utterbackia imbecillis* (mollusc) were sensitive to more than one PAH. Notably, these species are also connected with the PAHs in the stressor-species network constructed based on bioconcentration factors (Figure 6.12). Furthermore, we observed *P. pugio* is highly sensitive to B[a]P and Fluoranthene, and *A. bahia* is highly sensitive to Phenanthrene and Pyrene.

In a nutshell, we leveraged acute toxicity endpoints from ECOTOX to construct SSDs for the eight priority PAHs and derived the corresponding HC05 values using the model averaging method in both US EPA SSD Toolbox and ssdtools. The derived HC05 values were similar across both tools and helped identify a toxicity order for the eight PAHs.

6.3 Discussion

PHs are released into the environment through various human activities or accidental oil spills, where they can persist and pose long-term ecological risks. Thus, it is imperative to study the effects of PH exposure on species inhabiting the contaminated environments. In this chapter, we utilized network-based approaches to investigate PH-induced toxicity in ecological species (Figure 6.22). In addition to constructing stressor-AOP, undirected and directed AOP networks (similar to those reported in previous two Chapters 4 and 5), this chapter explored stressor-species networks, developed using toxicity concentration and bioconcentration factor data of PHs (Figure 6.22). While the AOP networks helped elucidate the adverse effects associated with the PH exposure, the stressor-species network revealed the species or species groups documented to be most affected by the PHs. Notably, constructing SSDs for priority PAHs not only helped derive their hazard concentrations that are not harmful to a large fraction of species but also revealed the species most sensitive to PAH exposure in aquatic environments.

To study the ecotoxicological effects induced by PHs, we relied on biological end-

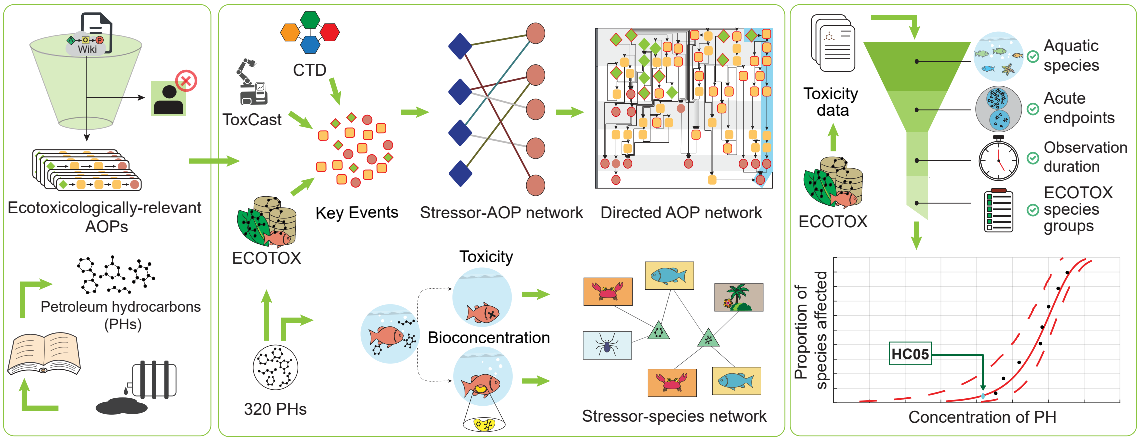


Figure 6.22: Schematic summary of our network-based investigation of PH-induced ecotoxicological effects and their risk assessment.

point data from ECOTOX and non-human endpoint data from other toxicological resources. To this end, we present a systematic workflow to curate ecotoxicologically-relevant high quality and complete AOPs by leveraging their taxonomic applicability information. Subsequently, integrating heterogeneous toxicological data, we constructed a stressor-AOP network associated with PH-induced ecotoxicity. Specifically, in the stressor-AOP network, incorporating toxicological data from ECOTOX allowed us to identify and expand the coverage of the stressor-AOP associations relevant to ecotoxicity. For example, the stressor-AOP network constructed in this chapter identified 29 highly relevant AOPs associated with B[a]P, compared to the 28 AOPs reported in Chapter 5, with 12 of these AOPs being common to both the studies.

However, we found that many AOPs in AOP-Wiki lack taxonomic applicability information, resulting in the curation of an incomplete set of ecotoxicologically-relevant AOPs. Further, we noted that the derived hazard concentrations for the priority PAHs may not be applicable to specific aquatic environments due to the limited information on test locations for the underlying toxicity data. Nonetheless, this chapter utilizes various network-based approaches along with toxicological data to elucidate the risks associated with PH exposure in ecosystem, thereby assisting in their effective regulation.

Supplementary Information

Supplementary Tables S6.1-S6.13 associated with this chapter are available for download from the GitHub repository: https://github.com/asamallab/PhDThesis-Ajaya_Kumar_Sahoo/blob/main/SI/ST_Chapter6.xlsx.

Code Availability

The computer programs used to perform the computations reported in this chapter are available in the following GitHub repository:

https://github.com/asamallab/PhDThesis-Ajaya_Kumar_Sahoo/tree/main/Codes.

Serial number	Chemical name	Model averaged HC05 value from SSD Toolbox	Model averaged HC05 standard error from SSD Toolbox	Model averaged HC05 value from ssdtools	Model averaged HC05 standard error from ssdtools
1	Fluoranthene	0.010708	0.002079	0.010712	0.002074
2	Naphthalene	0.69307	0.20508	0.693079	0.19385
3	Phenanthrene	0.053411	0.024022	0.053633	0.02168
4	Benzo[a]pyrene	0.000659	0.003752	0.000659	0.003746
5	Acenaphthene	0.22909	0.11273	0.22909	0.102988
6	Pyrene	0.003423	0.002693	0.003424	0.002594
7	Anthracene	0.00365	0.003635	0.003678	0.003424
8	9H-Fluorene	0.25044	0.17644	0.252645	0.166426

Table 6.1: Model averaged HC05 values and the corresponding standard errors computed by both US EPA SSD Toolbox and ssdtools for eight priority PAHs. The values are given in equivalent ppm units.

Chapter 7

Summary and future outlook

It is folly to think that we can destroy one species and ecosystem after another and not affect humanity. When we save species, we're actually saving ourselves.

- Joel Sartore

7.1 Summary

Endocrine disrupting chemicals (EDCs) target various protein receptors, can mimic or block the functioning of the natural hormones, and thereby disrupt normal physiological functions, which can lead to wide range of adverse health effects [13]. Therefore, it is crucial to accurately predict such EDCs in the chemical exposome. However, the heterogeneity in the structure-activity landscape of such environmental chemicals presents a significant challenge in the development of highly accurate predictive models. In this thesis, we address this knowledge gap by investigating the structure-activity landscape of environmental chemicals which can bind to two prominent endocrine receptors namely, androgen receptor (AR) and thyroid stimulating hormone receptor (TSHR). We employed several computational approaches to analyze the activity landscape of such environmental chemicals and identified activity cliffs in two separate studies (Figure 7.1; Chapters 2 and 3). Additionally, we analyzed the heterogeneity in the structure-mechanism relationships of the TSHR binding chemicals and identified mechanism of action (MOA)-cliffs (Figure 7.1; Chapter 3).

Further, we leveraged heterogeneous toxicological datasets and integrated them using computational approaches to associate the environmental chemicals with their adverse effects in humans and ecosystem. In particular, we focused on three different classes of environmental chemicals namely, heavy metal (cadmium), plastic additives and petroleum hydrocarbons (PHs) (Figure 7.2). Notably, these chemicals are well-known environmental contaminants commonly found in soil, air and aquatic ecosystems, and can cause a wide range of toxicities across multiple species, including humans. To investigate the adverse effects induced by these chemicals, we primarily relied on the adverse outcome pathway (AOP) framework, which enabled a systematic integration of relevant biological endpoints data from existing toxicological resources. Following an integrative data-centric approach, we constructed different AOP networks for inorganic cadmium (Chapter 4), and extended this approach to construct stressor-centric AOP networks for plastic additives (Chapter 5) and for PHs (Chapter 6) (Figure 7.2). These networks enabled exploration of chemical-induced adverse effects in humans and ecological species. Additionally, we constructed stressor-species networks for PHs which helped in assessing the ecological species most affected upon chemical exposure (Figure 7.2). Lastly, we utilized the toxicity endpoint data for PHs and performed their risk assessment in aquatic ecosystem. In sum, the research reported in this thesis employs network-based approaches along with integrative data-centric investigation to link the chemical exposome with human and ecosystem health. Below, we provide a summary and limitations of the research reported across different chapters of this thesis. Thereafter, we conclude with a short section on possible future directions of research based on work reported in this thesis.

In Chapter 2, we employed several computational approaches to visualize and explore the chemical space, and to compare the global diversity of the whole library (ALL) and its three identified clusters (C1, C2 and C3) among the 144 AR binding chemicals. Further, using the structure-activity similarity (SAS) map based approach, we identified 86 activity

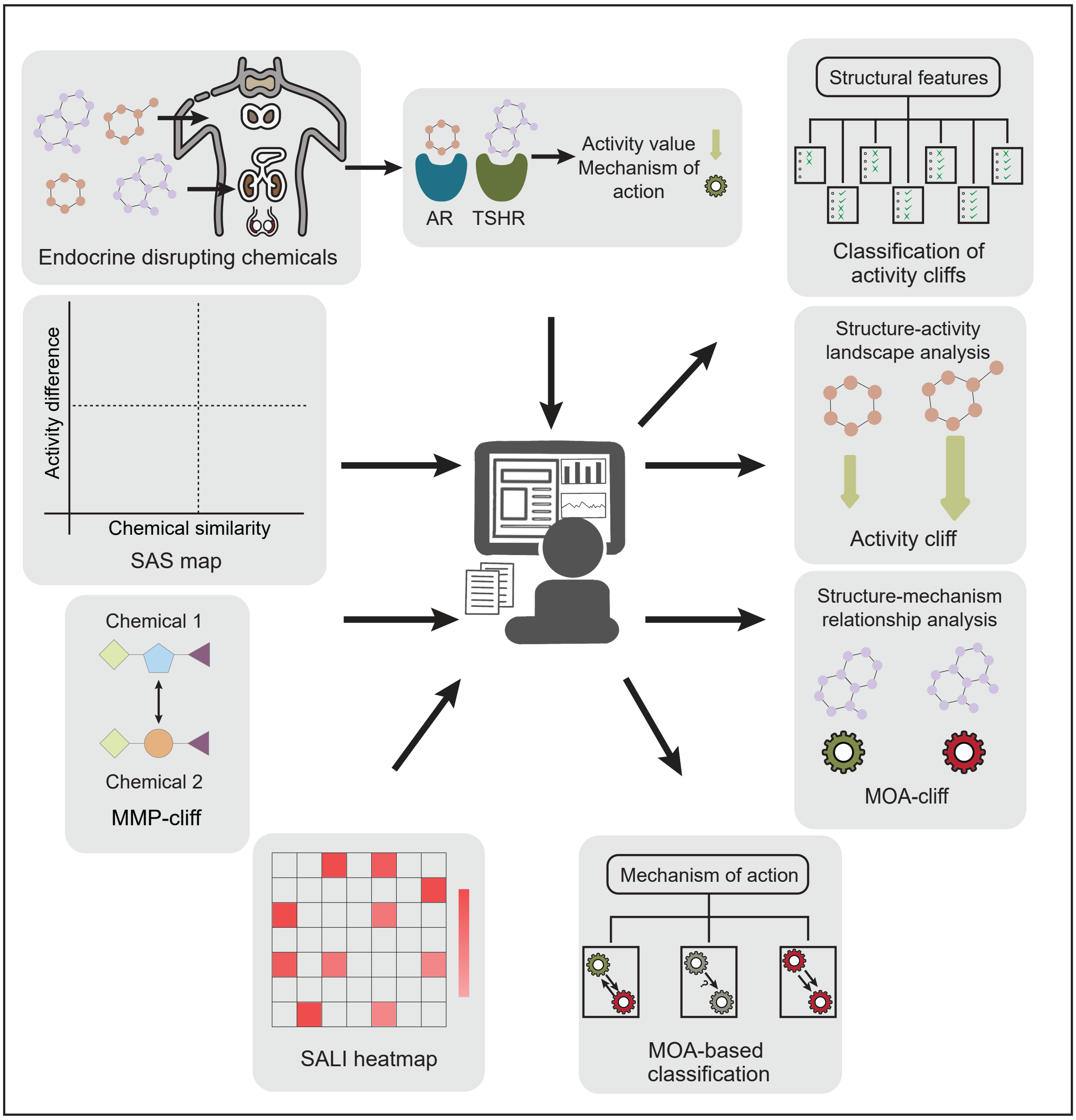


Figure 7.1: Schematic summary of our investigations of structure-activity and structure-mechanism relationships of environmental chemicals to identify activity cliffs and MOA-cliffs, respectively, reported in Chapters 2 and 3.

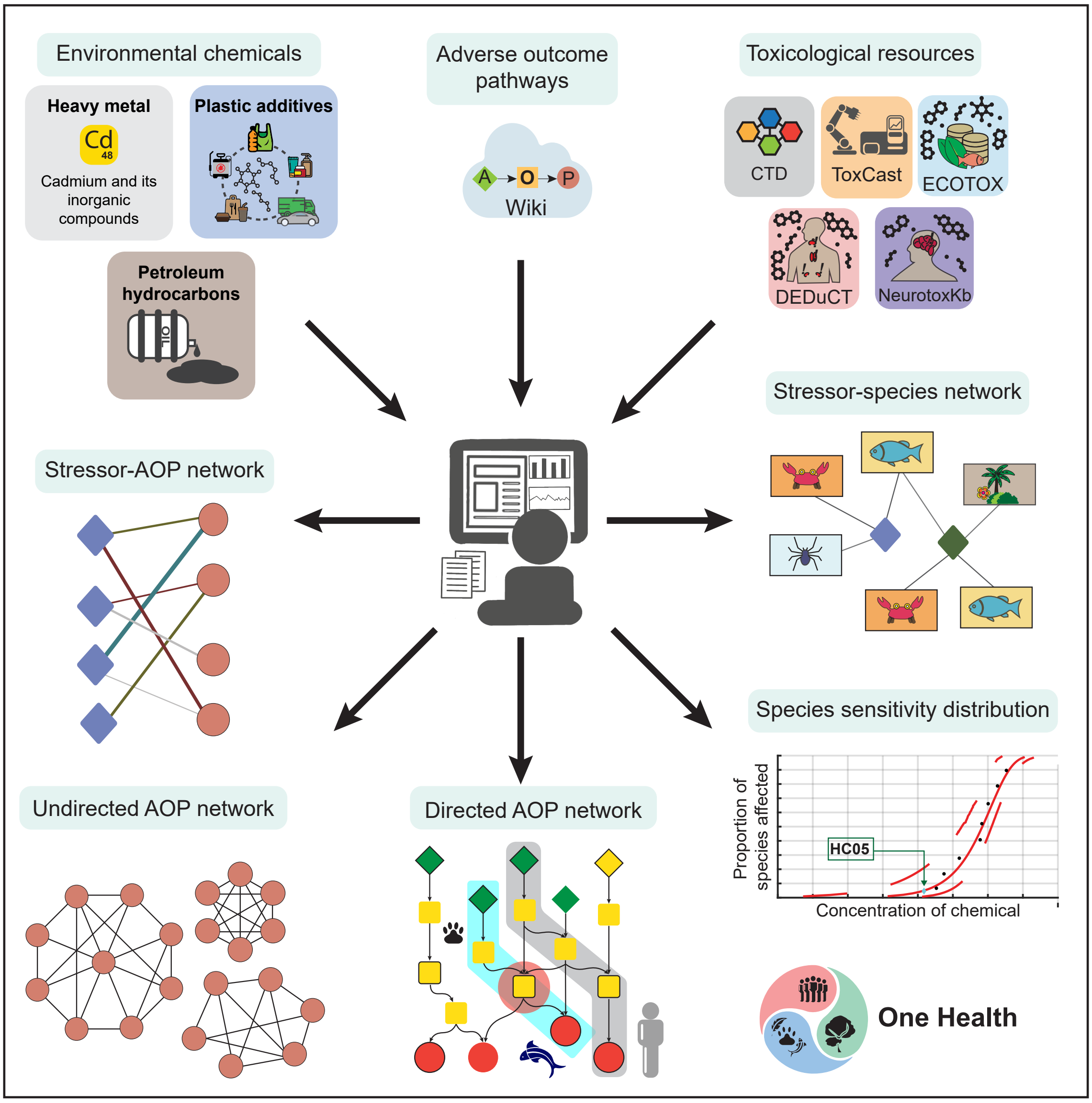


Figure 7.2: Schematic summary of our investigations of the chemical-induced toxicities in human and ecosystem using network-based and integrative toxicological data-centric approaches reported in Chapters 4, 5 and 6.

cliffs in the whole library and found that the chemicals forming activity cliffs belong exclusively to one of the three clusters, i.e., cluster C2 which is dominated by 'Steroids and steroid derivatives'. 14 chemicals that simultaneously form several activity cliffs were identified as activity cliff generators (ACGs). Additionally, a detailed inspection of the structure-activity landscape index (SALI) heatmap revealed that most of the activity cliffs identified from the SAS map have high SALI scores. Lastly, we classified the activity cliffs into six categories by considering the chemical structure information of the AR binders at different levels.

Importantly, we used three established computational approaches namely, a 2D visualization of the activity landscape (SAS map), a numerical scoring based approach (SALI), and structure based classification of activity cliffs, to reveal the structure-activity landscape of AR binders [36, 38, 39, 44, 347]. While these three methods have individually been reported in published literature for the identification of activity cliffs, we here employed a combined approach leveraging the three methods to best characterize the structure-activity landscape of the AR binders. From our analysis, we observed that SAS map, which takes into account the pairwise comparison of structural similarity and activity difference, seems a better way to analyze the heterogeneity of the activity landscapes. Further, the SALI based approach was helpful in numerically quantifying the activity cliffs. Though some chemical pairs with high SALI scores were not identified as activity cliffs from the SAS map, the majority of the chemical pairs with high SALI scores were identified as activity cliffs.

In their analysis of the activity landscape of estrogen receptor binding chemicals, Naveja *et al.* had additionally interpreted some of the identified activity cliffs from SAS map using experimentally determined estrogen receptor protein structures co-crystallized with chemicals forming activity cliff pairs [38]. Naveja *et al.* had performed this analysis with the aim of elucidating the molecular mechanisms behind the activity difference between a pair of estrogen receptor binding chemicals constituting an activity cliff. Though we would have also liked to perform a similar analysis, unfortunately there were no available experimentally determined co-crystallized AR protein structures for rat or mouse or human in the Protein Data Bank (PDB) [130] for any pair of AR binding chemicals that constitute an activity cliff identified from SAS map. On the other hand, we present the structural categorization of the activity cliffs proposed by Hu and Bajorath to aid in structural interpretation of the activity cliffs [44]. We believe that the insights from these analyses will be crucial in informing the presence of heterogeneity in the structure-activity landscape of environmental chemicals targeting AR, which will aid in the development of improved predictors of EDCs in the chemical exposome.

In Chapter 3, we explored and analyzed the activity landscape of chemicals in curated datasets of TSHR agonists (TSHR agonist dataset) and antagonists (TSHR antagonist dataset) compiled from the ToxCast library. By leveraging the established fingerprint-based approach and a substructure-based approach, we identified 79 activity cliffs in the TSHR agonist dataset and 69 activity cliffs in the TSHR antagonist dataset. Furthermore, we classified the resultant activity cliffs based on the information on chemical structures. Additionally, we analyzed the differences in the mechanism of action (MOA) of the TSHR binding chemicals and identified 3 Strong MOA-cliffs and 19 Weak MOA-cliffs.

However, our workflow does not account for the stereoisomeric information of the chemical structures in identification of activity cliffs and MOA-cliffs. Moreover, we were unable to quantify the differences in binding affinities of chemicals forming MOA-cliffs as their affinity values are obtained from two different assays. Like the case of activity cliffs among the AR binding chemicals in Chapter 2, we were unable to provide a mechanistic interpretation behind the formation of activity cliffs and MOA-cliffs due to the lack of co-crystallized protein-ligand complex for TSHR in public domain. Nonetheless, our efforts highlight the presence of activity cliffs and MOA-cliffs in a large chemical dataset such as ToxCast, and their identification will aid in development of robust toxicity predictors.

In Chapter 4, we constructed and analyzed an AOP network relevant to inorganic cadmium-induced toxicity. To construct the AOP network, we first extracted AOPs from AOP-Wiki [53] and systematically curated 309 high confidence AOPs. Simultaneously, we leveraged 5 exposome-relevant resources namely, AOP-Wiki, Comparative toxicogenomic database (CTD) [27], ToxCast [16], DEDuCT [8, 15] and NeurotoxKb [29], and integrated the heterogeneous data to identify 312 key events (KEs) present in AOP-Wiki to be associated with inorganic cadmium. Subsequently, we integrated the cadmium associated KEs with high confidence AOPs and identified 30 AOPs relevant for cadmium-induced toxicity (cadmium-AOPs). Thereafter, we constructed the AOP network using the 30 cadmium-AOPs and identified 3 connected components, with the largest component containing 18 cadmium-AOPs. We employed graph-theoretic approaches to analyze the 59 unique KEs present in the largest component and observed that the cadmium-induced molecular initiating event (MIE), 'Activation, AhR' (KE:18), leads to every adverse outcome (AO) present. Finally, we leveraged an artificial intelligence (AI) based tool namely AOP-helpFinder [183, 184] and Abstract Sifter [185] to curate supporting evidence for cadmium associations with each of the KEs present in the largest component.

However, we focused only on AOPs from AOP-Wiki to construct the cadmium-AOP network. Among the 18 cadmium-AOPs within the largest component, only 2 AOPs (AOP:21 and AOP:150) have been endorsed by Organisation for Economic Co-operation and Development (OECD). We also observed that 9 of these 18 cadmium-AOPs do not compile any evidence for their key event relationships (KERs). Upon closer inspection, we noted that some of the KEs lacked action information or were duplicated, and some KERs directly linked MIE to AO. This can be attributed to the fact that many of these AOPs are under development. Furthermore, we observed that only 66 of the 163 disease terms from CTD associated with cadmium toxicity were mapped to KEs within AOP-Wiki. This could be attributed to the fact that AOP-Wiki is a collaborative resource with contributions from research groups across the globe, each with varied interests. Therefore, it may not exhaustively capture AOPs for all possible adverse outcomes induced by cadmium toxicity.

Nonetheless, AOP-Wiki is the most up-to-date and comprehensive resource on AOPs developed globally, and therefore we leveraged AOP-Wiki to present the first ever AOP network specific to cadmium-induced toxicity. Our integrative data-centric approach helped in identifying KEs (including MIEs and AOs) associated with inorganic cadmium, which were otherwise not documented within AOP-Wiki. We additionally provide auxiliary evidence for the association of KEs with inorganic cadmium in the directed AOP network. Further experimentation is required to characterize the points-of-departure of cadmium toxicity which will aid in strengthening its regulation. In sum, the derivation and characterization of the AOP network in Chapter 4 will aid in the regulation of cadmium and its inorganic compounds.

In Chapter 5, we constructed and analyzed stressor-AOP network to explore the toxicities induced by plastic additives. We first relied on the United Nations Environment Programme (UNEP) report titled 'Chemicals in Plastics – A Technical Report' [20] and identified 6470 plastic additives based on the reported chemical functions. Next, we systematically integrated heterogeneous toxicogenomics and biological endpoints data from five exposome-relevant resources namely, ToxCast, CTD, DEDuCT, NeurotoxKb and AOP-Wiki, and identified 688 KEs within AOP-Wiki to be associated with 1314 plastic additives. Further, we systematically curated 328 high confidence AOPs from AOP-Wiki and linked them to plastic additives based on overlapping KE associations. In this study, we identified 322 high confidence AOPs to be associated with 1287 plastic additives while AOP-Wiki only documented 37 of the 1287 plastic additives to be associated with 27 of the 322 high confidence AOPs. Next, we constructed the stressor-AOP network for plastic additives (plastic additives-AOP network) with varying levels of associations, where the plastic additives are categorized into 10 priority use sectors and the AOPs are linked with 27 disease classes. We visualized the plastic additives-AOP network for each of the 1287 plastic additives and made them available in a dedicated website: https://cb.imsc.res.in/saopadditives/. Finally, we showed the utility of the constructed plastic additives-AOP network by identifying highly relevant AOPs (with Level 5 relevance and coverage score threshold of 0.4) associated with plastic additives. In particular, we identified highly relevant AOPs associated with Benzo[a]pyrene (B[a]P), Bisphenol A (BPA) and bis(2-ethylhexyl) phthalate (DEHP), and relied on published experimental evidence to explore human- and ecotoxicology-relevant toxicity pathways.

However, the functional annotations of chemicals as plastic additives provided by the UNEP report may be inaccurate [227]. For example, B[a]P is annotated as a plasticizer, cross-linker, lubricant and filler in the UNEP report, but it has been reported as a byproduct or a contaminant resulting from the use of other plastic additives during plastic production [254, 255]. Similarly, other chemicals may have been misidentified due to inaccuracies in the functional annotations provided by the UNEP report. Moreover, due to limited information on their presence in various use sectors, we were able to identify only 4309 of the 6470 plastic additives across 10 priority use sectors. Further, due to the paucity of plastic additive exposure studies, we were able to associate only 1287 of the 6470 plastic additives to AOPs within AOP-Wiki. We observed that 197 of the 322 high confidence AOPs (associated with the 1287 plastic additives) capture toxicity pathways leading to human relevant adverse effects. Moreover, the scope of this study is limited to the toxicological events in humans and other mammals as the toxicogenomics approach primarily relied on mammalian-centric biological data.

Nonetheless, Chapter 5 presents the first and most comprehensive stressor-AOP network for plastic additives. The constructed plastic additives-AOP network was useful in the identification of highly relevant AOPs for plastic additives which highlighted plastic additives-induced emergent toxicity pathways. In sum, Chapter 5 utilizes the AOP framework to explore the various adverse effects associated with plastic additives, assisting in their risk assessment and contributing towards their regulatory decision-making.

In Chapter 6, we leveraged network-based approaches along with ecotoxicological data to investigate petroleum hydrocarbon (PH)-induced toxicities and the associated risks for ecological species. We relied on the reports published by the Total Petroleum Hydrocarbon Criteria Working Group (TPHCWG) [97, 98] and curated a list of 320 PHs that were experimentally identified to be present in various fuel oils. We utilized this list to explore the mechanism of PH-induced ecotoxicity, and specifically focused on the 16 polycyclic aromatic hydrocarbons (PAHs) prioritized by United States Environmental Protection Agency (US EPA), and assessed their risks in aquatic environments. First, we curated a list of 195 ecotoxicologically-relevant AOPs from AOP-Wiki. Subsequently, we systematically integrated biological endpoint data from three sources namely, ToxCast, CTD and ECOTOX [28], and identified 206 KEs from 177 ecotoxicologically-relevant AOPs to be associated with 75 of 320 PHs. We constructed a stressor-centric AOP network comprising these 75 PHs and 177 ecotoxicologically-relevant AOPs linked through 3265 edges with varying levels of relevance and coverage scores. We leveraged this stressor-AOP network and identified 29 AOPs relevant for B[a]P-induced toxicity, constructed an AOP network, and performed a case study to understand its transgenerational effects in ecological species. Notably, compared to the 28 highly relevant B[a]P associated AOPs (B[a]P-AOPs) identified in Chapter 5, incorporating toxicological data from ECOTOX expanded the coverage to 29 B[a]P-AOPs, with 12 AOPs common to networks constructed in both chapters.

Next, we utilized the acute toxicity data within ECOTOX, constructed a stressor-species network comprising 80 PHs linked to 221 species, and observed that 'Crustaceans' species group was documented to be affected by many of these PHs. Similarly, we utilized the bioconcentration factors data within ECOTOX, constructed a stressor-species network comprising 28 PHs linked to 59 species, and observed that 'Crustaceans' species group was documented to bioaccumulate many of these PHs. Finally, we utilized the acute toxicity data within ECOTOX available for eight EPA priority PAHs, constructed their Species Sensitivity Distributions (SSDs), and derived corresponding hazard concentrations (HC05) that is not harmful to 95% of the species in aquatic environments.

However, the scope of the study reported in Chapter 6 is restricted by several limitations on the available data. The curated list of PHs may be incomplete due to the limitations of current analytical methods in determining the full molecular composition of fuel oils [348]. We observed that many of the AOPs within AOP-Wiki have no taxonomic applicability annotation thereby leading to the curation of an incomplete set of ecotoxicologically-relevant AOPs. Further, we observed that the derived HC05 value for some of the EPA priority PAHs have wider confidence intervals implying the need for more stressor-specific toxicity studies across diverse ecological species in order to accurately derive the hazard concentrations [349]. Importantly, the derived hazard concentrations of the priority PAHs may not be applicable to a specific aquatic environment as the information on the test location for the underlying toxicity data was sparse.

Nonetheless, Chapter 6 advances our understanding of the ecotoxicological effects of individual PHs by leveraging network-based approaches. The stressor-AOP network constructed using ecotoxicologically-relevant endpoints for PHs has facilitated the investigation of toxicity pathways leading to PH-specific adverse outcomes. Further, the inferences from the study reported in this chapter using acute toxicity data can aid in assessing the risks associated with events such as oil spills, which result in a sudden increase in PH concentrations in ecosystems. In sum, Chapter 6 explores the ecotoxicological effects and risks associated with PH exposure in ecosystems, thereby assisting in their regulation and enabling the formulation of effective mitigation and remediation strategies for various PH contamination.

7.2 Future outlook

> *To halt the decline of an ecosystem, it is necessary to think like an ecosystem.*
>
> \- Douglas P. Wheeler

Of late, the development of new scientific methods has led to significant increase in the research on chemical exposome, and this has led to deeper insights into the effects of environmental chemical exposures on both human and ecosystem health. In particular, computational approaches enabling integration of existing high quality toxicological datasets have become powerful tools for investigating environmental chemicals and their potential harmful effects [10, 12]. To this end, in this thesis, we present our computational investigations of the space of environmental chemicals and their adverse health impacts.

In Chapters 2 and 3, we employed established computational methods to investigate the heterogeneity in the structure-activity and structure-mechanism relationships of chemicals binding to specific endocrine receptors. In future, these investigations can be extended to analyze multitarget activity landscape [350] for environmental chemicals targeting multiple endocrine receptors. Additionally, the recently developed chemical similarity methods, such as extended similarity indices (n-ary comparison), which has potential to simultaneously compare more than two chemicals, can be used to address the computational complexity arising from the pairwise comparison of chemicals in large datasets [351, 352]. Further, quantitative structure–activity relationship (QSAR)-based machine learning (ML) models have been reported to perform poorly in predicting activity cliffs among chemicals, often leading to errors in overall predictions [353]. Towards this, van Tilborg *et al.* [354] have evaluated the performance of different ML models in predicting biological activity of thousands of chemicals. They recommended using 'activity-cliff-centered' metrics instead of traditional fingerprints to better capture the discontinuities in the structure-activity relationship, in order to improve the performance of the model [354]. Moreover, Dablander *et al.* [353] introduced a data splitting technique in their QSAR models, which improved the prediction of activity cliffs in specific scenarios. Thus, the analyses reported in Chapters 2 and 3 will provide insights into the structural features of the activity cliffs, which in turn will lead to better strategies for the development of models with high predictive power.

Furthermore, in Chapters 4, 5 and 6, this thesis utilizes the AOP framework in conjunction with diverse toxicological datasets to explore the adverse health effects of diverse environmental chemicals. In Chapter 4, we have built the first AOP network for the prominent heavy metal cadmium, and a similar workflow can be employed to construct AOP networks for other important heavy metals such as arsenic, lead and mercury. Additionally, the workflow to build AOP networks for heavy metals can be adapted to incorporate toxicity data from ECOTOX [28] for capturing the ecotoxicological effects associated with heavy metals. In Chapter 5, we have built stressor-AOP networks for 1287 plastic additives, and in future, this network can be expanded by using a recently published larger and more detailed dataset of plastic additives [355]. Moreover, the stressor-AOP networks can be enhanced by incorporating toxicity data from ECOTOX [28] to better capture the ecotoxicological effects of plastic additives. Importantly, AOP networks can be used to study the toxic effects of chemical mixtures [58], and the networks constructed in Chapter 6 can similarly be leveraged to investigate the effects of oil contamination by considering mixtures of different PHs. Moreover, the compiled dosage information of chemicals can provide empirical support for dose-response relationship, and thereby, enable the development of quantitative AOPs that could potentially aid in regulatory decision-making of the environmental chemicals [212, 213] studied in this thesis. Notably, AOP-based ML models can screen chemicals for similar toxicity mechanisms by utilizing the chemical structures and their interactions with biological targets [356]. Additionally, such models can incorporate *in vitro* data, including the mode of action of chemicals to predict potential adverse outcomes which will aid in chemical risk assessment [356]. Lastly, the data-centric approaches employed in this thesis to analyze chemical-induced toxicities can also be adapted to investigate the adverse effects associated with exposure to non-chemical stressors. [357]. We believe that, the detailed computational analyses of the environmental chemicals and their health effects presented in this thesis, provide valuable insights into the associated risks for both human and ecological species, and moreover, the acquired insights can assist in chemical regulation. In sum, this thesis presents a systematic investigation of diverse environmental chemical spaces and their adverse impacts on human and ecosystem health, thereby providing a holistic overview of the chemical exposome and its implications on health from a ‘One Health’ [358] perspective (Figure 7.2).

References

[1] Wild, C. P. Complementing the Genome with an “Exposome”: The Outstanding Challenge of Environmental Exposure Measurement in Molecular Epidemiology. *Cancer Epidemiology, Biomarkers & Prevention* **14**, 1847–1850 (2005).

[2] Wei, X., Huang, Z., Jiang, L., Li, Y., Zhang, X., Leng, Y. & Jiang, C. Charting the landscape of the environmental exposome. *iMeta* **1**, e50 (2022).

[3] Miller, G. W. *The exposome: a new paradigm for the environment and health (Second Edition)* (Academic Press, 2020).

[4] Wild, C. P. The exposome: from concept to utility. *International Journal of Epidemiology* **41**, 24–32 (2012).

[5] Samanipour, S., Barron, L. P., van Herwerden, D., Praetorius, A., Thomas, K. V. & O’Brien, J. W. Exploring the Chemical Space of the Exposome: How Far Have We Gone? *JACS Au* **4**, 2412–2425 (2024).

[6] Meeker, J. D., Sathyanarayana, S. & Swan, S. H. Phthalates and other additives in plastics: human exposure and associated health outcomes. *Philosophical Transactions of the Royal Society B: Biological Sciences* **364**, 2097–2113 (2009).

[7] Landis, W., Sofield, R., Yu, M.-H., Landis, W. G. & Yu, M.-H. *Introduction to Environmental Toxicology* (CRC Press, 2003).

[8] Karthikeyan, B. S., Ravichandran, J., Mohanraj, K., Vivek-Ananth, R. & Samal, A. A curated knowledgebase on endocrine disrupting chemicals and their biolog-

ical systems-level perturbations. *Science of the Total Environment* **692**, 281–296 (2019).

[9] Beyer, J., Trannum, H. C., Bakke, T., Hodson, P. V. & Collier, T. K. Environmental effects of the Deepwater Horizon oil spill: A review. *Marine Pollution Bulletin* **110**, 28–51 (2016).

[10] National Research Council. *Toxicity Testing in the 21st Century: A Vision and a Strategy* (The National Academies Press, Washington, DC, 2007).

[11] Rusyn Ivan & Daston George P. Computational Toxicology: Realizing the Promise of the Toxicity Testing in the 21st Century. *Environmental Health Perspectives* **118**, 1047–1050 (2010).

[12] National Research Council. *Using 21st Century Science to Improve Risk-Related Evaluations* (The National Academies Press, Washington, DC, 2017).

[13] Diamanti-Kandarakis, E., Bourguignon, J.-P., Giudice, L. C., Hauser, R., Prins, G. S., Soto, A. M., Zoeller, R. T. & Gore, A. C. Endocrine-Disrupting Chemicals: An Endocrine Society Scientific Statement. *Endocrine Reviews* **30**, 293–342 (2009).

[14] Bertram, M. G., Gore, A. C., Tyler, C. R. & Brodin, T. Endocrine-disrupting chemicals. *Current Biology* **32**, R727–R730 (2022).

[15] Karthikeyan, B. S., Ravichandran, J., Aparna, S. & Samal, A. DEDuCT 2.0: An updated knowledgebase and an exploration of the current regulations and guidelines from the perspective of endocrine disrupting chemicals. *Chemosphere* **267**, 128898 (2021).

[16] Dix, D. J., Houck, K. A., Martin, M. T., Richard, A. M., Setzer, R. W. & Kavlock, R. J. The ToxCast Program for Prioritizing Toxicity Testing of Environmental Chemicals. *Toxicological Sciences* **95**, 5–12 (2007).

[17] Tchounwou, P. B., Yedjou, C. G., Patlolla, A. K. & Sutton, D. J. Heavy Metal Toxicity and the Environment. In Luch, A. (ed.) *Molecular, Clinical and Environmental Toxicology: Volume 3: Environmental Toxicology*, 133–164 (Springer Basel, 2012).

[18] Bello, S., Nasiru, R., Garba, N. & Adeyemo, D. Carcinogenic and non-carcinogenic health risk assessment of heavy metals exposure from Shanono and Bagwai artisanal gold mines, Kano state, Nigeria. *Scientific African* **6**, e00197 (2019).

[19] Geyer, R., Jambeck, J. R. & Law, K. L. Production, use, and fate of all plastics ever made. *Science Advances* **3**, e1700782 (2017).

[20] UNEP. Chemicals in Plastics - A Technical Report. https://www.unep.org/resources/report/chemicals-plastics-technical-report (2023).

[21] Maddela, N. R., Kakarla, D., Venkateswarlu, K. & Megharaj, M. Additives of plastics: Entry into the environment and potential risks to human and ecological health. *Journal of Environmental Management* **348**, 119364 (2023).

[22] Kuppusamy, S., Maddela, N. R., Megharaj, M. & Venkateswarlu, K. An Overview of Total Petroleum Hydrocarbons. In Kuppusamy, S., Maddela, N. R., Megharaj, M. & Venkateswarlu, K. (eds.) *Total Petroleum Hydrocarbons: Environmental Fate, Toxicity, and Remediation*, 1–27 (Springer International Publishing, Cham, 2020).

[23] Almeda, R., Wambaugh, Z., Chai, C., Wang, Z., Liu, Z. & Buskey, E. J. Effects of Crude Oil Exposure on Bioaccumulation of Polycyclic Aromatic Hydrocarbons and Survival of Adult and Larval Stages of Gelatinous Zooplankton. *PLoS ONE* **8**, e74476 (2013).

[24] Almeda, R., Wambaugh, Z., Wang, Z., Hyatt, C., Liu, Z. & Buskey, E. J. Interactions between Zooplankton and Crude Oil: Toxic Effects and Bioaccumulation of Polycyclic Aromatic Hydrocarbons. *PLoS ONE* **8**, e67212 (2013).

[25] Pritsos, K. L. *et al.* Dietary intake of Deepwater Horizon oil-injected live food fish by double-crested cormorants resulted in oxidative stress. *Ecotoxicology and Environmental Safety* **146**, 62–67 (2017).

[26] Takeshita, R. *et al.* A review of the toxicology of oil in vertebrates: what we have learned following the Deepwater Horizon oil spill. *Journal of Toxicology and Environmental Health, Part B* **24**, 355–394 (2021).

[27] Davis, A. P., Wiegers, T. C., Johnson, R. J., Sciaky, D., Wiegers, J. & Mattingly, C. Comparative Toxicogenomics Database (CTD): update 2023. *Nucleic Acids Research* **51**, D1257–D1262 (2023).

[28] Olker, J. H. *et al.* The ECOTOXicology Knowledgebase: A Curated Database of Ecologically Relevant Toxicity Tests to Support Environmental Research and Risk Assessment. *Environmental Toxicology and Chemistry* **41**, 1520–1539 (2022).

[29] Ravichandran, J., Karthikeyan, B. S., Singla, P., Aparna, S. & Samal, A. NeurotoxKb 1.0: Compilation, curation and exploration of a knowledgebase of environmental neurotoxicants specific to mammals. *Chemosphere* **278**, 130387 (2021).

[30] McKinney, J. D., Richard, A., Waller, C., Newman, M. C. & Gerberick, F. The Practice of Structure Activity Relationships (SAR) in Toxicology. *Toxicological Sciences* **56**, 8–17 (2000).

[31] Guha, R. Exploring structure–activity data using the landscape paradigm. *WIREs Computational Molecular Science* **2**, 829–841 (2012).

[32] Cruz-Monteagudo, M., Medina-Franco, J. L., Pérez-Castillo, Y., Nicolotti, O., Cordeiro, M. N. D. & Borges, F. Activity cliffs in drug discovery: Dr Jekyll or Mr Hyde? *Drug Discovery Today* **19**, 1069–1080 (2014).

[33] Maggiora, G. M. On Outliers and Activity Cliffs–Why QSAR Often Disappoints. *Journal of Chemical Information and Modeling* **46**, 1535 (2006).

[34] Medina-Franco, J. L. Activity Cliffs: Facts or Artifacts? *Chemical Biology & Drug Design* **81**, 553–556 (2013).

[35] Shanmugasundaram, V. & Maggiora, G. Characterizing property and activity landscapes using an information-theoretic approach. vol. 222, Abstract No. 77 (222nd American Chemical Society National Meeting, Division of Chemical Information, 2001).

[36] Méndez-Lucio, O., Pérez-Villanueva, J., Castillo, R. & Medina-Franco, J. L. Identifying Activity Cliff Generators of PPAR Ligands Using SAS Maps. *Molecular Informatics* **31**, 837–846 (2012).

[37] Naveja, J. J. & Medina-Franco, J. L. Activity landscape sweeping: insights into the mechanism of inhibition and optimization of DNMT1 inhibitors. *RSC Advances* **5**, 63882–63895 (2015).

[38] Naveja, J. J., Norinder, U., Mucs, D., López-López, E. & Medina-Franco, J. L. Chemical space, diversity and activity landscape analysis of estrogen receptor binders. *RSC Advances* **8**, 38229–38237 (2018).

[39] Guha, R. & Van Drie, J. H. Structure–activity landscape index: Identifying and quantifying activity cliffs. *Journal of Chemical Information and Modeling* **48**, 646–658 (2008).

[40] Wassermann, A. M., Wawer, M. & Bajorath, J. Activity Landscape Representations for StructureActivity Relationship Analysis. *Journal of Medicinal Chemistry* **53**, 8209–8223 (2010).

[41] Hussain, J. & Rea, C. Computationally Efficient Algorithm to Identify Matched Molecular Pairs (MMPs) in Large Data Sets. *Journal of Chemical Information and Modeling* **50**, 339–348 (2010).

[42] Dalke, A., Hert, J. & Kramer, C. mmpdb: An Open-Source Matched Molecular Pair Platform for Large Multiproperty Data Sets. *Journal of Chemical Information and Modeling* **58**, 902–910 (2018).

[43] Hu, X., Hu, Y., Vogt, M., Stumpfe, D. & Bajorath, J. MMP-Cliffs: Systematic Identification of Activity Cliffs on the Basis of Matched Molecular Pairs. *Journal of Chemical Information and Modeling* **52**, 1138–1145 (2012).

[44] Hu, Y. & Bajorath, J. Extending the Activity Cliff Concept: Structural Categorization of Activity Cliffs and Systematic Identification of Different Types of Cliffs in the ChEMBL Database. *Journal of Chemical Information and Modeling* **52**, 1806–1811 (2012).

[45] Vivek-Ananth, R., Sahoo, A. K., Baskaran, S. P., Ravichandran, J. & Samal, A. Identification of activity cliffs in structure-activity landscape of androgen receptor binding chemicals. *Science of the Total Environment* **873**, 162263 (2023).

[46] Sahoo, A. K., Baskaran, S. P., Chivukula, N., Kumar, K. & Samal, A. Analysis of structure–activity and structure–mechanism relationships among thyroid stimulating hormone receptor binding chemicals by leveraging the ToxCast library. *RSC Advances* **13**, 23461–23471 (2023).

[47] Hao, M., Bryant, S. H. & Wang, Y. Cheminformatics analysis of the AR agonist and antagonist datasets in PubChem. *Journal of Cheminformatics* **8**, 37 (2016).

[48] Ankley, G. T. *et al.* Adverse outcome pathways: A conceptual framework to support ecotoxicology research and risk assessment. *Environmental Toxicology and Chemistry* **29**, 730–741 (2010).

[49] Tollefsen, K. E. *et al.* Applying Adverse Outcome Pathways (AOPs) to support Integrated Approaches to Testing and Assessment (IATA). *Regulatory Toxicology and Pharmacology* **70**, 629–640 (2014).

[50] Ankley, G. T. & Edwards, S. W. The adverse outcome pathway: A multifaceted framework supporting 21st century toxicology. *Current Opinion in Toxicology* **9**, 1–7 (2018).

[51] Jin, Y. *et al.* High throughput data-based, toxicity pathway-oriented development of a quantitative adverse outcome pathway network linking AHR activation to lung damages. *Journal of Hazardous Materials* **425**, 128041 (2022).

[52] Sakuratani, Y., Horie, M. & Leinala, E. Integrated Approaches to Testing and Assessment: OECD Activities on the Development and Use of Adverse Outcome Pathways and Case Studies. *Basic & Clinical Pharmacology & Toxicology* **123**, 20–28 (2018).

[53] AOP-Wiki. https://aopwiki.org/.

[54] Villeneuve, D. L. *et al.* Adverse Outcome Pathway (AOP) Development I: Strategies and Principles. *Toxicological Sciences* **142**, 312–320 (2014).

[55] Villeneuve, D. L. *et al.* Adverse Outcome Pathway Development II: Best Practices. *Toxicological Sciences* **142**, 321–330 (2014).

[56] Vinken, M., Knapen, D., Vergauwen, L., Hengstler, J. G., Angrish, M. & Whelan, M. Adverse outcome pathways: a concise introduction for toxicologists. *Archives of Toxicology* **91**, 3697–3707 (2017).

[57] OECD. Revised Guidance Document on Developing and Assessing Adverse Outcome Pathways, Series on Testing and Assessment No. 184. *OECD Publishing, Paris* (2017).

[58] Knapen, D. *et al.* Adverse outcome pathway networks I: Development and applications. *Environmental Toxicology and Chemistry* **37**, 1723–1733 (2018).

[59] Villeneuve, D. L. *et al.* Adverse outcome pathway networks II: Network analytics. *Environmental Toxicology and Chemistry* **37**, 1734–1748 (2018).

[60] Coady, K., Browne, P., Embry, M., Hill, T., Leinala, E., Steeger, T., Maślankiewicz, L. & Hutchinson, T. When Are Adverse Outcome Pathways and Associated Assays "Fit for Purpose" for Regulatory Decision-Making and Management of Chemicals? *Integrated Environmental Assessment and Management* **15**, 633–647 (2019).

[61] Hecker, M. & LaLone, C. A. Adverse Outcome Pathways: Moving from a Scientific Concept to an Internationally Accepted Framework. *Environmental Toxicology and Chemistry* **38**, 1152–1163 (2019).

[62] Khadka, K. K., Chen, M., Liu, Z., Tong, W. & Wang, D. Integrating adverse outcome pathways (AOPs) and high throughput in vitro assays for better risk evaluations, a study with drug-induced liver injury (DILI). *ALTEX - Alternatives to animal experimentation* **37**, 187–196 (2020).

[63] Pollesch, N. L., Villeneuve, D. L. & O'Brien, J. M. Extracting and Benchmarking Emerging Adverse Outcome Pathway Knowledge. *Toxicological Sciences* **168**, 349–364 (2019).

[64] Chai, Z. *et al.* Generating adverse outcome pathway (AOP) of inorganic arsenic-induced adult male reproductive impairment via integration of phenotypic analysis in comparative toxicogenomics database (CTD) and AOP wiki. *Toxicology and Applied Pharmacology* **411**, 115370 (2021).

[65] Howdeshell, K. L., Hotchkiss, A. K. & Gray, L. E. Cumulative effects of antiandrogenic chemical mixtures and their relevance to human health risk assessment. *International Journal of Hygiene and Environmental Health* **220**, 179–188 (2017).

[66] Knapen, D., Vergauwen, L., Villeneuve, D. L. & Ankley, G. T. The potential of AOP networks for reproductive and developmental toxicity assay development. *Reproductive Toxicology* **56**, 52–55 (2015).

[67] Pogrmic-Majkic, K., Samardzija Nenadov, D., Tesic, B., Fa Nedeljkovic, S., Kokai, D., Stanic, B. & Andric, N. Mapping DEHP to the adverse outcome pathway network for human female reproductive toxicity. *Archives of Toxicology* **96**, 2799–2813 (2022).

[68] Hogberg, H. T. *et al.* The Adverse Outcome Pathway Framework Applied to Neurological Symptoms of COVID-19. *Cells* **11**, 3411 (2022).

[69] Mustieles, V. *et al.* Bisphenol A and its analogues: A comprehensive review to identify and prioritize effect biomarkers for human biomonitoring. *Environment International* **144**, 105811 (2020).

[70] Spinu, N., Bal-Price, A., Cronin, M. T. D., Enoch, S. J., Madden, J. C. & Worth, A. P. Development and analysis of an adverse outcome pathway network for human neurotoxicity. *Archives of Toxicology* **93**, 2759–2772 (2019).

[71] Tsamou, M. & Roggen, E. L. Building a Network of Adverse Outcome Pathways (AOPs) Incorporating the Tau-Driven AOP Toward Memory Loss (AOP429). *Journal of Alzheimer's Disease Reports* **6**, 271–296 (2022).

[72] Ankley, G. T. *et al.* Adverse Outcome Pathway Network–Based Assessment of the Interactive Effects of an Androgen Receptor Agonist and an Aromatase Inhibitor on Fish Endocrine Function. *Environmental Toxicology and Chemistry* **39**, 913–922 (2020).

[73] Gölz, L., Baumann, L., Pannetier, P., Braunbeck, T., Knapen, D. & Vergauwen, L. AOP Report: Thyroperoxidase Inhibition Leading to Altered Visual Function in Fish Via Altered Retinal Layer Structure. *Environmental Toxicology and Chemistry* **41**, 2632–2648 (2022).

[74] Haigis, A.-C., Vergauwen, L., LaLone, C. A., Villeneuve, D. L., O'Brien, J. M. & Knapen, D. Cross-species applicability of an adverse outcome pathway network for thyroid hormone system disruption. *Toxicological Sciences* **195**, 1–27 (2023).

[75] Holbech, H. *et al.* ERGO: Breaking Down the Wall between Human Health and Environmental Testing of Endocrine Disrupters. *International Journal of Molecular Sciences* **21**, 2954 (2020).

[76] Pípal, M., Wiklund, L., Caccia, S. & Beronius, A. Assessment of endocrine disruptive properties of PFOS: EFSA/ECHA guidance case study utilising AOP networks and alternative methods. *EFSA Journal* **20**, e200418 (2022).

[77] Ravichandran, J., Karthikeyan, B. S. & Samal, A. Investigation of a derived adverse outcome pathway (AOP) network for endocrine-mediated perturbations. *Science of the Total Environment* **826**, 154112 (2022).

[78] Ankley, G. T., Santana-Rodriguez, K., Jensen, K. M., Miller, D. H. & Villeneuve, D. L. AOP Report: Adverse Outcome Pathways for Aromatase Inhibition or Androgen Receptor Agonism Leading to Male-Biased Sex Ratio and Population Decline in Fish. *Environmental Toxicology and Chemistry* **42**, 747–756 (2023).

[79] Pistollato, F., de Gyves, E. M., Carpi, D., Bopp, S. K., Nunes, C., Worth, A. & Bal-Price, A. Assessment of developmental neurotoxicity induced by chemical mixtures using an adverse outcome pathway concept. *Environmental Health* **19**, 23 (2020).

[80] Arnesdotter, E., Spinu, N., Firman, J., Ebbrell, D., Cronin, M. T., Vanhaecke, T. & Vinken, M. Derivation, characterisation and analysis of an adverse outcome pathway network for human hepatotoxicity. *Toxicology* **459**, 152856 (2021).

[81] Escher, S. E. *et al.* Integrate mechanistic evidence from new approach methodologies (NAMs) into a read-across assessment to characterise trends in shared mode of action. *Toxicology in Vitro* **79**, 105269 (2022).

[82] Halappanavar, S. *et al.* Adverse outcome pathways as a tool for the design of testing strategies to support the safety assessment of emerging advanced materials at the nanoscale. *Particle and Fibre Toxicology* **17**, 16 (2020).

[83] Jeong, J., Kim, D. & Choi, J. Integrative Data Mining Approach: Case Study with Adverse Outcome Pathway Network Leading to Pulmonary Fibrosis. *Chemical Research in Toxicology* **36**, 838–847 (2023).

[84] Luettich, K., Sharma, M., Yepiskoposyan, H., Breheny, D. & Lowe, F. J. An Adverse Outcome Pathway for Decreased Lung Function Focusing on Mechanisms of Impaired Mucociliary Clearance Following Inhalation Exposure. *Frontiers in Toxicology* **3**, 750254 (2021).

[85] Cho, E. *et al.* AOP report: Development of an adverse outcome pathway for oxidative DNA damage leading to mutations and chromosomal aberrations. *Environmental and Molecular Mutagenesis* **63**, 118–134 (2022).

[86] Del'haye, G. G. *et al.* Development of an adverse outcome pathway network for breast cancer: a comprehensive representation of the pathogenesis, complexity and diversity of the disease. *Archives of Toxicology* **96**, 2881–2897 (2022).

[87] Edwards, S. W., Tan, Y.-M., Villeneuve, D. L., Meek, M. E. & McQueen, C. A. Adverse Outcome Pathways—Organizing Toxicological Information to Improve Decision Making. *Journal of Pharmacology and Experimental Therapeutics* **356**, 170–181 (2016).

[88] EFSA Panel on Plant Protection Products and their Residues (PPR) *et al.* Development of adverse outcome pathways relevant for the identification of substances having endocrine disruption properties Uterine adenocarcinoma as adverse outcome. *EFSA Journal* **21**, 7744 (2023).

[89] LaLone, C. A., Villeneuve, D. L., Wu-Smart, J., Milsk, R. Y., Sappington, K., Garber, K. V., Housenger, J. & Ankley, G. T. Weight of evidence evaluation of a network of adverse outcome pathways linking activation of the nicotinic acetylcholine receptor in honey bees to colony death. *Science of the Total Environment* **584-585**, 751–775 (2017).

[90] Villeneuve, D. L. *et al.* Representing the Process of Inflammation as Key Events in Adverse Outcome Pathways. *Toxicological Sciences* **163**, 346–352 (2018).

[91] Wiklund, L., Caccia, S., Pípal, M., Nymark, P. & Beronius, A. Development of a data-driven approach to Adverse Outcome Pathway network generation: a case study on the EATS-modalities. *Frontiers in Toxicology* **5**, 1183824 (2023).

[92] Wu, Q., Bagdad, Y., Taboureau, O. & Audouze, K. Capturing a Comprehensive Picture of Biological Events From Adverse Outcome Pathways in the Drug Exposome. *Frontiers in Public Health* **9**, 763962 (2021).

[93] Aguayo-Orozco, A., Audouze, K., Siggaard, T., Barouki, R., Brunak, S. & Taboureau, O. sAOP: linking chemical stressors to adverse outcomes pathway networks. *Bioinformatics* **35**, 5391–5392 (2019).

[94] Wang, S., Li, C., Zhang, L., Chen, Q. & Wang, S. Assessing the ecological impacts of polycyclic aromatic hydrocarbons petroleum pollutants using a network toxicity model. *Environmental Research* **245**, 117901 (2024).

[95] Sahoo, A. K., Chivukula, N., Ramesh, K., Singha, J., Marigoudar, S. R., Sharma, K. V. & Samal, A. An integrative data-centric approach to derivation and charac-

terization of an adverse outcome pathway network for cadmium-induced toxicity. *Science of The Total Environment* **920**, 170968 (2024).

[96] Sahoo, A. K., Chivukula, N., Madgaonkar, S. R., Ramesh, K., Marigoudar, S. R., Sharma, K. V. & Samal, A. Leveraging integrative toxicogenomic approach towards development of stressor-centric adverse outcome pathway networks for plastic additives. *Archives of Toxicology* **98**, 3299–3321 (2024).

[97] Gustafson, J. B., Tell, J. G. & Orem, D. *Selection of Representative TPH Fractions Based on Fate and Transport Considerations*, vol. 3 of *Total Petroleum Hydrocarbon Criteria Working Group Series* (Amherst Scientific Publishers, Amherst, Massachusetts, USA, 1997).

[98] Potter, T. L. & Simmons, K. E. *Composition of Petroleum Mixtures*, vol. 2 of *Total Petroleum Hydrocarbon Criteria Working Group Series* (Amherst Scientific Publishers, Amherst, Massachusetts, USA, 1998).

[99] Posthuma, L., Suter II, G. W. & Traas, T. P. (eds.) *Species Sensitivity Distributions in Ecotoxicology* (CRC Press, Boca Raton, 2001).

[100] Sahoo, A. K., Madgaonkar, S. R., Chivukula, N., Karthikeyan, P., Ramesh, K., Marigoudar, S. R., Sharma, K. V. & Samal, A. Network-based investigation of petroleum hydrocarbons-induced ecotoxicological effects and their risk assessment. *bioRxiv* 2024.07.18.604159 (2024).

[101] Davey, R. A. & Grossmann, M. Androgen Receptor Structure, Function and Biology: From Bench to Bedside. *The Clinical Biochemist. Reviews* **37**, 3–15 (2016).

[102] Tan, M. E., Li, J., Xu, H. E., Melcher, K. & Yong, E.-l. Androgen receptor: structure, role in prostate cancer and drug discovery. *Acta Pharmacologica Sinica* **36**, 3–23 (2015).

[103] Jeng, H. A. Exposure to Endocrine Disrupting Chemicals and Male Reproductive Health. *Frontiers in Public Health* **2**, 55 (2014).

[104] Rehman, S., Usman, Z., Rehman, S., AlDraihem, M., Rehman, N., Rehman, I. & Ahmad, G. Endocrine disrupting chemicals and impact on male reproductive health. *Translational Andrology and Urology* **7**, 490–503 (2018).

[105] Rodprasert, W., Toppari, J. & Virtanen, H. E. Endocrine Disrupting Chemicals and Reproductive Health in Boys and Men. *Frontiers in Endocrinology* **12**, 706532 (2021).

[106] UNEP & WHO. State of the science of endocrine disrupting chemicals 2012. https://www.who.int/publications/i/item/9789241505031 (2013).

[107] Fang, H. *et al.* Study of 202 Natural, Synthetic, and Environmental Chemicals for Binding to the Androgen Receptor. *Chemical Research in Toxicology* **16**, 1338–1358 (2003).

[108] ChemIDplus. https://pubchem.ncbi.nlm.nih.gov/source/ChemIDplus.

[109] Sud, M. MayaChemTools: An Open Source Package for Computational Drug Discovery. *Journal of Chemical Information and Modeling* **56**, 2292–2297 (2016).

[110] Djoumbou Feunang, Y. *et al.* ClassyFire: automated chemical classification with a comprehensive, computable taxonomy. *Journal of Cheminformatics* **8**, 61 (2016).

[111] Morgan, H. L. The Generation of a Unique Machine Description for Chemical Structures-A Technique Developed at Chemical Abstracts Service. *Journal of Chemical Documentation* **5**, 107–113 (1965).

[112] Rogers, D. & Hahn, M. Extended-Connectivity Fingerprints. *Journal of Chemical Information and Modeling* **50**, 742–754 (2010).

[113] González-Medina, M., Prieto-Martínez, F. D., Owen, J. R. & Medina-Franco, J. L. Consensus Diversity Plots: a global diversity analysis of chemical libraries. *Journal of Cheminformatics* **8**, 63 (2016).

[114] Bemis, G. W. & Murcko, M. A. The Properties of Known Drugs. 1. Molecular Frameworks. *Journal of Medicinal Chemistry* **39**, 2887–2893 (1996).

[115] Tanimoto, T. *An Elementary Mathematical Theory of Classification and Prediction* (International Business Machines Corporation, 1958).

[116] Jolliffe, I. T. *Principal Component Analysis*. Springer Series in Statistics (Springer New York, New York, NY, 1986).

[117] Bastian, M., Heymann, S. & Jacomy, M. Gephi: An Open Source Software for Exploring and Manipulating Networks. *Proceedings of the International AAAI Conference on Web and Social Media* **3**, 361–362 (2009).

[118] Blondel, V. D., Guillaume, J.-L., Lambiotte, R. & Lefebvre, E. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment* **2008**, P10008 (2008).

[119] Lipkus, A. H., Yuan, Q., Lucas, K. A., Funk, S. A., Bartelt, W. F., Schenck, R. J. & Trippe, A. J. Structural Diversity of Organic Chemistry. A Scaffold Analysis of the CAS Registry. *The Journal of Organic Chemistry* **73**, 4443–4451 (2008).

[120] Vivek-Ananth, R., Sahoo, A. K., Baskaran, S. P. & Samal, A. Scaffold and structural diversity of the secondary metabolite space of medicinal fungi. *ACS Omega* **8**, 3102–3113 (2023).

[121] González-Medina, M., Owen, J. R., El-Elimat, T., Pearce, C. J., Oberlies, N. H., Figueroa, M. & Medina-Franco, J. L. Scaffold Diversity of Fungal Metabolites. *Frontiers in Pharmacology* **8**, 180 (2017).

[122] Hunter, J. D. Matplotlib: A 2D Graphics Environment. *Computing in Science & Engineering* **9**, 90–95 (2007).

[123] RDKit: Open-Source Cheminformatics Software. https://www.rdkit.org/.

[124] California Office of Environmental Health Hazard Assessment. The Proposition 65 List. https://oehha.ca.gov/proposition-65/proposition-65-list.

[125] European Chemicals Agency. Substances restricted under REACH. https://echa.europa.eu/substances-restricted-under-reach.

[126] European Chemicals Agency. Candidate List of Substances of Very High Concern (SVHC) for Authorisation. https://echa.europa.eu/candidate-list-table.

[127] Toxic chemicals restricted to be imported or exported in China. https://www.cirs-reach.com/China_Chemical_Regulation/Registration_of_import_export_of_toxic_chemicals_in_China.html.

[128] European Chemicals Agency. EU list of sub- stances prohibited in cosmetic products. https://echa.europa.eu/cosmetics-prohibited-substances.

[129] Schedule 1 hazardous chemicals list in India. http://moef.gov.in/wp-content/uploads/2019/08/SCHEDULE-I.html.

[130] RCSB Protein Data Bank. https://www.rcsb.org/.

[131] Bassett, J. D. & Williams, G. R. Critical role of the hypothalamic–pituitary–thyroid axis in bone. *Bone* **43**, 418–426 (2008).

[132] Feldt-Rasmussen, U., Effraimidis, G. & Klose, M. The hypothalamus-pituitary-thyroid (HPT)-axis and its role in physiology and pathophysiology of other hypothalamus-pituitary functions. *Molecular and Cellular Endocrinology* **525**, 111173 (2021).

[133] Ortiga-Carvalho, T. M., Chiamolera, M. I., Pazos-Moura, C. C. & Wondisford, F. E. Hypothalamus-Pituitary-Thyroid Axis. *Comprehensive Physiology* **6**, 1387–1428 (2016).

[134] Fekete, C. & Lechan, R. M. Central Regulation of Hypothalamic-Pituitary-Thyroid Axis Under Physiological and Pathophysiological Conditions. *Endocrine Reviews* **35**, 159–194 (2014).

[135] Schmutzler Cornelia *et al.* Endocrine Disruptors and the Thyroid Gland—A Combined in Vitro and in Vivo Analysis of Potential New Biomarkers. *Environmental Health Perspectives* **115**, 77–83 (2007).

[136] Thambirajah, A. A., Wade, M. G., Verreault, J., Buisine, N., Alves, V. A., Langlois, V. S. & Helbing, C. C. Disruption by stealth - Interference of endocrine disrupting chemicals on hormonal crosstalk with thyroid axis function in humans and other animals. *Environmental Research* **203**, 111906 (2022).

[137] Schug, T. T. *et al.* Designing endocrine disruption out of the next generation of chemicals. *Green Chemistry* **15**, 181–198 (2013).

[138] Chen, X. *et al.* Tralopyril induces developmental toxicity in zebrafish embryo (Danio rerio) by disrupting the thyroid system and metabolism. *Science of the Total Environment* **746**, 141860 (2020).

[139] Lee, S., Lee, J.-S., Kho, Y. & Ji, K. Effects of methylisothiazolinone and octylisothiazolinone on development and thyroid endocrine system in zebrafish larvae. *Journal of Hazardous Materials* **425**, 127994 (2022).

[140] Teng, M., Zhu, W., Wang, D., Yan, J., Qi, S., Song, M. & Wang, C. Acute exposure of zebrafish embryo (Danio rerio) to flutolanil reveals its developmental mechanism of toxicity via disrupting the thyroid system and metabolism. *Environmental Pollution* **242**, 1157–1165 (2018).

[141] Sanders, J., Evans, M., Premawardhana, L., Depraetere, H., Jeffreys, J., Richards, T., Furmaniak, J. & Smith, B. R. Human monoclonal thyroid stimulating autoantibody. *The Lancet* **362**, 126–128 (2003).

[142] Evans, M. *et al.* Monoclonal autoantibodies to the TSH receptor, one with stimulating activity and one with blocking activity, obtained from the same blood sample. *Clinical Endocrinology* **73**, 404–412 (2010).

[143] Garcia de Lomana, M., Weber, A. G., Birk, B., Landsiedel, R., Achenbach, J., Schleifer, K.-J., Mathea, M. & Kirchmair, J. In Silico Models to Predict the Perturbation of Molecular Initiating Events Related to Thyroid Hormone Homeostasis. *Chemical Research in Toxicology* **34**, 396–411 (2021).

[144] Kurosaki, K., Wu, R. & Uesawa, Y. A Toxicity Prediction Tool for Potential Agonist/Antagonist Activities in Molecular Initiating Events Based on Chemical Structures. *International Journal of Molecular Sciences* **21**, 7853 (2020).

[145] PubChem. https://pubchem.ncbi.nlm.nih.gov/.

[146] OECD Existing Chemicals Database. https://hpvchemicals.oecd.org/ui/Search.aspx.

[147] US EPA: High Production Volume List. https://comptox.epa.gov/dashboard/chemical-lists/EPAHPV.

[148] Dionisio, K. L., Phillips, K., Price, P. S., Grulke, C. M., Williams, A., Biryol, D., Hong, T. & Isaacs, K. K. The Chemical and Products Database, a resource for exposure-relevant data on chemicals in consumer products. *Scientific Data* **5**, 180125 (2018).

[149] Pérez-Villanueva, J., Santos, R., Hernández-Campos, A., Giulianotti, M. A., Castillo, R. & Medina-Franco, J. L. Structure–activity relationships of benzim-

idazole derivatives as antiparasitic agents: Dual activity-difference (DAD) maps. *MedChemComm* **2**, 44–49 (2011).

[150] Medina-Franco, J. L. Scanning Structure–Activity Relationships with Structure–Activity Similarity and Related Maps: From Consensus Activity Cliffs to Selectivity Switches. *Journal of Chemical Information and Modeling* **52**, 2485–2493 (2012).

[151] Richard, A. M. *et al.* The Tox21 10K Compound Library: Collaborative Chemistry Advancing Toxicology. *Chemical Research in Toxicology* **34**, 189–216 (2021).

[152] Richard, A. M. *et al.* ToxCast Chemical Landscape: Paving the Road to 21st Century Toxicology. *Chemical Research in Toxicology* **29**, 1225–1251 (2016).

[153] Jeong, J., Kim, D. & Choi, J. Application of ToxCast/Tox21 data for toxicity mechanism-based evaluation and prioritization of environmental chemicals: Perspective and limitations. *Toxicology in Vitro* **84**, 105451 (2022).

[154] Balali-Mood, M., Naseri, K., Tahergorabi, Z., Khazdair, M. R. & Sadeghi, M. Toxic Mechanisms of Five Heavy Metals: Mercury, Lead, Chromium, Cadmium, and Arsenic. *Frontiers in Pharmacology* **12**, 643972 (2021).

[155] Rehman, K., Fatima, F., Waheed, I. & Akash, M. S. H. Prevalence of exposure of heavy metals and their impact on health consequences. *Journal of Cellular Biochemistry* **119**, 157–184 (2018).

[156] Genchi, G., Sinicropi, M. S., Lauria, G., Carocci, A. & Catalano, A. The Effects of Cadmium Toxicity. *International Journal of Environmental Research and Public Health* **17**, 3782 (2020).

[157] Kubier, A., Wilkin, R. T. & Pichler, T. Cadmium in soils and groundwater: A review. *Applied Geochemistry* **108**, 104388 (2019).

[158] Tamele, I. J. & Vázquez Loureiro, P. Lead, Mercury and Cadmium in Fish and Shellfish from the Indian Ocean and Red Sea (African Countries): Public Health Challenges. *Journal of Marine Science and Engineering* **8**, 344 (2020).

[159] Åkesson, A. *et al.* Tubular and Glomerular Kidney Effects in Swedish Women with Low Environmental Cadmium Exposure. *Environmental Health Perspectives* **113**, 1627–1631 (2005).

[160] ATSDR. *Toxicological profile for cadmium* (Public Health Service, U.S. Department of Health and Human Services, Atlanta, GA, USA, 2012).

[161] Kumar, S. & Sharma, A. Cadmium toxicity: effects on human reproduction and fertility. **34**, 327–338 (2019).

[162] Prozialeck, W. C., Edwards, J. R. & Woods, J. M. The vascular endothelium as a target of cadmium toxicity. *Life Sciences* **79**, 1493–1506 (2006).

[163] Wang, B. & Du, Y. Cadmium and Its Neurotoxic Effects. *Oxidative Medicine and Cellular Longevity* **2013**, 898034 (2013).

[164] Chen, Y. Y. & Chan, K. M. Modulations of TCDD-mediated induction of zebrafish cyp1a1 and the AHR pathway by administering Cd2+ in vivo. *Chemosphere* **210**, 577–587 (2018).

[165] Guével, R. L., Petit, F., Goff, P. L., Métivier, R., Valotaire, Y. & Pakdel, F. Inhibition of Rainbow Trout (Oncorhynchus mykiss) Estrogen Receptor Activity by Cadmium1. *Biology of Reproduction* **63**, 259–266 (2000).

[166] Liu, Y., Chen, Q., Li, Y., Bi, L., Jin, L. & Peng, R. Toxic Effects of Cadmium on Fish. *Toxics* **10**, 622 (2022).

[167] Nesatyy, V. J., Ammann, A. A., Rutishauser, B. V. & Suter, M. J.-F. Effect of Cadmium on the Interaction of 17β-Estradiol with the Rainbow Trout Estrogen Receptor. *Environmental Science & Technology* **40**, 1358–1363 (2006).

[168] Tilton, S. C., Foran, C. M. & Benson, W. H. Effects of cadmium on the reproductive axis of Japanese medaka (Oryzias latipes). *Comparative Biochemistry and Physiology Part C: Toxicology & Pharmacology* **136**, 265–276 (2003).

[169] US EPA. Priority Pollutant List. https://www.epa.gov/sites/default/files/2015-09/documents/priority-pollutant-list-epa.pdf.

[170] IARC Working Group on the Evaluation of Carcinogenic Risks to Humans. Cadmium and cadmium compounds. In *Beryllium, Cadmium, Mercury, and Exposures in the Glass Manufacturing Industry* (International Agency for Research on Cancer, 1993).

[171] Extended Advisory Group on Molecular Screening and Toxicogenomics. AOP Developers' Handbook. https://aopwiki.org/handbooks/4.

[172] OECD. Users' Handbook Supplement to the Guidance Document for Developing and Assessing AOPs, Series on Testing and Assessment No. 233, OECD Publishing, Paris (2022).

[173] Hagberg, A. A., Schult, D. A. & Swart, P. J. Exploring Network Structure, Dynamics, and Function using NetworkX. In Varoquaux, G., Vaught, T. & Millman, J. (eds.) *Proceedings of the 7th Python in Science Conference (SciPy 2008)*, 11–15 (2008).

[174] Davis, A. P., Wiegers, T. C., Grondin, C. J., Johnson, R. J., Sciaky, D., Wiegers, J. & Mattingly, C. J. Leveraging the Comparative Toxicogenomics Database to Fill in Knowledge Gaps for Environmental Health: A Test Case for Air Pollution-induced Cardiovascular Disease. *Toxicological Sciences* **177**, 392–404 (2020).

[175] NCBI Gene. https://www.ncbi.nlm.nih.gov/gene.

[176] The Gene Ontology Resource. Guide to GO evidence codes. https://geneontology.org/docs/guide-go-evidence-codes/.

[177] Fröhlich, H., Speer, N., Poustka, A. & Beißbarth, T. GOSim – an R-package for computation of information theoretic GO similarities between terms and gene products. *BMC Bioinformatics* **8**, 166 (2007).

[178] US EPA. Exploring ToxCast Data. https://www.epa.gov/chemical-research/exploring-toxcast-data.

[179] Feshuk, M., Kolaczkowski, L., Dunham, K., Davidson-Fritz, S. E., Carstens, K. E., Brown, J., Judson, R. S. & Paul Friedman, K. The ToxCast pipeline: updates to curve-fitting approaches and database structure. *Frontiers in Toxicology* **5**, 1275980 (2023).

[180] Shannon, P. *et al.* Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Research* **13**, 2498–2504 (2003).

[181] Takes, F. W. & Kosters, W. A. Determining the Diameter of Small World Networks. In *Proceedings of the 20th ACM International Conference on Information and Knowledge Management*, CIKM '11, 1191–1196 (Association for Computing Machinery, New York, NY, USA, 2011).

[182] Benoit, L. *et al.* Adverse outcome pathway from activation of the AhR to breast cancer-related death. *Environment International* **165**, 107323 (2022).

[183] Jaylet, T., Coustillet, T., Jornod, F., Margaritte-Jeannin, P. & Audouze, K. AOP-helpFinder 2.0: Integration of an event-event searches module. *Environment International* **177**, 108017 (2023).

[184] Jornod, F., Jaylet, T., Blaha, L., Sarigiannis, D., Tamisier, L. & Audouze, K. AOP-helpFinder webserver: a tool for comprehensive analysis of the literature to support adverse outcome pathways development. *Bioinformatics* **38**, 1173–1175 (2022).

[185] Baker, N., Knudsen, T. & Williams, A. Abstract Sifter: a comprehensive front-end system to PubMed. *F1000Research* **6**, 2164 (2017).

[186] PubMed. https://pubmed.ncbi.nlm.nih.gov/.

[187] Dimitriadis, E. *et al.* Pre-eclampsia. *Nature Reviews Disease Primers* **9**, 8 (2023).

[188] Liu, T., Zhang, M., Guallar, E., Wang, G., Hong, X., Wang, X. & Mueller, N. T. Trace Minerals, Heavy Metals, and Preeclampsia: Findings from the Boston Birth Cohort. *Journal of the American Heart Association* **8**, e012436 (2019).

[189] Smarr, M. M., Mirzaei Salehabadi, S., Boyd Barr, D., Buck Louis, G. M. & Sundaram, R. A multi-pollutant assessment of preconception persistent endocrine disrupting chemicals and incident pregnancy loss. *Environment International* **157**, 106788 (2021).

[190] Kluxen, F. M., Diel, P., Höfer, N., Becker, E. & Degen, G. H. The metallohormone cadmium modulates AhR-associated gene expression in the small intestine of rats similar to ethinyl-estradiol. *Archives of Toxicology* **87**, 633–643 (2013).

[191] Larigot, L., Juricek, L., Dairou, J. & Coumoul, X. AhR signaling pathways and regulatory functions. *Biochimie Open* **7**, 1–9 (2018).

[192] Chun, Y.-S., Choi, E., Kim, G.-T., Choi, H., Kim, C.-H., Lee, M.-J., Kim, M.-S. & Park, J.-W. Cadmium blocks hypoxia-inducible factor (HIF)-1-mediated response to hypoxia by stimulating the proteasome-dependent degradation of HIF-1. *European Journal of Biochemistry* **267**, 4198–4204 (2000).

[193] Genbacev, O., Zhou, Y., Ludlow, J. W. & Fisher, S. J. Regulation of Human Placental Development by Oxygen Tension. *Science* **277**, 1669–1672 (1997).

[194] Iyer, N. V. *et al.* Cellular and developmental control of O2 homeostasis by hypoxia-inducible factor 1. *Genes & development* **12**, 149–162 (1998).

[195] Xiong, Y.-W. *et al.* Environmental exposure to cadmium impairs fetal growth and placental angiogenesis via GCN-2-mediated mitochondrial stress. *Journal of Hazardous Materials* **401**, 123438 (2021).

[196] Kim, J. *et al.* The effects of cadmium on VEGF-mediated angiogenesis in HUVECs. *Journal of Applied Toxicology* **32**, 342–349 (2012).

[197] Kozlosky, D., Lu, A., Doherty, C., Buckley, B., Goedken, M. J., Miller, R. K., Barrett, E. S. & Aleksunes, L. M. Cadmium reduces growth of male fetuses by impairing development of the placental vasculature and reducing expression of nutrient transporters. *Toxicology and Applied Pharmacology* **475**, 116636 (2023).

[198] Guo, M.-Y., Wang, H., Chen, Y.-H., Xia, M.-Z., Zhang, C. & Xu, D.-X. N-acetylcysteine alleviates cadmium-induced placental endoplasmic reticulum stress and fetal growth restriction in mice. *PLOS ONE* **13**, e0191667 (2018).

[199] Xiong, Y.-W. *et al.* Maternal cadmium exposure during late pregnancy causes fetal growth restriction via inhibiting placental progesterone synthesis. *Ecotoxicology and Environmental Safety* **187**, 109879 (2020).

[200] Laine, J. E., Ray, P., Bodnar, W., Cable, P. H., Boggess, K., Offenbacher, S. & Fry, R. C. Placental Cadmium Levels Are Associated with Increased Preeclampsia Risk. *PLOS ONE* **10**, e0139341 (2015).

[201] Zhang, Q. *et al.* Cadmium-induced immune abnormality is a key pathogenic event in human and rat models of preeclampsia. *Environmental Pollution* **218**, 770–782 (2016).

[202] Li, X. *et al.* Maternal cadmium exposure impairs placental angiogenesis in preeclampsia through disturbing thyroid hormone receptor signaling. *Ecotoxicology and Environmental Safety* **244**, 114055 (2022).

[203] Amiard-Triquet, C. Chapter 1 - Introduction. In Amiard-Triquet, C., Amiard, J.-C. & Mouneyrac, C. (eds.) *Aquatic Ecotoxicology*, 1–23 (Academic Press, 2015).

[204] Blechinger, S. R., Warren, J. T., Kuwada, J. Y. & Krone, P. H. Developmental toxicology of cadmium in living embryos of a stable transgenic zebrafish line. *Environmental Health Perspectives* **110**, 1041–1046 (2002).

[205] Shankar, P., Dasgupta, S., Hahn, M. E. & Tanguay, R. L. A Review of the Functional Roles of the Zebrafish Aryl Hydrocarbon Receptors. *Toxicological Sciences* **178**, 215–238 (2020).

[206] Liang, H. *et al.* Cadmium exposure induces endothelial dysfunction via disturbing lipid metabolism in human microvascular endothelial cells. *Journal of Applied Toxicology* **41**, 775–788 (2021).

[207] Messner, B. *et al.* Cadmium Is a Novel and Independent Risk Factor for Early Atherosclerosis Mechanisms and In Vivo Relevance. *Arteriosclerosis, Thrombosis, and Vascular Biology* **29**, 1392–1398 (2009).

[208] Tang, L., Su, J. & Liang, P. Modeling cadmium-induced endothelial toxicity using human pluripotent stem cell-derived endothelial cells. *Scientific Reports* **7**, 14811 (2017).

[209] Omidi, M., Niknahad, H., Noorafshan, A., Fardid, R., Nadimi, E., Naderi, S., Bakhtari, A. & Mohammadi-Bardbori, A. Co-exposure to an Aryl Hydrocarbon Receptor Endogenous Ligand, 6-Formylindolo[3,2-b]carbazole (FICZ), and Cadmium Induces Cardiovascular Developmental Abnormalities in Mice. *Biological Trace Element Research* **187**, 442–451 (2019).

[210] Wu, X., Chen, Y., Luz, A., Hu, G. & Tokar, E. J. Cardiac Development in the Presence of Cadmium: An *in Vitro* Study Using Human Embryonic Stem Cells and Cardiac Organoids. *Environmental Health Perspectives* **130**, 117002 (2022).

[211] Witeska, M., Sarnowski, P., Ługowska, K. & Kowal, E. The effects of cadmium and copper on embryonic and larval development of ide Leuciscus idus L. *Fish Physiology and Biochemistry* **40**, 151–163 (2014).

[212] Conolly, R. B., Ankley, G. T., Cheng, W., Mayo, M. L., Miller, D. H., Perkins, E. J., Villeneuve, D. L. & Watanabe, K. H. Quantitative Adverse Outcome Pathways and Their Application to Predictive Toxicology. *Environmental Science & Technology* **51**, 4661–4672 (2017).

[213] Perkins, E. J. *et al.* Building and Applying Quantitative Adverse Outcome Pathway Models for Chemical Hazard and Risk Assessment. *Environmental Toxicology and Chemistry* **38**, 1850–1865 (2019).

[214] Thompson, R. C., Moore, C. J., vom Saal, F. S. & Swan, S. H. Plastics, the environment and human health: current consensus and future trends. *Philosophical Transactions of the Royal Society B: Biological Sciences* **364**, 2153–2166 (2009).

[215] Parthasarathy, A., Tyler, A. C., Hoffman, M. J., Savka, M. A. & Hudson, A. O. Is Plastic Pollution in Aquatic and Terrestrial Environments a Driver for the Transmission of Pathogens and the Evolution of Antibiotic Resistance? *Environmental Science & Technology* **53**, 1744–1745 (2019).

[216] Li, P., Wang, X., Su, M., Zou, X., Duan, L. & Zhang, H. Characteristics of Plastic Pollution in the Environment: A Review. *Bulletin of Environmental Contamination and Toxicology* **107**, 577–584 (2021).

[217] Hecker, M., Peijnenburg, W., Lam, P. K. S. & Brinkmann, M. Emerging issues in aquatic toxicology – Plastic pollution. *Aquatic Toxicology* **264**, 106729 (2023).

[218] Hahladakis, J. N., Velis, C. A., Weber, R., Iacovidou, E. & Purnell, P. An overview of chemical additives present in plastics: Migration, release, fate and environmental impact during their use, disposal and recycling. *Journal of Hazardous Materials* **344**, 179–199 (2018).

[219] Maes, T. *et al.* A recipe for plastic: Expert insights on plastic additives in the marine environment. *Marine Pollution Bulletin* **196**, 115633 (2023).

[220] Hermabessiere, L., Dehaut, A., Paul-Pont, I., Lacroix, C., Jezequel, R., Soudant, P. & Duflos, G. Occurrence and effects of plastic additives on marine environments and organisms: A review. *Chemosphere* **182**, 781–793 (2017).

[221] Oehlmann, J., Oetken, M. & Schulte-Oehlmann, U. A critical evaluation of the environmental risk assessment for plasticizers in the freshwater environment in Europe, with special emphasis on bisphenol A and endocrine disruption. *Environmental Research* **108**, 140–149 (2008).

[222] Sendra, M., Pereiro, P., Figueras, A. & Novoa, B. An integrative toxicogenomic analysis of plastic additives. *Journal of Hazardous Materials* **409**, 124975 (2021).

[223] Stevens, S., McPartland, M., Bartosova, Z., Skåland, H. S., Völker, J. & Wagner, M. Plastic Food Packaging from Five Countries Contains Endocrine- and Metabolism-Disrupting Chemicals. *Environmental Science & Technology* **58**, 4859–4871 (2024).

[224] Herrera, A. *et al.* Bioaccumulation of additives and chemical contaminants from environmental microplastics in European seabass (Dicentrarchus labrax). *Science of the Total Environment* **822**, 153396 (2022).

[225] Costa, J. P. d., Avellan, A., Mouneyrac, C., Duarte, A. & Rocha-Santos, T. Plastic additives and microplastics as emerging contaminants: Mechanisms and analytical assessment. *TrAC Trends in Analytical Chemistry* **158**, 116898 (2023).

[226] Aurisano, N., Weber, R. & Fantke, P. Enabling a circular economy for chemicals in plastics. *Current Opinion in Green and Sustainable Chemistry* **31**, 100513 (2021).

[227] Wiesinger, H., Wang, Z. & Hellweg, S. Deep Dive into Plastic Monomers, Additives, and Processing Aids. *Environmental Science & Technology* **55**, 9339–9351 (2021).

[228] CAS Common Chemistry. https://commonchemistry.cas.org/.

[229] Koleske, J. V., Springate, R. & Brezinski, D. Additives handbook. **6**, 1–64 (2011).

[230] Pritchard, G. (ed.) *Plastics Additives An A-Z reference* (Springer Dordrecht, 2012).

[231] Al-Malaika, S., Axtell, F., Rothon, R. & Gilbert, M. Chapter 7 - Additives for Plastics. In Gilbert, M. (ed.) *Brydson's Plastics Materials (Eighth Edition)*, 127–168 (Butterworth-Heinemann, 2017).

[232] Coleman, E. A. 21 - Plastics Additives. In Kutz, M. (ed.) *Applied Plastics Engineering Handbook (Second Edition)*, 489–500 (William Andrew Publishing, 2017).

[233] BASF. Additives for Adhesives and Sealants. https://www.basf.com/global/documents/en/products-and-industries/architectural-coatings/20130501_Additives_for_Adhesives_and_Sealants_Catalogue_Europe.pdf.assetinline.pdf.

[234] De Frond, H., Rubinovitz, R. & Rochman, C. M. ATR-FTIR Spectral Libraries of Plastic Particles (FLOPP and FLOPP-e) for the Analysis of Microplastics. *Analytical Chemistry* **93**, 15878–15885 (2021).

[235] Jeong, J. & Choi, J. Development of AOP relevant to microplastics based on toxicity mechanisms of chemical additives using ToxCast™ and deep learning models combined approach. *Environment International* **137**, 105557 (2020).

[236] Judson, R. *et al.* Analysis of the Effects of Cell Stress and Cytotoxicity on In Vitro Assay Activity Across a Diverse Chemical and Assay Space. *Toxicological Sciences* **152**, 323–339 (2016).

[237] Saarimäki, L. A., Fratello, M., Pavel, A., Korpilähde, S., Leppänen, J., Serra, A. & Greco, D. A curated gene and biological system annotation of adverse outcome pathways related to human health. *Scientific Data* **10**, 409 (2023).

[238] Williams, A. The Chemical and Products Database (CPDat) MySQL Data File. https://epa.figshare.com/articles/dataset/The_Chemical_and_Products_Database_CPDat_MySQL_Data_File/5352997.

[239] Williams, A. J. *et al.* The CompTox Chemistry Dashboard: a community data resource for environmental chemistry. *Journal of Cheminformatics* **9**, 61 (2017).

[240] Williams, A. J., Lambert, J. C., Thayer, K. & Dorne, J.-L. C. Sourcing data on chemical properties and hazard data from the US-EPA CompTox Chemicals Dashboard: A practical guide for human risk assessment. *Environment International* **154**, 106566 (2021).

[241] Qian, S. *et al.* Detection and quantification analysis of chemical migrants in plastic food contact products. *PLOS ONE* **13**, e0208467 (2018).

[242] Groh, K. J., Geueke, B., Martin, O., Maffini, M. & Muncke, J. Overview of intentionally used food contact chemicals and their hazards. *Environment International* **150**, 106225 (2021).

[243] Groh, K. J. *et al.* Overview of known plastic packaging-associated chemicals and their hazards. *Science of the Total Environment* **651**, 3253–3268 (2019).

[244] Lewis, K. A., Tzilivakis, J., Warner, D. J. & Green, A. An international database for pesticide risk assessments and management. *Human and Ecological Risk Assessment: An International Journal* **22**, 1050–1064 (2016).

[245] Menger, F., Boström, G., Jonsson, O., Ahrens, L., Wiberg, K., Kreuger, J. & Gago-Ferrero, P. Identification of Pesticide Transformation Products in Surface Water Using Suspect Screening Combined with National Monitoring Data. *Environmental Science & Technology* **55**, 10343–10353 (2021).

[246] Aurisano, N., Huang, L., Milà i Canals, L., Jolliet, O. & Fantke, P. Chemicals of concern in plastic toys. *Environment International* **146**, 106194 (2021).

[247] Ravichandran, J., Karthikeyan, B. S., Jost, J. & Samal, A. An atlas of fragrance chemicals in children's products. *Science of the Total Environment* **818**, 151682 (2022).

[248] IARC. IARC Monographs on the Identification of Carcinogenic Hazards to Humans. https://monographs.iarc.who.int/list-of-classifications/.

[249] Ravichandran, J., Karthikeyan, B. S., Aparna, S. & Samal, A. Network biology approach to human tissue-specific chemical exposome. *The Journal of Steroid Biochemistry and Molecular Biology* **214**, 105998 (2021).

[250] Neveu, V., Nicolas, G., Salek, R. M., Wishart, D. S. & Scalbert, A. Exposome-Explorer 2.0: an update incorporating candidate dietary biomarkers and dietary associations with cancer risk. *Nucleic Acids Research* **48**, D908–D912 (2020).

[251] Schriml, L. M. *et al.* The Human Disease Ontology 2022 update. *Nucleic Acids Research* **50**, D1255–D1261 (2022).

[252] Rugard, M., Coumoul, X., Carvaillo, J.-C., Barouki, R. & Audouze, K. Deciphering Adverse Outcome Pathway Network Linked to Bisphenol F Using Text Mining and Systems Toxicology Approaches. *Toxicological Sciences* **173**, 32–40 (2020).

[253] Yang, W. *et al.* Benzo[a]pyrene inhibits testosterone biosynthesis via NDUFA10-mediated mitochondrial compromise in mouse Leydig cells: Integrating experimental and in silico toxicological approaches. *Ecotoxicology and Environmental Safety* **244**, 114075 (2022).

[254] Lassen, P., Hoffmann, L. & Thomsen, M. PAHs in toys and childcare products. https://www2.mst.dk/udgiv/publications/2012/01/978-87-92779-49-6.pdf (2011).

[255] Alassali, A., Calmano, W., Gidarakos, E. & Kuchta, K. The degree and source of plastic recyclates contamination with polycyclic aromatic hydrocarbons. *RSC Advances* **10**, 44989–44996 (2020).

[256] Bataller, R. & Brenner, D. A. Liver fibrosis. *The Journal of Clinical Investigation* **115**, 209–218 (2005).

[257] Cheung, A. C., Walker, D. I., Juran, B. D., Miller, G. W. & Lazaridis, K. N. Studying the Exposome to Understand the Environmental Determinants of Complex Liver Diseases. *Hepatology* **71**, 352–362 (2020).

[258] Barouki, R., Samson, M., Blanc, E. B., Colombo, M., Zucman-Rossi, J., Lazaridis, K. N., Miller, G. W. & Coumoul, X. The exposome and liver disease - how environmental factors affect liver health. *Journal of Hepatology* **79**, 492–505 (2023).

[259] Tsai, C.-H., Li, C.-H., Liao, P.-L., Cheng, Y.-W., Lin, C.-H., Huang, S.-H. & Kang, J.-J. NcoA2-Dependent Inhibition of HIF-1 Activation Is Regulated via AhR. *Toxicological Sciences* **148**, 517–530 (2015).

[260] Lou, W. *et al.* Molecular mechanism of benzo [a] pyrene regulating lipid metabolism via aryl hydrocarbon receptor. *Lipids in Health and Disease* **21**, 13 (2022).

[261] Almendarez-Reyna, C. I., de la Trinidad Chacón, C. G., Ochoa-Martínez, C., Rico-Guerrero, L. A. & Pérez-Maldonado, I. N. The aryl hydrocarbon receptor (AhR) activation mediates benzo(a)pyrene-induced overexpression of AQP3 and Notch1 in HaCaT cells. *Environmental and Molecular Mutagenesis* **64**, 466–472 (2023).

[262] Malik, D.-e.-s., David, R. M. & Gooderham, N. J. Mechanistic evidence that benzo[a]pyrene promotes an inflammatory microenvironment that drives the metastatic potential of human mammary cells. *Archives of Toxicology* **92**, 3223–3239 (2018).

[263] Zheng, Z., Park, J. K., Kwon, O. W., Ahn, S. H., Kwon, Y. J., Jiang, L., Zhu, S. & Park, B. H. The Risk of Gastrointestinal Cancer on Daily Intake of Low-Dose BaP in C57BL/6 for 60 Days. *The Korean Academy of Medical Sciences* **37**, e235 (2022).

[264] Gorria, M. *et al.* A new lactoferrin- and iron-dependent lysosomal death pathway is induced by benzo[a]pyrene in hepatic epithelial cells. *Toxicology and Applied Pharmacology* **228**, 212–224 (2008).

[265] Li, J., Bai, J., Si, X., Jia, H. & Wu, Z. Benzo[a]pyrene induces epithelial tight junction disruption and apoptosis via inhibiting the initiation of autophagy in intestinal porcine epithelial cells. *Chemico-Biological Interactions* **374**, 110386 (2023).

[266] Yan, L., Messner, C. J., Zhang, X. & Suter-Dick, L. Assessment of fibrotic pathways induced by environmental chemicals using 3D-human liver microtissue model. *Environmental Research* **194**, 110679 (2021).

[267] Hill, C. & Wang, Y. Autophagy in pulmonary fibrosis: friend or foe? *Genes & Diseases* **9**, 1594–1607 (2022).

[268] Bukowska, B., Mokra, K. & Michałowicz, J. Benzo[a]pyrene—Environmental Occurrence, Human Exposure, and Mechanisms of Toxicity. *International Journal of Molecular Sciences* **23**, 6348 (2022).

[269] Sathikumaran, R., Madhuvandhi, J., Priya, K., Sridevi, A., Krishnamurthy, R. & Thilagam, H. Evaluation of benzo[a]pyrene-induced toxicity in the estuarine thornfish Therapon jarbua. *Toxicology Reports* **9**, 720–727 (2022).

[270] Nacci, D. E., Kohan, M., Pelletier, M. & George, E. Effects of benzo[a]pyrene exposure on a fish population resistant to the toxic effects of dioxin-like compounds. *Aquatic Toxicology* **57**, 203–215 (2002).

[271] Seemann, F., Peterson, D. R., Witten, P. E., Guo, B.-S., Shanthanagouda, A. H., Ye, R. R., Zhang, G. & Au, D. W. Insight into the transgenerational effect of benzo[a]pyrene on bone formation in a teleost fish (Oryzias latipes). *Comparative Biochemistry and Physiology Part C: Toxicology & Pharmacology* **178**, 60–67 (2015).

[272] Bugiak, B. & Weber, L. P. Hepatic and vascular mRNA expression in adult zebrafish (Danio rerio) following exposure to benzo-a-pyrene and 2,3,7,8-tetrachlorodibenzo-p-dioxin. *Zebrafish Issue* **95**, 299–306 (2009).

[273] Wang, H., Pan, L., Zhang, X., Ji, R., Si, L. & Cao, Y. The molecular mechanism of AhR-ARNT-XREs signaling pathway in the detoxification response induced by polycyclic aromatic hydrocarbons (PAHs) in clam Ruditapes philippinarum. *Environmental Research* **183**, 109165 (2020).

[274] Garcia, G. R., Shankar, P., Dunham, C. L., Garcia, A., Du, J. K. L., Truong, L., Tilton, S. C. & Tanguay, R. L. Signaling Events Downstream of AHR Activation That Contribute to Toxic Responses: The Functional Role of an AHR-Dependent Long Noncoding RNA (*slincR*) Using the Zebrafish Model. *Environmental Health Perspectives* **126**, 117002 (2018).

[275] Dalcq, J., Pasque, V., Ghaye, A., Larbuisson, A., Motte, P., Martial, J. A. & Muller, M. RUNX3, EGR1 and SOX9B Form a Regulatory Cascade Required to Modulate BMP-Signaling during Cranial Cartilage Development in Zebrafish. *PLOS ONE* **7**, e50140 (2012).

[276] He, C., Zuo, Z., Shi, X., Li, R., Chen, D., Huang, X., Chen, Y. & Wang, C. Effects of benzo(a)pyrene on the skeletal development of Sebastiscus marmoratus embryos and the molecular mechanism involved. *Aquatic Toxicology* **101**, 335–341 (2011).

[277] Hague, S. M., Klaffke, S. & Bandmann, O. Neurodegenerative disorders: Parkinson's disease and Huntington's disease. *Journal of Neurology, Neurosurgery & Psychiatry* **76**, 1058–1063 (2005).

[278] Przedborski, S. Neurodegeneration. In Gendelman, H. E. & Ikezu, T. (eds.) *Neuroimmune Pharmacology*, 229–237 (Springer US, Boston, MA, 2008).

[279] Hanioka, N., Jinno, H., Tanaka-Kagawa, T., Nishimura, T. & Ando, M. Interaction of bisphenol A with rat hepatic cytochrome P450 enzymes. *Chemosphere* **41**, 973–978 (2000).

[280] Shi, R., Liu, Z. & Liu, T. The antagonistic effect of bisphenol A and nonylphenol on liver and kidney injury in rats. *Immunopharmacology and Immunotoxicology* **43**, 527–535 (2021).

[281] Costa, H. E. & Cairrao, E. Effect of bisphenol A on the neurological system: a review update. *Archives of Toxicology* **98**, 1–73 (2024).

[282] Valencia-Olvera, A. C., Morán, J., Camacho-Carranza, R., Prospéro-García, O. & Espinosa-Aguirre, J. J. CYP2E1 induction leads to oxidative stress and cytotoxicity in glutathione-depleted cerebellar granule neurons. *Toxicology in Vitro* **28**, 1206–1214 (2014).

[283] DeGracia, D. J., Kumar, R., Owen, C. R., Krause, G. S. & White, B. C. Molecular Pathways of Protein Synthesis Inhibition during Brain Reperfusion: Implications for Neuronal Survival or Death. *Journal of Cerebral Blood Flow & Metabolism* **22**, 127–140 (2002).

[284] Coimbra-Costa, D., Alva, N., Duran, M., Carbonell, T. & Rama, R. Oxidative stress and apoptosis after acute respiratory hypoxia and reoxygenation in rat brain. *Redox Biology* **12**, 216–225 (2017).

[285] Bosch-Panadero, E. *et al.* Bisphenol A is an exogenous toxin that promotes mitochondrial injury and death in tubular cells. *Environmental Toxicology* **33**, 325–332 (2018).

[286] Xia, T., Guo, J., Zhang, B., Song, C., Zhao, Q., Cui, B. & Liu, Y. Bisphenol A Promotes the Progression of Colon Cancer Through Dual-Targeting of NADPH Oxidase and Mitochondrial Electron-Transport Chain to Produce ROS and Activating HIF-1/VEGF/PI3K/AKT Axis. *Frontiers in Endocrinology* **13**, 933051 (2022).

[287] Biswas, S., Ghosh, S., Samanta, A., Das, S., Mukherjee, U. & Maitra, S. Bisphenol A impairs reproductive fitness in zebrafish ovary: Potential involvement of oxidative/nitrosative stress, inflammatory and apoptotic mediators. *Environmental Pollution* **267**, 115692 (2020).

[288] Zhao, Z.-b., Ji, K., Shen, X.-y., Zhang, W.-w., Wang, R., Xu, W.-p. & Wei, W. Di(2-ethylhexyl) phthalate promotes hepatic fibrosis by regulation of oxidative stress and inflammation responses in rats. *Environmental Toxicology and Pharmacology* **68**, 109–119 (2019).

[289] Li, G., Zhao, C.-Y., Wu, Q., Guan, S.-y., Jin, H.-W., Na, X.-L. & Zhang, Y.-B. Integrated metabolomics and transcriptomics reveal di(2-ethylhexyl) phthalate-induced mitochondrial dysfunction and glucose metabolism disorder through oxidative stress in rat liver. *Ecotoxicology and Environmental Safety* **228**, 112988 (2021).

[290] Amara, I., Timoumi, R., Annabi, E., Di Rosa, G., Scuto, M., Najjar, M. F., Calabrese, V. & Abid-Essefi, S. Di (2-ethylhexyl) phthalate targets the thioredoxin system and the oxidative branch of the pentose phosphate pathway in liver of Balb/c mice. *Environmental Toxicology* **35**, 78–86 (2020).

[291] Maloney, E. K. & Waxman, D. J. trans-Activation of PPAR and PPAR by Structurally Diverse Environmental Chemicals. *Toxicology and Applied Pharmacology* **161**, 209–218 (1999).

[292] Lapinskas, P. J., Brown, S., Leesnitzer, L. M., Blanchard, S., Swanson, C., Cattley, R. C. & Corton, J. C. Role of PPAR in mediating the effects of phthalates and metabolites in the liver. *Toxicology* **207**, 149–163 (2005).

[293] Sant, K. E. *et al.* Embryonic exposures to mono-2-ethylhexyl phthalate induce larval steatosis in zebrafish independent of Nrf2a signaling. *Journal of Developmental Origins of Health and Disease* **12**, 132–140 (2021).

[294] Golshan, M. *et al.* Di-(2-ethylhexyl)-phthalate disrupts pituitary and testicular hormonal functions to reduce sperm quality in mature goldfish. *Aquatic Toxicology* **163**, 16–26 (2015).

[295] Zhang, Q., Ye, D., Wang, H., Wang, Y., Hu, W. & Sun, Y. Zebrafish cyp11c1 Knockout Reveals the Roles of 11-ketotestosterone and Cortisol in Sexual Development and Reproduction. *Endocrinology* **161**, bqaa048 (2020).

[296] Sorci, G. & Loiseau, C. Should we worry about the accumulation of microplastics in human organs? *eBioMedicine* **82**, 104191 (2022).

[297] Tornero, V. & Hanke, G. Chemical contaminants entering the marine environment from sea-based sources: A review with a focus on European seas. *Marine Pollution Bulletin* **112**, 17–38 (2016).

[298] Zheng, G. J. & Richardson, B. J. Petroleum hydrocarbons and polycyclic aromatic hydrocarbons (PAHs) in Hong Kong marine sediments. *Chemosphere* **38**, 2625–2632 (1999).

[299] Alford, J. B., Peterson, M. S. & Green, C. C. (eds.) *Impacts of Oil Spill Disasters on Marine Habitats and Fisheries in North America* (CRC Press, Boca Raton, 2014), 1st edn.

[300] Al-Hawash, A. B., Dragh, M. A., Li, S., Alhujaily, A., Abbood, H. A., Zhang, X. & Ma, F. Principles of microbial degradation of petroleum hydrocarbons in the environment. *Egyptian Journal of Aquatic Research* **44**, 71–76 (2018).

[301] Keith, L. H. The Source of U.S. EPA's Sixteen PAH Priority Pollutants. *Polycyclic Aromatic Compounds* **35**, 147–160 (2015).

[302] 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 Pollution* **223**, 5535–5548 (2012).

[303] Lawal, A. T. Polycyclic aromatic hydrocarbons. A review. *Cogent Environmental Science* **3**, 1339841 (2017).

[304] Barron, M. G., Vivian, D. N., Heintz, R. A. & Yim, U. H. Long-Term Ecological Impacts from Oil Spills: Comparison of Exxon Valdez, Hebei Spirit, and Deepwater Horizon. *Environmental Science & Technology* **54**, 6456–6467 (2020).

[305] Pasparakis, C., Esbaugh, A. J., Burggren, W. & Grosell, M. Physiological impacts of Deepwater Horizon oil on fish. *Comparative Biochemistry and Physiology Part C: Toxicology & Pharmacology* **224**, 108558 (2019).

[306] Begum, M. *et al.* Assessment of Ennore Oil Spill 2017 on Chennai Coastal Water and Biota. In *OCEANS 2022 - Chennai*, 1–7 (IEEE, 2022).

[307] Sammarco, P. W., Kolian, S. R., Warby, R. A., Bouldin, J. L., Subra, W. A. & Porter, S. A. Distribution and concentrations of petroleum hydrocarbons associated with the BP/Deepwater Horizon Oil Spill, Gulf of Mexico. *Marine Pollution Bulletin* **73**, 129–143 (2013).

[308] Sivagami, K., Jaa Vignesh, V., Tamizhdurai, P., Rajasekhar, B., Sakthipriya, N. & Nambi, I. M. Studies on short term weathering of spilled oil along Chennai coast in South India. *Journal of Cleaner Production* **230**, 1410–1420 (2019).

[309] Weisman, W. *Analysis of Petroleum Hydrocarbons in Environmental Media*, vol. 1 of *Total Petroleum Hydrocarbon Criteria Working Group Series* (Amherst Scientific Publishers, Amherst, Massachusetts, USA, 1998).

[310] Jagiello, K., Judzinska, B., Sosnowska, A., Lynch, I., Halappanavar, S. & Puzyn, T. Using AOP-Wiki to support the ecotoxicological risk assessment of nanomaterials:

first steps in the development of novel adverse outcome pathways. *Environmental Science: Nano* **9**, 1675–1684 (2022).

[311] US EPA. Water-related environmental fate of 129 priority pollutants: Volume II: Halogenated aliphatic hydrocarbons, halogenated ethers, monocyclic aromatics, phthalate esters, polycyclic aromatic hydrocarbons, nitrosamines, and miscellaneous compounds. http://nepis.epa.gov/Exe/ZyPURL.cgi?Dockey=2000K6JL.txt (1979).

[312] Kamo, M. Species Sensitivity Distribution in Ecological Risk Assessment. In Kamo, M. (ed.) *Theories in Ecological Risk Assessment*, 103–134 (Springer Nature Singapore, Singapore, 2023).

[313] Kooijman, S. A. L. M. A safety factor for LC50 values allowing for differences in sensitivity among species. *Water Research* **21**, 269–276 (1987).

[314] van Straalen, N. M. & Denneman, C. A. Ecotoxicological evaluation of soil quality criteria. *Ecotoxicology and Environmental Safety* **18**, 241–251 (1989).

[315] Dowse, R., Tang, D., Palmer, C. G. & Kefford, B. J. Risk assessment using the species sensitivity distribution method: Data quality versus data quantity. *Environmental Toxicology and Chemistry* **32**, 1360–1369 (2013).

[316] US EPA. Guidelines for Ecological Risk Assessment. https://www.epa.gov/risk/guidelines-ecological-risk-assessment (1998).

[317] Etterson, M. Technical Manual: SSD Toolbox Version 1.0. https://gaftp.epa.gov/comptox/Sustainable_Chemistry_Data/SSD_Toolbox/V1.0/SSDToolbox.TechnicalManual.March.2020_Tagged01.pdf (2020).

[318] Schwarz, C. & Tillmanns, A. Improving statistical methods to derive species sensitivity distribution. https://a100.gov.bc.ca/pub/acat/public/viewReport.do?reportId=57400 (2019).

[319] Wheeler, J., Grist, E., Leung, K., Morritt, D. & Crane, M. Species sensitivity distributions: data and model choice. *Marine Pollution Bulletin* **45**, 192–202 (2002).

[320] US EPA. Methods for Measuring the Acute Toxicity of Effluents and Receiving Waters to Freshwater and Marine Organisms. https://www.epa.gov/sites/default/files/2015-08/documents/acute-freshwater-and-marine-wet-manual_2002.pdf (2002).

[321] Belanger, S. *et al.* Future needs and recommendations in the development of species sensitivity distributions: Estimating toxicity thresholds for aquatic ecological communities and assessing impacts of chemical exposures. *Integrated Environmental Assessment and Management* **13**, 664–674 (2017).

[322] US EPA. Species Sensitivity Distribution (SSD) Toolbox. https://www.epa.gov/comptox-tools/species-sensitivity-distribution-ssd-toolbox.

[323] Eom, I., Rast, C., Veber, A. & Vasseur, P. Ecotoxicity of a polycyclic aromatic hydrocarbon (PAH)-contaminated soil. *Ecotoxicology and Environmental Safety* **67**, 190–205 (2007).

[324] Sopian, N. A., Jalaludin, J., Abu Bakar, S., Hamedon, T. R. & Latif, M. T. Exposure to Particulate PAHs on Potential Genotoxicity and Cancer Risk among School Children Living Near the Petrochemical Industry. *International Journal of Environmental Research and Public Health* **18**, 2575 (2021).

[325] Yan, J., Wang, L., Fu, P. P. & Yu, H. Photomutagenicity of 16 polycyclic aromatic hydrocarbons from the US EPA priority pollutant list. *Mutation Research/Genetic Toxicology and Environmental Mutagenesis* **557**, 99–108 (2004).

[326] Zelinkova, Z. & Wenzl, T. The Occurrence of 16 EPA PAHs in Food – A Review. *Polycyclic Aromatic Compounds* **35**, 248–284 (2015).

[327] Zhang, Y. & Tao, S. Global atmospheric emission inventory of polycyclic aromatic hydrocarbons (PAHs) for 2004. *Atmospheric Environment* **43**, 812–819 (2009).

[328] Zhuo, S. *et al.* Source-oriented risk assessment of inhalation exposure to ambient polycyclic aromatic hydrocarbons and contributions of non-priority isomers in urban Nanjing, a megacity located in Yangtze River Delta, China. *Environmental Pollution* **224**, 796–809 (2017).

[329] Logeshwaran, P., Megharaj, M., Chadalavada, S., Bowman, M. & Naidu, R. Petroleum hydrocarbons (PH) in groundwater aquifers: An overview of environmental fate, toxicity, microbial degradation and risk-based remediation approaches. *Environmental Technology & Innovation* **10**, 175–193 (2018).

[330] Honda, M. & Suzuki, N. Toxicities of Polycyclic Aromatic Hydrocarbons for Aquatic Animals. *International Journal of Environmental Research and Public Health* **17**, 1363 (2020).

[331] Booc, F., Thornton, C., Lister, A., MacLatchy, D. & Willett, K. L. Benzo[a]pyrene Effects on Reproductive Endpoints in Fundulus heteroclitus. *Toxicological Sciences* **140**, 73–82 (2014).

[332] Pandelides, Z., Sturgis, M., Thornton, C., Aluru, N. & Willett, K. Benzo[a]pyrene-induced multigenerational changes in gene expression, behavior, and DNA methylation are primarily influenced by paternal exposure. *Toxicology and Applied Pharmacology* **469**, 116545 (2023).

[333] Wan, T., Mo, J., Au, D. W.-T., Qin, X., Tam, N. Y.-K., Kong, R. Y.-C. & Seemann, F. The role of DNA methylation on gene expression in the vertebrae of ancestrally benzo[a]pyrene exposed F1 and F3 male medaka. *Epigenetics* **18**, 2222246 (2023).

[334] Wojciechowski, M. F. & Meehan, T. Inhibition of DNA methyltransferases in vitro by benzo[a]pyrene diol epoxide-modified substrates. *Journal of Biological Chemistry* **259**, 9711–9716 (1984).

[335] Yauk, C. L., Polyzos, A., Rowan-Carroll, A., Kortubash, I., Williams, A. & Kovalchuk, O. Tandem repeat mutation, global DNA methylation, and regulation of DNA methyltransferases in cultured mouse embryonic fibroblast cells chronically exposed to chemicals with different modes of action. *Environmental and Molecular Mutagenesis* **49**, 26–35 (2008).

[336] Corrales, J., Fang, X., Thornton, C., Mei, W., Barbazuk, W., Duke, M., Scheffler, B. & Willett, K. Effects on specific promoter DNA methylation in zebrafish embryos and larvae following benzo[a]pyrene exposure. *Comparative Biochemistry and Physiology Part C: Toxicology & Pharmacology* **163**, 37–46 (2014).

[337] Lin, S., Ren, A., Wang, L., Huang, Y., Wang, Y., Wang, C. & Greene, N. D. Oxidative Stress and Apoptosis in Benzo[a]pyrene-Induced Neural Tube Defects. *Free Radical Biology and Medicine* **116**, 149–158 (2018).

[338] Malott, K. F., Leon Parada, K., Lee, M., Swanson, E. & Luderer, U. Gestational Benzo[a]pyrene Exposure Destroys F1 Ovarian Germ Cells Through Mitochondrial Apoptosis Pathway and Diminishes Surviving Oocyte Quality. *Toxicological Sciences* **190**, 23–40 (2022).

[339] Sui, L. *et al.* Maternal benzo[a]pyrene exposure is correlated with the meiotic arrest and quality deterioration of offspring oocytes in mice. *Reproductive Toxicology* **93**, 10–18 (2020).

[340] Arnot, J. A. & Gobas, F. A. A review of bioconcentration factor (BCF) and bioaccumulation factor (BAF) assessments for organic chemicals in aquatic organisms. *Environmental Reviews* **14**, 257–297 (2006).

[341] Karthikeyan, P., Marigoudar, S. R., Mohan, D., Sharma, K. V. & Ramana Murthy, M. Prescribing sea water quality criteria for arsenic, cadmium and lead through species sensitivity distribution. *Ecotoxicology and Environmental Safety* **208**, 111612 (2021).

[342] Méndez-Fernández, L., Casado-Martínez, C., Martínez-Madrid, M., Moreno-Ocio, I., Costas, N., Pardo, I. & Rodriguez, P. Derivation of sediment Hg quality standards based on ecological assessment in river basins. *Environmental Pollution* **245**, 1000–1013 (2019).

[343] Cavanaugh, J. E. & Neath, A. A. The Akaike information criterion: Background, derivation, properties, application, interpretation, and refinements. *WIREs Computational Statistics* **11**, e1460 (2019).

[344] Fox, D. *et al.* Recent Developments in Species Sensitivity Distribution Modeling. *Environmental Toxicology and Chemistry* **40**, 293–308 (2021).

[345] Yanagihara, M., Hiki, K. & Iwasaki, Y. Can Chemical Toxicity in Saltwater Be Predicted from Toxicity in Freshwater? A Comprehensive Evaluation Using Species Sensitivity Distributions. *Environmental Toxicology and Chemistry* **41**, 2021–2027 (2022).

[346] Chen, J., Fan, B., Li, J., Wang, X., Li, W., Cui, L. & Liu, Z. Development of human health ambient water quality criteria of 12 polycyclic aromatic hydrocarbons (PAH) and risk assessment in China. *Chemosphere* **252**, 126590 (2020).

[347] Medina-Franco, J. L., Martínez-Mayorga, K., Bender, A., Marín, R. M., Giulianotti, M. A., Pinilla, C. & Houghten, R. A. Characterization of Activity Landscapes Using 2D and 3D Similarity Methods: Consensus Activity Cliffs. *Journal of Chemical Information and Modeling* **49**, 477–491 (2009).

[348] Bierkens, J. & Geerts, L. Environmental hazard and risk characterisation of petroleum substances: A guided "walking tour" of petroleum hydrocarbons. *Environment International* **66**, 182–193 (2014).

[349] Machin, D., Bryant, T., Altman, D. & Gardner, M. (eds.) *Statistics with Confidence: Confidence Intervals and Statistical Guidelines* (BMJ Books, 2000).

[350] Medina-Franco, J. L., Yongye, A. B., Pérez-Villanueva, J., Houghten, R. A. & Martínez-Mayorga, K. Multitarget Structure–Activity Relationships Characterized by Activity-Difference Maps and Consensus Similarity Measure. *Journal of Chemical Information and Modeling* **51**, 2427–2439 (2011).

[351] Miranda-Quintana, R. A., Bajusz, D., Rácz, A. & Héberger, K. Extended similarity indices: the benefits of comparing more than two objects simultaneously. Part 1: Theory and characteristics†. *Journal of Cheminformatics* **13**, 32 (2021).

[352] Miranda-Quintana, R. A., Rácz, A., Bajusz, D. & Héberger, K. Extended similarity indices: the benefits of comparing more than two objects simultaneously. Part 2: speed, consistency, diversity selection. *Journal of Cheminformatics* **13**, 33 (2021).

[353] Dablander, M., Hanser, T., Lambiotte, R. & Morris, G. M. Exploring QSAR models for activity-cliff prediction. *Journal of Cheminformatics* **15**, 47 (2023).

[354] van Tilborg, D., Alenicheva, A. & Grisoni, F. Exposing the Limitations of Molecular Machine Learning with Activity Cliffs. *Journal of Chemical Information and Modeling* **62**, 5938–5951 (2022).

[355] Wagner, M. *et al.* State of the science on plastic chemicals - Identifying and addressing chemicals and polymers of concern. https://zenodo.org/records/10701706 (2024).

[356] Shi, W., Zhang, R. & Tan, H. AOP-Based Machine Learning for Toxicity Prediction. In Hong, H. (ed.) *Machine Learning and Deep Learning in Computational Toxicology*, 141–157 (Springer International Publishing, Cham, 2023).

[357] Clerbaux, L.-A. *et al.* Beyond chemicals: Opportunities and challenges of integrating non-chemical stressors in adverse outcome pathways. *ALTEX - Alternatives to animal experimentation* **41**, 233–247 (2023).

[358] Gao, P. The Exposome in the Era of One Health. *Environmental Science & Technology* **55**, 2790–2799 (2021).