

## Research Article

# Computational Systems Biology as an Animal-Free Approach to Characterize Toxicological Effects of Persistent Organic Pollutants

Qier Wu, Rayane Achebouché and Karine Audouze

Université de Paris, Inserm UMR S-1124, Paris, France

### Abstract

Exposure to persistent organic pollutants (POPs), as defined by the Stockholm Convention, may alter biological systems and cause toxic effects. Computational studies appear to be a relevant approach to increase our understanding of the molecular mechanisms triggered by POPs. We investigated the use of a systems toxicology approach to explore the effects of POPs on human health. A protein-protein association network (PPAN) was developed based on known POP-protein interactions. This model was used to predict protein complexes for several candidate POPs, including dicofol, methoxychlor, and perfluorooctanoic acid (PFOA), that are listed or proposed to be listed as POPs by the Stockholm Convention. Integration of multiple data sources (pathways, disease annotations, adverse outcome pathways) involving the identified protein complexes was performed independently in order to reveal putative risk factors for human health. This approach revealed that several systems may be disturbed by these candidate POPs, mainly the reproductive, metabolic and nervous systems.

This study highlights that a computational systems toxicology approach may help to decipher putative biological mechanisms of poorly studied chemicals and link them to possible adverse effects with the aim to support regulatory assessment and trigger new epidemiological and experimental studies. In order to develop more accurate computational models as alternative methods to animal testing, the next challenge will be to integrate more data according to the findable, accessible, interoperable and reusable (FAIR) data principles.

## 1 Introduction

Persistent organic pollutants (POPs) are organic chemicals that persist for long time periods in the environment and that are resistant to environmental degradation by chemical, biological and photolytic processes. Therefore, they are capable of long-range transport, bioaccumulation in human and animal tissues (e.g., adipose tissue), and bio-magnification in food chains. POPs include dioxin, polychlorinated biphenyls (PCBs), polybrominated diphenyl ethers (PBDEs), and hexabromocyclodecane (HBCD).

Exposure to such chemicals may have significant impacts on human health, animals as well as on the environment. Previous studies (from epidemiology to *in vitro* studies) have shown that some POPs may contribute to health disorders such as certain

cancers, birth defects, dysfunctional immunity, reproductive and neurological disorders (Koual et al., 2019; Deng et al., 2019). With the aim of protecting human health and the environment, an international agreement, the Stockholm Convention on POPs<sup>1</sup>, was established and adopted in 2001, and is effective since May 2004. The main objectives of this convention are to eliminate or restrict the production and use of chemical substances defined as POPs and to reduce their unintentional release. In 2007, 16 substances were added to the 12 initial POPs covered by this convention<sup>2</sup>.

Besides taking measures to regulate the production and use of these chemicals, it is essential to gain a better understanding of their biological modes of action (MoA) and pathways of toxicity (PoT) to understand their impact on human health (Kleensang et

<sup>1</sup> <http://www.pops.int/TheConvention/Overview/TextoftheConvention/tabid/2232/Default.aspx>

<sup>2</sup> <http://chm.pops.int/tabid/3486/Default.aspx>

Received October 16, 2019; Accepted January 14, 2020;  
Epub January 21, 2020; © The Authors, 2020.

ALTEX 37(2), 287-299. doi:10.14573/altex.1910161

Correspondence: Karine Audouze, PhD  
Université de Paris, Inserm UMR-S 1124  
45 rue des Saints-Pères, 75006 Paris, France  
(karine.audouze@univ-paris-diderot.fr)

This is an Open Access article distributed under the terms of the Creative Commons Attribution 4.0 International license (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution and reproduction in any medium, provided the original work is appropriately cited.



al., 2014) and facilitate decision-making that supports regulation, innovation and competitiveness (Naderi et al., 2014). Further, the development of relevant, reliable and cost-effective methods can help to identify new substances with similar activities.

New approach methodologies (NAMs) are animal-free approaches that include computational models, *in vitro* high throughput screening (HTS), omics studies (Hartung et al., 2017), and the newly suggested 3S approach (systematic, systemic, and systems biology and toxicology) (Smirnova et al., 2018). NAMs promise to capture a better understanding of the MoA of chemicals in humans and may help to support chemical safety assessment (Andersen et al., 2019). While they may not be able to completely substitute experimental studies, they can help to prioritize further experiments and thus support the 3Rs principle (Replacement, Reduction, Refinement).

During recent years, toxicological and chemical databases have expanded substantially, and the relevance of integrative systems toxicological methods has been recognized (Corton et al., 2019; Audouze and Grandjean, 2011; Audouze et al., 2013). The development of computational approaches for future toxicity assessment has been recommended by several committees including the U.S. EPA (Knudsen, 2013), the National Research Council (NRC, 2007; Krewski et al., 2010), the European Chemicals Agency (ECHA), the REACH legislation (Registration, Evaluation and Authorisation of Chemicals), and the Organization for Economic Co-operation and Development (OECD).

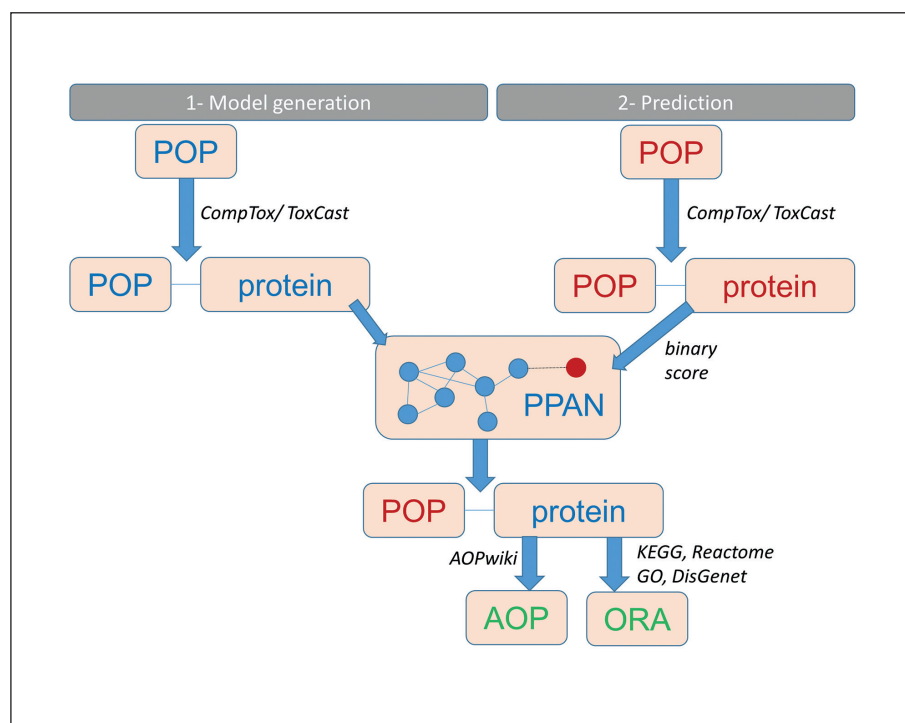
Many small molecules, including drugs and environmental chemicals, are known to affect multiple molecular targets and may deregulate proteins and, subsequently, functional pathways (Yildirim et al., 2007; Audouze et al., 2010). Here, we propose a computational model specific to the “biological space” of POPs

to predict yet uncharacterized links between POPs and biological targets that may cause toxic effects. Since, in biological systems, proteins do not work as individual units but more in functional modules made up of two or more interacting proteins, we applied a computational network biology approach developed previously that associates protein pairs based on chemical-protein information (PPAN) (Audouze et al., 2010) to generate a computational systems toxicology model for persistent organic pollutants (POP-PPAN, POP-protein-protein association network).

This first global model, created using large sets of toxicogenomics data to explore the full biological space, was benchmarked against high confidence experimental protein-protein interactions, and predictions were validated by literature searches. This concept has previously been applied to other areas, with successful experimental validation of some chemical-protein predictions (Taboureau and Audouze, 2017; Audouze et al., 2014). The ability of the developed POP-PPAN model to predict new links by calculation of a binary score is demonstrated with several case studies, including dicofol and perfluorooctanoic acid (PFOA), compounds listed as POP candidates in May 2019 under the Stockholm Convention on POPs.

## 2 Material and methods

A workflow of the strategy to develop the computational systems toxicology model is shown in Figure 1. First, existing biological knowledge was extracted from the ToxCast and CompTox databases (Richard et al., 2016; Williams et al., 2017). Specific information on POP-protein interactions was compiled in order to create the network-based model POP-PPAN. Only pro-



**Fig. 1: Workflow of the computational systems toxicology approach to predict adverse effects of suspected persistent organic pollutants (POPs)**

**1-Model generation:** Links between POPs from the Stockholm Convention list and proteins were extracted from the CompTox and ToxCast databases (in blue). The POP-PPAN model was created, based on a protein-protein association network procedure. **2-Prediction:** Known proteins linked to suspected POPs (in red) were compiled. The developed model was then screened for these proteins; binary scores were calculated between known (red) and uncharacterized (blue) proteins for the suspected POP, using a neighbor protein procedure. Biological interpretation of the results was done after integration and over representation analysis (ORA) of various datasets (adverse outcome pathways (AOPs), pathways and diseases).

**Tab. 1: Chemicals currently listed as POPs under the Stockholm convention**

For the compound classes (designated with an \*), specific chemicals that are good representatives of these chemical families were selected, resulting in a total of 35 compounds.

Initial POPs	CAS rn		Newer POPs	CAS rn	
Aldrin	309-00-2	Insecticide	Alpha Hexachlorocyclohexane	319-84-6	Insecticide
Chlordane	57-74-9	Pesticide	Beta Hexachlorocyclohexane	319-85-7	Insecticide
Dichlorodiphenyltrichloroethane (DDT)	50-29-3	Pesticide	Chlordecone (=Kepone)	143-50-0	Insecticide
Dieldrin	60-57-1	Insecticide	Decabromodiphenyl ether (DecaBDE = BDE 209)	1163-19-5	Flame retardant
Endrin	72-20-8	Pesticide	Hexabromobiphenyl (PBB 153)	36355-01-8	Flame retardant
Heptachlor	76-44-8	Pesticide	Hexabromocyclododecane	25637-99-4	Flame retardant
Hexachlorobenzene (HCB)	118-74-1	Fungicide, additive to fireworks, ammunition & synthetic rubber	Hexabromodiphenyl ether	36483-60-0	
Mirex	2385-85-5	Pesticide, fire retardant	Heptabromodiphenyl ether	68928-80-3	
Toxaphene	8001-35-2	Pesticide	Hexachlorobutadiene	87-68-3	
Polychlorinated biohenyls (PCB)*		Insulating fluid; paint; additive; lubricant	Lindane	58-89-9	Insecticide
– 2,2'-Dichlorobiphenyl	13029-08-8		Pentachlorobenzene	608-93-5	Pesticide
– 2,2',3,4',5,5',6-Heptachloro-1,1'-biphenyl	52663-68-0		Pentachlorophenol	87-86-5	
– 2,2',4,4',5,5'-Hexachlorobiphenyl	35065-27-1		Perfluorooctane sulfonic acid (PFOS)	1763-23-1	Protective coating
Polychlorinated dibenzo-p-dioxins (PCDD)*		Herbicide	Perfluorooctane sulfonyl fluoride (PFOSF)	307-35-7	Protective coating
– 2,3,7,8-Tetrachlorodibenzodioxin (=TCDD)	1746-01-6		Pentabromodiphenyl ether	32534-81-9	Flame retardant
Polychlorinated dibenzofurans (PCDF)*		Herbicide	Tetrabromodiphenyl ether	40088-47-9	
– 2,3,4,7,8-Pentachlorodibenzofuran	57117-31-4		Polychlorinated naphthalenes*		
– 1,2,3,7,8-Pentachlorodibenzofuran	57117-41-6		– 1-Chloronaphthalene	90-13-1	
			Endosulfan	115-29-7	Pesticide
			Short chain chlorinated parafins (SCCPs)*		
			– chlorinated paraffins: C12, 60% chlorine	no CAS rn (PubChem ID: 5284361)	
			– 1,2,3,4,6,7,10-Heptachlorododecane	1005111-47-6	

teins for which POPs are defined as “active” and with activity below their cytotoxicity threshold were included. A neighbor protein procedure was developed and applied, based on the binary interaction method used in high-throughput experiments to develop a scoring system that allows the prediction of uncharacterized links between 28 POP candidates (compounds listed

or under review in the Stockholm Convention or identified from literature studies) and proteins. As a last step, biological information (adverse outcome pathways (AOPs), pathways, gene ontology terms (GO) and diseases) was integrated to identify statistically significant connections between the chemicals and the biological events.

**Tab. 2: List of the 28 chemicals defined as POP candidates**

Chemicals that have proteins associated to them in the ToxCast database considering the cytotoxicity threshold are shown in grey. Chemicals that have more than 4 proteins associated to them, for which we performed an ORA analysis, are in bold font. NA, not analyzed for enrichment (proteins < 4)

Chemical name		CAS rn	Cytotoxicity	Number of proteins from assays	Number of proteins with PPAN
	1,2-Dichlorobenzene	95-50-1	1000 μM	7	18
BDE-47	2,2',4,4'-Tetrabromodiphenyl ether	5436-43-1	8.36 μM	no assays after cytotox threshold	NA
BDE-99	2,2',4,4',5-Pentabromodiphenyl ether	60348-60-9	6.94 μM	no assays after cytotox threshold	NA
BDE-153	2,2',4,4',5,5'-Hexabromodiphenyl ether	68631-49-2	6.66 μM	no assays after cytotox threshold	NA
PCB 118	2,3',4,4',5-Pentachlorobiphenyl	31508-00-6	no data	—	NA
	2,6-Dinitrotoluene	606-20-2	1000 μM	4	NA
	Atrazine	1912-24-9	1.25 μM	1	NA
	Benzene	71-43-2	1000 μM	—	NA
βHCH	beta-Hexachlorocyclohexane	319-85-7	11.38 μM	6	16
Dechlorane Plus	Bis(hexachlorocyclopentadieno)cyclooctane	13560-89-9	no data	—	NA
	Bisphenol-A	80-05-7	11.69 μM	33	39
	Chloramphenicol	56-75-7	1000 μM	4	NA
	Chlordimeform	6164-98-3	1000 μM	1	NA
	Chlorobenzene	108-90-7	1000 μM	1	NA
	Diazinon	333-41-5	1000 μM	30	35
p,p'-DDE	Dichlorodiphenyldichloroethylene	72-55-9	8.74 μM	4	NA
	Dimethomorph	110488-70-5	5.86 μM	13	21
	Methoxychlor	72-43-5	8.26 μM	11	18
	Naproxen	22204-53-1	1000 μM	0	NA
	Nitrobenzene	98-95-3	1000 μM	3	NA
	Parathion	56-38-2	8.74 μM	27	33
PFHxS	Perfluorohexane sulfonic acid	355-46-4	no data	—	NA
	Phenanthrene	85-01-8	8.76 μM	3	NA
	Phenol	108-95-2	1000 μM	3	NA
	Prometryne	7287-19-6	8.28 μM	16	25
	Pyrene	129-00-0	11.46 μM	3	NA
	Pyrimethanil	53112-28-0	1000 μM	34	41
PFOA	Perfluorooctanoic acid	335-67-1	5.35 μM	9	17
	Dicofol	115-32-2	7.89 μM	18	28

## 2.1 Chemical data

### Model generation

We extracted the chemical names of the 28 POPs from the Stockholm Convention list (as of January 2019)<sup>3</sup>, i.e., the 12 initial POPs and the 16 new POPs (Tab. 1). The 12 initial POPs are chemicals that have been recognized as causing adverse effects on humans and the ecosystem. They are categorized into three

classes, i.e., pesticides, industrial chemicals, and by-products. The 16 new chemicals are compounds with a risk profile; they are categorized in the same way as the 12 initial POPs.

### Prediction

A list of 28 chemicals suspected to be POPs or semi-persistent organic pollutants (sPOPs) was established, based on data from

<sup>3</sup> <http://chm.pops.int/TheConvention/ThePOPs/tabid/673/Default.aspx>

the literature, chemicals recommended for listing by the review committee of the Stockholm Convention, and chemicals under review by the Stockholm Convention (Tab. 2).

## 2.2 Chemical-protein associations

Chemical-protein associations were extracted from the publicly available CompTox and ToxCast databases (last accessed November 2019). The CompTox database (Williams et al., 2017), an online database maintained by the U.S. Environmental Protection Agency (EPA), contains information regarding multiple data types (physico-chemical properties, exposure, biological activities, etc.) for over 87,500 chemicals. The CompTox dashboard was used to identify the biological assays in which the POPs and sPOPs are active to determine the cytotoxicity information and respective links to AOPs. The ToxCast database is a U.S. EPA program that builds large collections of endpoint data of more than 700 high-throughput assays on chemicals to which humans are potentially exposed (Richard et al., 2016; Kavlock et al., 2012). The ToxCast dashboard was used to obtain information about the assays extracted from the CompTox database, for example the protein identifiers.

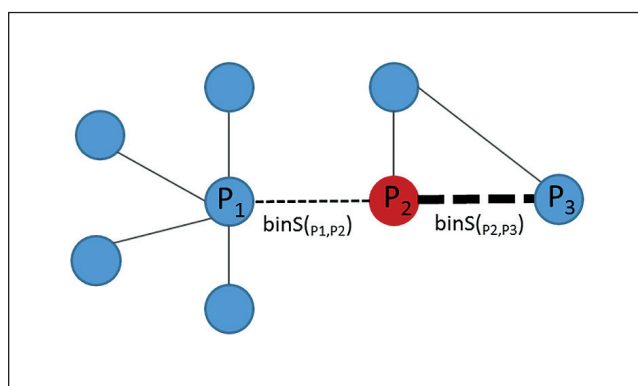
For the current investigation, only proteins passing the filtering process using the “cytotoxic-associated burst” region were retained (Judson et al., 2016) (Tab. S1<sup>4</sup>), as it has been demonstrated that compounds activate assays at concentration levels also observed for cytotoxicity or cell stress (Judson et al., 2016). Therefore, only assay activities (AC50) that occurred at a concentration below the “cytotoxic-associated burst” region were extracted.

## 2.3 Generating a high-confidence protein-protein associations network model

The relevant POP-protein links compiled and filtered from the CompTox and ToxCast databases were used to create the model. The POP-PPAN model was generated based on the previously published protein-protein association network (PPAN) approach (Audouze et al., 2010). Each protein (that linked to at least one POP) was represented as a node, and any protein-protein pairs for which at least one overlapping chemical was identified were linked by an edge. These protein-protein associations were converted into a non-redundant list of associations, i.e., if proteins A and B are linked, the network may have two associations A-B and B-A. In our approach, only one of these protein pairs was retained to create the model. To reduce noise and retain only the most relevant associations, each protein pair was scored (oS) based on the chemical overlapping information, i.e., counting shared POPs between each protein pair.

## 2.4 Neighbor protein procedure

To predict novel links between sPOPs and proteins, a neighbor protein procedure was developed, which is a three-step procedure: First, input proteins known to interact with the studied chemical were identified from the CompTox and ToxCast databases. Then, the developed POP-PPAN model was screened



**Fig. 2: Predictions of protein associations using a binary score (binS)**

An input protein (in red) known to interact with a studied chemical is screened into the POP-PPAN model. As results, network of first order interacting proteins are identified (in blue), and binS are calculated to rank and select high-confidence interactions. The  $\text{binS}(P_1, P_2)$  is a low-confidence score, as it has six unshared interaction partners (all the other blue ones). The  $\text{binS}(P_2, P_3)$  has a higher confidence score as it has only one unshared interaction partner, i.e., P1. Therefore, proteins P2 and P3 are more likely to be associated in the POP space, and protein P3 is predicted as a potential target for the chemical of interest.

for the input proteins in order to determine the first-order interacting proteins and the surrounding networks. This established a network of interacting proteins that contains the input proteins and proteins from the model. The last step was to calculate a binary score (binS) between each pair of proteins to rank and select putative proteins for the sPOPs. The developed binS is based on quality scoring of protein-protein interactions (PPIs) used in high-throughput experiments (such as yeast two-hybrid and protein-fragment complementation assay) to determine the reliability of the PPIs in large-scale datasets (Yu et al., 2008; Tarassov et al., 2008). A previous study has shown that the reliability of interactions between proteins correlates well with the number of non-shared interaction proteins for each interacting pair (de Lichtenberg et al., 2005). A binS was calculated for each protein pair as follows:

$$\text{binS}(P_1, P_2) = -\log_{10} ((N_{P1} + 1)(N_{P2} + 1)),$$

where  $N_{P1}$  and  $N_{P2}$  are the numbers of non-shared interaction partners for an association between proteins 1 ( $P_1$ ) and 2 ( $P_2$ ) (Fig. 2).

## 2.5 AOP-protein linkages

To characterize AOPs connected to the protein complexes, AOP information was integrated into the developed model using data from the CompTox and AOP-wiki databases<sup>5</sup> (August 28, 2019). In order to cover the biggest toxicological space possible,

<sup>4</sup> doi:10.14573/altex.1910161s

<sup>5</sup> https://aopwiki.org/





and since few AOPs had been validated by the date of the study, AOPs under development were also integrated. The AOP-wiki database, which is part of a knowledge structure of AOPs (Villeneuve et al., 2014), is an online database based on a collaborative program between the OECD and the European Commission.

## 2.6 Biological enrichment of the protein networks

The protein complexes associated with the sPOPs were independently tested for biological significance and disease associations. All proteins were converted to EntrezGeneID to facilitate analysis. To assess the relevance of protein -GO, -pathway and -disease linkages, each data source was individually tested using an over-representation analysis (ORA) based on a hypergeometric distribution. A significance level of 0.05 after Benjamini-Hochberg correction for multiple testing of  $p$ -values was used to select the most relevant associations.

### *Gene ontology-protein associations*

Investigation in terms of biological relevance was done for each of the GO categories, i.e., biological process, molecular function, and cellular component (Ashburner et al., 2000; The Gene Ontology Consortium, 2019).

### *Pathway-protein associations*

Functional pathways that may potentially be disturbed by chemicals were investigated using the Kyoto Encyclopedia of Genes and Genomes (KEGG) and the Reactome database (Kanehisa et al., 2019; Fabregat et al., 2018) (last access November, 2019).

### *Disease-protein enrichment*

For disease enrichment, the DisGeNET data source, which contains human protein-disease annotations, was used to identify disorders potentially connected with the studied chemicals (Piñero et al., 2017) (last access November, 2019). DisGeNET is the largest publicly available platform integrating knowledge on gene-disease linkages from diverse databases and published articles. The current version (v6.0) contains 628,685 gene-disease associations between 17,549 genes and 24,166 diseases and abnormal human phenotypes.

## 3 Results

Based on the CompTox and ToxCast databases, a PPAN model representative of the POP biological space was generated in order to predict potential toxic effects of suspected POPs (sPOPs).

We extracted the 28 POPs from the Stockholm Convention list. Among them, five are families of compounds; therefore, we selected specific chemicals that are good representatives of these chemical families to facilitate further analysis, resulting in a total of 35 compounds used to create the PPAN model (Tab. 1). Protein information could only be retrieved for 20 chemicals when their cytotoxicity thresholds, as defined by Judson et al. (2016), were taken into consideration (Tab. S1<sup>4</sup>). The cytotoxicity thresholds allow distinction between toxicity due to the disruption of a spe-

cific biomolecular function and toxicity due to general cell stress. These 20 chemicals were linked to 95 unique proteins via 247 associations; chlordecone was associated with 56 proteins (via 97 assays), and the human estrogen receptor 1 (ESR1) was associated with 10 compounds (via 29 assays) (Tab. S1<sup>4</sup>). We were not able to retrieve bioactive records for hexachlorobenzene (HCB), hexabromobiphenyl (PBB 153), hexabromocyclododecane (HBCD), and some polybrominated diphenyl ethers (PBDEs). To build the model, 198 unique associations between the 20 chemicals and 95 proteins were used.

## 3.1 Generating a high-confidence protein-protein association network (PPAN) in the biological space of POPs

Using the compiled chemical-protein information, we constructed a POP-PPAN model following the procedure developed previously (Audouze et al., 2010). This procedure is based on the assumption that if two proteins are biologically affected by the same POPs (defined as shared compounds), they are likely to be involved in a common mechanism of action of the chemicals. Therefore, two biological targets are linked to each other if they are associated with at least one common POP. The resulting model consists of 47 associations between 18 proteins (Fig. 3, Tab. S2<sup>4</sup>), meaning that many of the extracted proteins are not, or are not yet known to be, targeted by common POPs.

To highlight the most significant associations, an overlapping score (oS) was assigned to each pair of proteins based on the calculation of overlapping POPs for each protein pair. The smallest oS value being 3 (for example, ESR1 and ESR2 are targeted by the 3 POPs bHCH, chlordane, and chlordecone); the highest score was 8 (ESR1 and NR1H2 are targeted by endosulfan, diel-drin, bHCH, endrin, DDT, chlordane, chlordecone, and aldrin).

In order to evaluate the POP-PPAN in the toxicological space, we went one step further and integrated biological events from the known AOPs into the protein complex network using both the CompTox and AOP-wiki databases. We were able to link 7 of the 18 proteins to 12 MIEs and 7 KEs (the event ID 122 is defined as a MIE and a KE in AOPwiki) that are involved in 21 different AOPs, mainly ones related to reproductive dysfunctions (see Tab. S3<sup>4</sup> for full names of the AOP events, and links to MIEs and KEs). Only one AOP (AOP 167, Early-life estrogen receptor activity leading to endometrial carcinoma in the mouse) is connected to two events (MIE 1064, Pre-pubertal increase, estrogen receptor (ER) activity and KE 1065, Activation, estrogen receptor alpha). Four MIEs are connected to two AOPs each, i.e., MIE 111 (Agonism, estrogen receptor), MIE 122 (Activation, glucocorticoid receptor), MIE 245 (Activation, PXR/SXR), and MIE 1396 (Increased, glucocorticoid receptor activity) linked respectively to AOPs 29/52 (Estrogen receptor agonism leading to reproductive dysfunction / ER agonism leading to skewed sex ratios due to altered sexual differentiation in males), AOPs 14/214 (Glucocorticoid Receptor Activation Leading to Increased Disease Susceptibility / Network of SSRIs (selective serotonin reuptake inhibitors), AOPs 11/60 (Pentachlorophenol Acute Response by Percellome / PXR activation leading to hepatic steatosis), and AOPs 221/222 (Mental stress to depression / Mental stress to agitation).

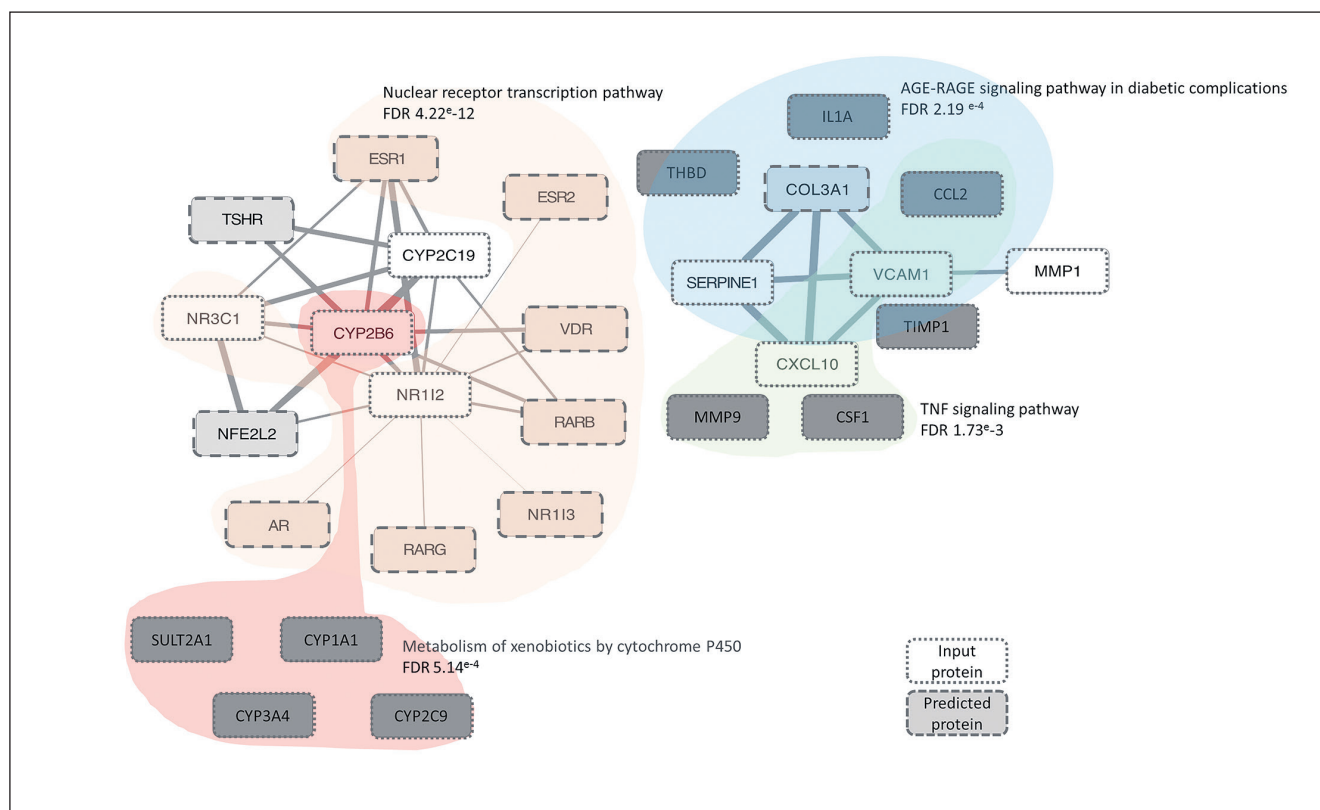
An edge is placed between two proteins (grey nodes) if they share at least one chemical (POP), and their width is proportional to the number of chemicals that are linked to both proteins (based on the calculated overlapping score (oS)). For example, the ESR1 and RARB share more POPs (oS = 6) than RARB and RARG (oS = 4). The white nodes circled in grey correspond to the biological events, and the white nodes circled in black to the AOPs extracted from the CompTox and AOP-wiki databases (for both, edges represent only a known interaction). The ID numbering is from the AOP-wiki database.

an sPOP and proteins. This neighbor protein procedure was applied to all 28 sPOPs; detailed analyses are presented for dicofol and PFOA below.

### 3.2.1 Systems toxicology prediction of dicofol

Dicofol is an organochlorine pesticide, structurally related to DDT, that is known to be involved in human adverse effects such as reproductive and neurobehavioral outcomes (Audouze and Grandjean, 2011). Dicofol is used as a pesticide and acaricide on a variety of fruits, vegetables, and field crops in many countries around the world. In May 2019, it was added to the Stockholm Convention on POPs due to its ability to bioaccumulate in living organisms and evidence of persistent criteria (half-life of 85 days in water and of 313 days in aerobic soil) (Kelly et al., 2007;<sup>6</sup>). Moreover, this substance has a high potential for long-range environmental transport, for example, it has been detected in the Arctic environment (Zhong et al., 2012). Even if the MoA of dicofol are not yet fully identified, adverse effects on human health and other living organisms have been reported, including effects on liver, kidney, and adrenal glands (Liu and Liu, 2012).

293



**Fig. 4: Protein-protein association network for Dicofol in the POP space**

Each node represents a protein, the input proteins being the ones extracted from the CompTox database, to which associated proteins from the POP-PPAN model were added (predicted proteins). The width of edges is according to the binary score (binS) calculated for each protein pair. Colored forms represent some significant pathways annotations from the KEGG and Reactome databases based on an over-representation analysis.

Dicofol may disturb biological functions, some being related to endocrine disruption effects (Vinggaard et al., 2000), and has been shown to exert generational perturbations (MacLellan et al., 1996). The World Health Organization (WHO) classifies dicofol as a Level II, “moderately hazardous” pesticide; the US EPA classifies dicofol in Group C, as “possible human carcinogen”. Based on these concerns, we decided to explore the potential toxicity of dicofol using the developed POP-PPAN model in order to propose potential linkages between this substance and adverse health effects.

Proteins known to be deregulated by dicofol were extracted from the CompTox database (as of July 17, 2019). Considering the cytotoxicity threshold (7.89  $\mu$ M, CompTox database as of July 2019), we were able to identify 18 biological targets (Tab. S4<sup>4</sup>). The POP-PPAN model was scanned with these 18 proteins, using the neighbor protein procedure. As a result, a protein complex of 18 proteins was identified that includes 10 proteins predicted to be associated with 8 of the known proteins, meaning that 10 proteins known to be linked to dicofol were not present in the protein complex (Tab. S4<sup>4</sup>). The resulting protein complex for dicofol, which contains 33 protein-protein associations (PPAs), was connected to 9 biological KEs of the know targets and to 9 KEs of the predicted proteins, linking respectively to 9 and 12 AOPs

(Tab. S4<sup>4</sup>). Most of the AOPs identified are linked to the reproductive and the nervous systems. Interestingly, the predicted glucocorticoid receptor NR3C1 (nuclear receptor subfamily 3 group C member 1) was associated with 7 AOPs, among these 3 are involved in male fertility dysfunctions and 3 in stress. A recent study from Nordkap et al. (2017) suggested a link between stress and testicular function that may be modulated by NR3C1, which is expressed in human testis (Chihara et al., 2016).

To obtain an overview of the pathways that may be dysregulated by the identified proteins, an ORA was performed on the full protein complex related to dicofol using the KEGG and Reactome databases (Fig. 4, Tab. S5<sup>4</sup>). The KEGG analysis reveals that some proteins are involved in 20 pathways, and among them 12 were statistically significantly associated with proteins (false discovery rate (FDR) > 5%). The most significant KEGG pathway was AGE-RAGE signaling pathway in diabetic complications (FDR  $2.19 \times 10^{-4}$ , via CCL2; COL3A1; IL1A; SERPINE1; THBD; VCAM1), a well-studied cascade of signaling mechanisms that initiates metabolic disorders such as diabetic complications (Kay et al., 2016). Among the other KEGG pathways, several were associated with metabolism (metabolism of xenobiotics by cytochrome P450, FDR  $5.14 \times 10^{-4}$  via CYP1A1; CYP2B6; CYP2C9; CYP3A4; SULT2A1; and linoleic acid metabolism,



FDR  $4.65 \times 10^{-3}$  via CYP2C19; CYP2C9; CYP3A4). Another significantly associated pathway was the signaling pathway of TNF (tumor necrosis factor), which is a multifunctional proinflammatory cytokine with effects on lipid metabolism, coagulation, insulin resistance, and endothelial function (FDR  $1.73 \times 10^{-3}$  via CCL2; CSF1; CXCL10; MMP9; VCAM1). The Reactome main pathway analysis (Tab. S5<sup>4</sup>) further indicates that 9 of the 28 proteins are annotated to the nuclear receptor transcription pathway (FDR  $4.22 \times 10^{-12}$ , AR; ESR1; ESR2; NR1I2; NR1I3; NR3C1; RARB; RARG; VDR). In total, 20 significant pathways were identified with the Reactome ORA analysis, most of them involving the different cytochromes associated with dicofol.

The enrichment analysis performed with the GO terms revealed that 27 of the 28 proteins are involved in metabolic processes (GO:0008152) and 10 are linked to reproduction (GO:0000003) (all biological processes, molecular functions, and cellular components are shown in Tab. S5<sup>4</sup>).

As a last step, disease enrichment was performed on the identified protein complex using the DisGeNET database. This allows us to highlight potential links between dicofol and disorders of the nervous system (depressive symptoms, unipolar depression, major depressive disorder, and brain ischemia with respective FDRs of  $1.11 \times 10^{-6}$ ,  $3.22 \times 10^{-3}$ ,  $3.40 \times 10^{-3}$ , and 0.03) and to the reproductive system (prostatic neoplasms, mammary neoplasms, female infertility, prostatic intraepithelial neoplasias, and endometrial neoplasm, with respective FDRs of  $1.15 \times 10^{-5}$ ,  $3.88 \times 10^{-4}$ ,  $7.98 \times 10^{-4}$ ,  $1.65 \times 10^{-3}$ , and  $4.98 \times 10^{-3}$ ) (Tab. S5<sup>4</sup>). Moreover, linkages were found for metabolic syndrome X, which is a cluster of conditions that occur together, leading to increased risk of heart disease, stroke, and type 2 diabetes as well as diabetic nephropathy, which is kidney damage that results from having diabetes.

### 3.2.2 Systems toxicology prediction of perfluorooctanoic acid (PFOA)

PFOA and its related substances are widely used in applications and consumer products such as fire-fighting foams, wetting agents, and cleaners. PFOA can be found in textiles, food packaging, and various paints to mention but a few sources. This chemical is highly persistent and does not undergo any further abiotic or biotic degradation under environmental conditions owing to its hydrolytically stable properties. Moreover, studies have shown that PFOA bio-accumulates in humans (Haug et al., 2010, 2011), and experimental and epidemiological evidence shows that PFOA is linked to adverse effects in humans and wildlife, e.g., affecting the endocrine system, leading to reproductive disorders<sup>7</sup> (Environment Canada and Health Canada, 2012). For example, PFOA, which is quickly absorbed, can be transferred to the fetus through the placenta, and to infants via breast milk<sup>8</sup>.

Potential toxicity events were explored using the POP-PPAN model. First, data were compiled from the CompTox database (as of July 18, 2019). Considering the cytotoxicity threshold ( $5.35 \mu\text{M}$ ), nine proteins were extracted (Tab. S6<sup>4</sup>). Using the POP-PPAN model, eight proteins were added, allowing the formation of a protein complex of 17 proteins linked to PFOA (Tab. S6<sup>4</sup>) that

contains 8 protein-protein associations. Only 3 AOPs (related to hepatic steatosis, breast cancer, and fetal testis) were identified via 3 targets, i.e., NR1I2, ESR1 and NR3C1.

ORA studies were performed on the diverse data sources. Among the 17 proteins involved in the complex linked to PFOA, two were not mapped (CYP2A2 and CYP2C11), meaning that these proteins are not yet annotated in the databases. The results of the GO terms investigation showed that the 15 proteins are classified in the metabolic process category (Tab. S7<sup>4</sup>). Among them, 8 are assigned to the developmental process and 4 to growth and reproductive processes. The KEGG pathway enrichment analysis did not reveal any significant result, meaning that no pathway was highly connected to PFOA. By exploring the non-significant associations, several metabolism-related pathways could be reported as “drug metabolism” (Tab. S7<sup>4</sup>). The enrichment analysis with the Reactome database identified 6 significant pathways, the most significant being the nuclear receptor transcription pathway (FDR of  $2.40 \times 10^{-8}$  via ESR1; NR1I2; NR3C1; RARB; RORA and VDR). Interestingly, the circadian clock, which is part of regulatory pathways, was also identified (FDR  $2.20 \times 10^{-2}$  via NR3C1; PPP1CA; RORA). No diseases were associated after enrichment of the protein complex using the DisGeNET database due to the small size of the protein complex. More knowledge and data integration are necessary in order to predict toxicity of such compounds. Looking at other existing databases such as the Comparative Toxicogenomics Database (CTD), PFOA was found to be connected to many more proteins, but not necessarily directly, and no cytotoxicity data were found (Davis et al., 2017). For instance, the biological target PPARA is highly associated with PFOA (in CTD as of September 8, 2019), via interactions such as “PPARA protein inhibits the reaction PFOA results in decreased expression of ADH7 mRNA” (Rosen et al., 2008), which is very difficult to integrate into a toxicological model.

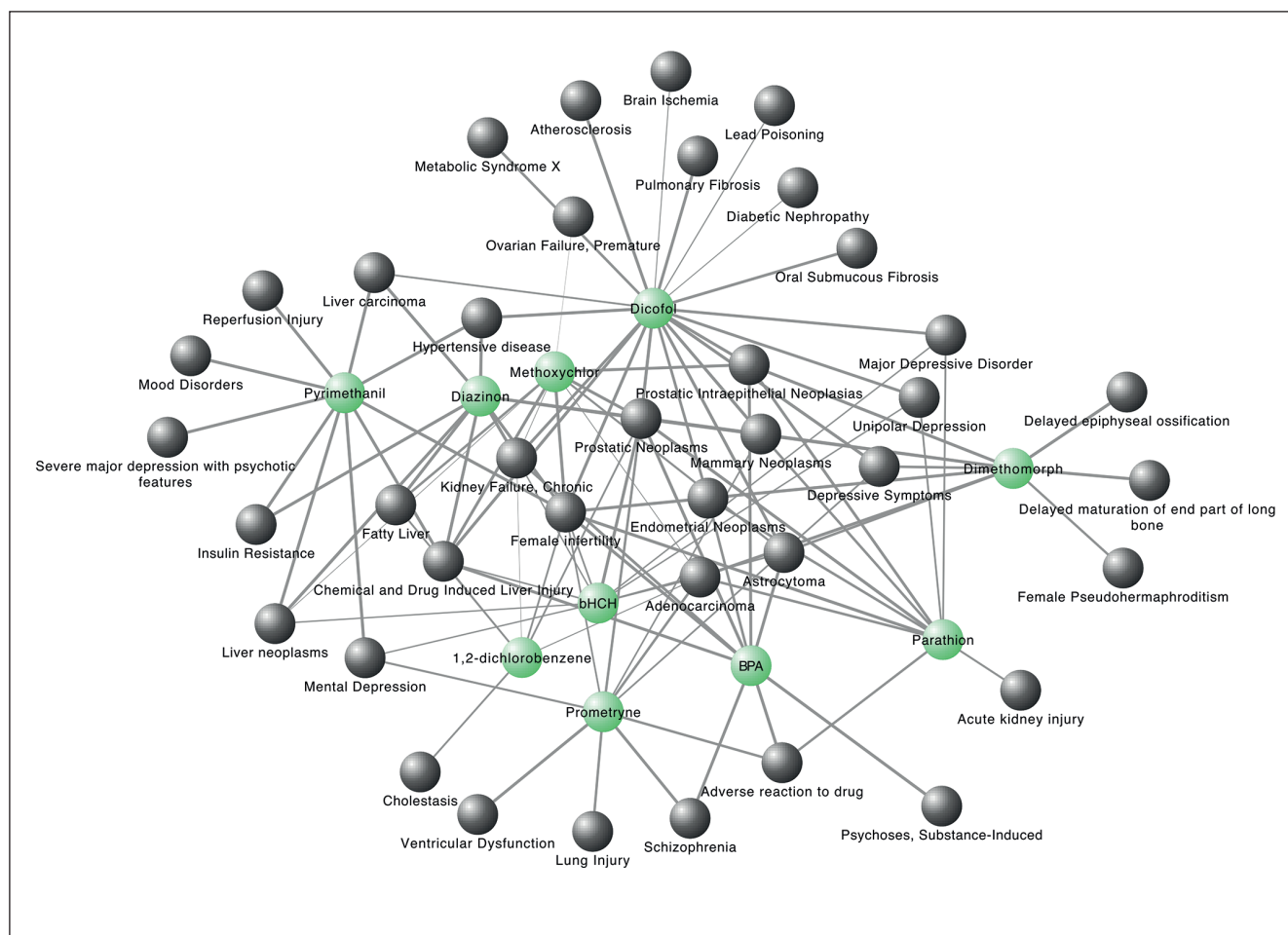
### 3.2.3 Global mapping of the 28 sPOPs in biological and disease spaces

Potential toxicities were explored using the POP-PPAN model for the other 26 listed sPOPs using the approach described above for dicofol and PFOA. First, sPOP-protein data were compiled from the CompTox database (as of November 2019), including the cytotoxicity thresholds (Tab. 2). Among the 28 sPOPs, no data was retrieved for three compounds, i.e., PCB 118, dechlorane plus, and perfluorohexane sulfonic acid (PFHxS). Three other chemicals did not pass the cytotoxicity threshold (BDE 47, BDE 99 and BDE 153), and fewer than 5 proteins were identified for 11 compounds. Therefore, only 11 sPOPs could undergo the systems toxicology analyses. Using the POP-PPAN model, first order interactor proteins were added individually to the 11 sPOP-proteins to form protein complexes (Tab. 2), and ORA analyses were performed.

In the GO analyses, the main biological processes involving proteins from the different identified protein complexes were the metabolic, the developmental and the reproductive processes. From a pathway point of view, based on the KEGG ORA, var-

<sup>7</sup> [https://www.ec.gc.ca/ese-ees/370AB133-3972-454F-A03A-F18890B58277/PFOA\\_EN.pdf](https://www.ec.gc.ca/ese-ees/370AB133-3972-454F-A03A-F18890B58277/PFOA_EN.pdf)

<sup>8</sup> <https://echa.europa.eu/documents/10162/8059e342-1092-410f-bd85-80118a5526f5>



**Fig. 5: Mapping of 10 POP candidates in the disease space**

Each green node represents a chemical, and each grey node a disease. Linkages between chemicals and disorders were obtained by an over-representation analysis using the DisGeNet database on the protein complexes associated to each chemical that were identified using the developed POP-protein network. The width of the edges is according to statistical significance.

ious metabolism pathways were retrieved as statistically significant, such as linoleic acid, retinol, arachidonic acid or tryptophan metabolism. Within the Reactome enrichment, mainly lipid metabolism-related pathways were identified. The last analysis was performed using the DisGeNet database in order to decipher linkage between the sPOPs and diseases. Two major systems appear to be affected by 10 of these candidate POPs (no statistical significance was found for PFOA), i.e., the nervous system and reproductive system (Fig. 5). Such findings are well in line with previous studies. A recent review highlighted the pivotal role of lipids in neuronal functions, emphasizing the potential use of lipids as biomarkers for major depressive disorders (Parekh et al., 2017). Our model allowed us to identify which sPOPs (dicofol, parathion, beta-hexachlorocyclohexane ( $\beta$ -HCH)) may be associated with major depressive disorders as well as other neural diseases. For example, to date, among the 405 available publications on dicofol, none mention a link between dicofol and major depression disorder (PubMed, as of December 09, 2019).

Seven chemicals were statistically associated with the reproductive disorder female infertility ( $2.56 \times 10^{-6} < \text{FDR} < 7.98 \times 10^{-4}$ ). Some links between the reproductive system and sPOPs are already known (methoxychlor), others have not been characterized yet (diazinon, dimethomorph, pyrimethanil) (PubMed, as of December 9, 2019).

## 4 Discussion

The proposed computational network-based model is different and complementary to existing molecular modelling approaches such as quantitative structure activity relationships (QSAR) and quantitative structure property relationships (QSPR) that are based on structural information of chemicals. For example, efficient QSAR and QSPR models have been developed to predict, in a quantitative manner, eco-toxicological, half-life or chemical properties of environmental interest that can be used for regula-

tory purposes and in order to design alternative chemicals that are safer for the environment (Khan et al., 2019a,b; Mansouri et al., 2018). Another strategy is the use of the read-across approach to develop predictive models to derive reference points for hazard characterization of untested chemicals from the available experimental data for structurally-similar compounds (Benfenati et al., 2019). A network-based model has the advantage of being independent of the chemical's structural information and can be used to propose a hypothesis for toxicological profiling of chemical substances. Such systems modeling methods combine various types of information, including chemical-protein interactions, protein-protein interactions, protein-disease annotations, chemical-adverse effect evidence or chemical exposure information (Knudsen et al., 2013; Kongsbak et al., 2014; Taboureau and Audouze, 2017). A network of complex diseases was developed by Gohlke et al. (2009) by integrating molecular pathways associated with both genetic and environmental factors.

In the presented approach, known POP-protein interactions were used to develop a selective and predictive POP-PPAN model to which diverse data types, such as AOP information, biological pathways, and protein-disease annotations, were integrated for biological interpretation. As results, the model identified, in a qualitative manner, putative links between candidate POP chemicals and mainly three adverse effect classes related to the reproductive, the nervous, and the metabolism systems.

Nevertheless, the development of such a NAM is still limited in terms of available data regarding chemical-protein interactions when taking into account the cytotoxicity levels. Indeed, only part of the available information was considered in order to create a more accurate and predictive model. An extension of the proposed model with integration of more data, including quantitative information such as dose levels, would be beneficial to fully evaluate the potential toxicity of emerging chemicals, including POPs, on human health.

Furthermore, one must take into consideration the so-called "Matthew effect", which describes the bias, in terms of available documentation and scientific literature, between well investigated chemicals and less well studied environmental compounds. This may impact computational models that are based on existing information, and therefore the potential innovation and discovery in research (Grandjean et al., 2011). Hence, the lack of association between a chemical and a biological event/disease may be the result of less available data used to create a model, rather than the lack of effect. Nevertheless, studies such as the one proposed here, highlight the chemicals that need further attention and more testing. To address the need for new data, increasing efforts based on alternative test systems are on-going. New technologies, including the continued expansion of HTS (e.g., Tox21) and multi-omics (e.g., toxicogenomics) allow the production of large amounts of data (Richard et al., 2016).

Another important issue to take into consideration when working with multi-data integration, is the harmonization and the availability of the data. To generate relevant and predictive models, the interoperability of the data based on the goals established by findable, accessible, interoperable and reusable (FAIR) data principles, is essential to support computational toxicology and chemical safety evaluation (Watford et al., 2019).

With the advances in data generation and bioinformatics technology, more heterogeneous data will be publicly accessible and could be integrated, such as qualitative and quantitative information and know-how from animal, clinical, epidemiological, exposure and biomonitoring studies, allowing the development of emerging approaches to help understand the effects of POPs on human health. The challenge will be to map and integrate all these data into a common platform (Pawar et al., 2019).

The use of innovative informatics technology is also useful as this will allow the identification, in a systemic manner, of already known and published information. For example, a novel tool based on artificial intelligence, AOP-helpFinder, uses text mining to screen abstracts of published articles to identify co-mentioned terms, which could be a chemical and an AOP event (Carvaillo et al., 2019; Rugard et al., 2019). Key information can also be identified by integrating multiple data sources from HTS studies by using frequent itemset mining (FIM) to develop a computational predictive model (Oki and Edwards, 2016).

## 5 Conclusion

To support chemical safety assessment, there is a need to develop and apply novel innovative and validated methods that are animal-free approaches in order to decipher the toxicological profiling of chemical substances such as POPs. While *in silico* testing cannot substitute for *in vitro* or *in vivo* testing, it can help focus on particular substances and targets to allow priority-setting for more efficient testing that can greatly help early risk assessment of POPs.

## References

- Andersen, M. E., McMullen, P. D., Phillips, M. B. et al. (2019). Developing context appropriate toxicity testing approaches using new alternative methods (NAM). *ALTEX* 36, 523-534. doi:10.14573/altex.1906261
- Ashburner, M., Ball, C. A., Blake, J. A. et al. (2000). Gene ontology: Tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* 25, 25-29. doi:10.1038/75556
- Audouze, K., Juncker, A. S., Roque, F. J. et al. (2010). Deciphering diseases and biological targets for environmental chemicals using toxicogenomics networks. *PLoS Comput Biol* 6, e1000788. doi:10.1371/journal.pcbi.1000788
- Audouze, K. and Grandjean, P. (2011). Application of computational systems biology to explore environmental toxicity hazards. *Environ Health Perspect* 119, 1754-1759. doi:10.1289/ehp.1103533
- Audouze, K., Brunak, S. and Grandjean, P. (2013). A computational approach to chemical etiologies of diabetes. *Sci Rep* 3, 2712. doi:10.1038/srep02712
- Audouze, K., Tromelin, A., Le Bon, A. M. et al. (2014). Identification of odorant-receptor interactions by global mapping of the human odorome. *PLoS One* 9, e93037. doi:10.1371/journal.pone.0093037
- Benfenati, E., Chaudhry, Q., Gini, G. et al. (2019). Integrating in silico models and read-across methods for predicting toxicity





- of chemicals: A step-wise strategy. *Environ Int* 131, 105060. doi:10.1016/j.envint.2019.105060
- Carvaillo, J.-C., Barouki, R., Coumoul, X. et al. (2019). Linking bisphenol S to adverse outcome pathways using a combined text mining and systems biology approach. *Environ Health Perspect* 127, 47005. doi:10.1289/EHP4200
- Chihara, M., Yoshihara, K., Ishiguro, T. et al. (2016). Association of NR3C1/glucocorticoid receptor gene SNP with azoospermia in Japanese men. *J Obstet Gynaecol Res* 42, 59-66. doi:10.1111/jog.12877
- Corton, J. C., Kleinstreuer, N. C. and Judson, R. S. (2019). Identification of potential endocrine disrupting chemicals using gene expression biomarkers. *Toxicol Appl Pharmacol* 380, 114683. doi:10.1016/j.taap.2019.114683
- Davis, A. P., Grondin, C. J., Johnson, R. J. et al. (2017). The comparative toxicogenomics database: Update 2017. *Nucleic Acids Res* 45, D972-D978. doi:10.1093/nar/gkw838
- de Lichtenberg, U., Jensen, L. J., Brunak, S. et al. (2005). Dynamic complex formation during the yeast cell cycle. *Science* 307, 724-727. doi:10.1126/science.1105103
- Deng, P., Li, X., Petriello, M. C. et al. (2019). Application of metabolomics to characterize environmental pollutant toxicity and disease risks. *Rev Environ Health* 34, 251-259. doi:10.1515/reveh-2019-0030
- Fabregat, A., Jupe, S., Matthews, L. et al. (2018). The reactome pathway knowledgebase. *Nucleic Acids Res* 46, D649-D655. doi:10.1093/nar/gkx1132
- Gohlke, J. M., Thomas, R., Zhang, Y. et al. (2009). Genetic and environmental pathways to complex diseases. *BMC Syst Biol* 3, 46. doi:10.1186/1752-0509-3-46
- Grandjean, P., Eriksen, M. L., Ellegaard, O. et al. (2011). The Matthew effect in environmental science publication: A bibliometric analysis of chemical substances in journal articles. *Environ Health* 10, 96. doi:10.1186/1476-069X-10-96
- Hartung, T., FitzGerald, R. E., Jennings, P. et al. (2017). Systems toxicology: Real world applications and opportunities. *Chem Res Toxicol* 30, 870-882. doi:10.1021/acs.chemrestox.7b00003
- Haug, L. S., Thomsen, C., Brantsaeter, A. L. et al. (2010). Diet and particularly seafood are major sources of perfluorinated compounds in humans. *Environ Int* 36, 772-778. doi:10.1016/j.envint.2010.05.016
- Haug, L. S., Huber, S., Becher, G. et al. (2011). Characterisation of human exposure pathways to perfluorinated compounds – Comparing exposure estimates with biomarkers of exposure. *Environ Int* 37, 687-693. doi:10.1016/j.envint.2011.01.011
- Judson, R., Houck, K., Martin, M. et al. (2016). Analysis of the effects of cell stress and cytotoxicity on in vitro assay activity across a diverse chemical and assay space. *Toxicol Sci* 153, 409. doi:10.1093/toxsci/kfw148
- Kanehisa, M., Sato, Y., Furumichi, M. et al. (2019). New approach for understanding genome variations in KEGG. *Nucleic Acids Res* 47, D590-D595. doi:10.1093/nar/gky962
- Kavlock, R., Chandler, K., Houck, K. et al. (2012). Update on EPA's ToxCast program: Providing high throughput decision support tools for chemical risk management. *Chem Res Toxicol* 25, 1287-1302. doi:10.1021/tx3000939
- Kay, A. M., Simpson, C. L., Stewart, J. A. et al. (2016). The role of AGE/RAGE signaling in diabetes-mediated vascular calcification. *J Diabetes Res* 2016, 6809703. doi:10.1155/2016/6809703
- Kelly, B. C., Ikononou, M. G., Blair, J. D. et al. (2007). Food web-specific biomagnification of persistent organic pollutants. *Science* 317, 236-239. doi:10.1126/science.1138275
- Khan, K., Baderna, D., Cappelli, C. et al. (2019a). Ecotoxicological QSAR modeling of organic compounds against fish: Application of fragment based descriptors in feature analysis. *Aquat Toxicol* 212, 162-174. doi:10.1016/j.aquatox.2019.05.011
- Khan, P. M., Baderna, D., Lombardo, A. et al. (2019b). Chemometric modeling to predict air half-life of persistent organic pollutants (POPs). *J Hazard Mater* 382, 121035. doi:10.1016/j.jhazmat.2019.121035
- Kleensang, A., Maertens, A., Rosenberg, M. et al. (2014). Pathways of toxicity. *ALTEX* 31, 53-61. doi:10.14573/altex.1309261
- Knudsen, T., Martin, M., Chandler, K. et al. (2013). Predictive models and computational toxicology. *Methods Mol Biol* 947, 343-374. doi:10.1007/978-1-62703-131-8\_26
- Kongsbak, K., Vinggaard, A. M., Hadrup, N. et al. (2014). A computational approach to mechanistic and predictive toxicology of pesticides. *ALTEX* 31, 11-22. doi:10.14573/altex.1304241
- Koual, M., Cano-Sancho, G., Bats, A.-S. et al. (2019). Associations between persistent organic pollutants and risk of breast cancer metastasis. *Environ Int* 132, 105028. doi:10.1016/j.envint.2019.105028
- Krewski, D., Acosta, D., Jr., Andersen, M. et al. (2010). Toxicity testing in the 21<sup>st</sup> century: A vision and a strategy. *J Toxicol Environ Health B* 13, 51-138. doi:10.1080/10937404.2010.483176
- Liu, Y. and Liu, R. (2012). The interaction of  $\alpha$ -chymotrypsin with one persistent organic pollutant (dicofol): Spectroscopy and molecular modeling identification. *Food Chem Toxicol* 50, 3298-3305. doi:10.1016/j.fct.2012.06.037
- MacLellan, K. N., Bird, D. M., Fry, D. M. et al. (1996). Reproductive and morphological effects of o,p'-dicofol on two generations of captive American kestrels. *Arch Environ Contam Toxicol* 30, 364-372. doi:10.1007/bf00212295
- Mansouri, K., Grulke, C. M., Judson, R. S. et al. (2018). OPERA models for predicting physicochemical properties and environmental fate endpoints. *J Cheminform* 10, 10. doi:10.1186/s13321-018-0263-1
- Naderi, M., Wong, M. Y. L. and Gholami, F. (2014). Developmental exposure of zebrafish (*Danio rerio*) to bisphenol-S impairs subsequent reproduction potential and hormonal balance in adults. *Aquat Toxicol* 148, 195-203. doi:10.1016/j.aquatox.2014.01.009
- Nordkap, L., Almstrup, K., Nielsen, J. E. et al. (2017). Possible involvement of the glucocorticoid receptor (NR3C1) and selected NR3C1 gene variants in regulation of human testicular function. *Andrology* 5, 1105-1114. doi:10.1111/andr.12418
- NRC – National Research Council (2007). *Toxicity Testing in the 21<sup>st</sup> Century: A Vision and a Strategy*. Washington, DC, USA: The National Academies Press. doi:10.17226/11970

- Oki, N. O. and Edwards, S. W. (2016). An integrative data mining approach to identifying adverse outcome pathway signatures. *Toxicology* 350-352, 49-61. doi:10.1016/j.tox.2016.04.004
- Parekh, A., Smeeth, D., Milner, Y. et al. (2017). The role of lipid biomarkers in major depression. *Healthcare (Basel)* 5, 5. doi:10.3390/healthcare5010005
- Pawar, G., Madden, J. C., Ebbrell, D. et al. (2019). In silico toxicology data resources to support read-across and (Q)SAR. *Front Pharmacol* 10, 561. doi:10.3389/fphar.2019.00561
- Piñero, J., Bravo, À., Queralt-Rosinach, N. et al. (2017). DisGeNET: A comprehensive platform integrating information on human disease-associated genes and variants. *Nucleic Acids Res* 45, D833-D839. doi:10.1093/nar/gkw943
- Richard, A. M., Judson, R. S., Houck, K. A. et al. (2016). ToxicCast chemical landscape: Paving the road to 21<sup>st</sup> century toxicology. *Chem Res Toxicol* 29, 1225-1251. doi:10.1021/acs.chemrestox.6b00135
- Rosen, M. B., Abbott, B. D., Wolf, D. C. et al. (2008). Gene profiling in the livers of wild-type and PPARalpha-null mice exposed to perfluorooctanoic acid. *Toxicol Pathol* 36, 592-607. doi:10.1177/0192623308318208
- Rugard, M., Courmoul, X., Carvaillo, J.-C. et al. (2019). Deciphering adverse outcome pathway network linked to bisphenol F using text mining and systems toxicology approaches. *Toxicol Sci* 173, 32-40. doi:10.1093/toxsci/kfz214
- Smirnova, L., Kleinstreuer, N., Corvi, R. et al. (2018). 3S – Systematic, systemic, and systems biology and toxicology. *ALTEX* 35, 139-162. doi:10.14573/altex.1804051
- Taboureau, O. and Audouze, K. (2017). Human environmental disease network: A computational model to assess toxicology of contaminants. *ALTEX* 34, 289-300. doi:10.14573/altex.1607201
- Tarassov, K., Messier, V., Landry, C. R. et al. (2008). An in vivo map of the yeast protein interactome. *Science* 320, 1465-1470. doi:10.1126/science.1153878
- The Gene Ontology Consortium (2019). The gene ontology resource: 20 years and still GOing strong. *Nucleic Acids Res* 47, D330-D338. doi:10.1093/nar/gky1055
- Villeneuve, D. L., Crump, D., Garcia-Reyero, N. et al. (2014). Adverse outcome pathway (AOP) development I: Strategies and principles. *Toxicol Sci* 142, 312-320. doi:10.1093/toxsci/kfu199
- Vinggaard, A. M., Hnida, C., Breinholt, V. et al. (2000). Screening of selected pesticides for inhibition of CYP19 aromatase activity in vitro. *Toxicol In Vitro* 14, 227-234. doi:10.1016/S0887-2333(00)00018-7
- Watford, S., Edwards, S., Angrish, M. et al. (2019). Progress in data interoperability to support computational toxicology and chemical safety evaluation. *Toxicol Appl Pharmacol* 380, 114707. doi:10.1016/j.taap.2019.114707
- Williams, A. J., Grulke, C. M., Edwards, J. et al. (2017). The CompTox chemistry dashboard: A community data resource for environmental chemistry. *J Cheminform* 9, 61. doi:10.1186/s13321-017-0247-6
- Yildirim, M. A., Goh, K.-I., Cusick, M. E. et al. (2007). Drug-target network. *Nat Biotechnol* 25, 1119-1126. doi:10.1038/nbt1338
- Yu, H., Braun, P., Yildirim, M. A. et al. (2008). High-quality binary protein interaction map of the yeast interactome network. *Science* 322, 104-110. doi:10.1126/science.1158684
- Zhong, G., Xie, Z., Cai, M. et al. (2012). Distribution and air-sea exchange of current-use pesticides (CUPs) from East Asia to the high Arctic Ocean. *Environ Sci Technol* 46, 259-267. doi:10.1021/es202655k

#### Conflict of interest

The authors declare they have no actual or potential competing financial interests.

#### Acknowledgments

The authors would like to acknowledge the CREATiVE project. This project has received funding from the French National Research Agency (ANR-18-CE34-0001-01). This work was also supported by the Université de Paris and INSERM.