Bioinformatics, 36(15), 2020, 4379–4381
doi: 10.1093/bioinformatics/btaa545
Advance Access Publication Date: 28 May 2020
Applications Note

Data and text mining

AOP4EUpest: mapping of pesticides in adverse outcome pathways using a text mining tool

Florence Jornod[1], Marylène Rugard[1], Luc Tamisier[2], Xavier Coumoul[1], Helle R. Andersen[3], Robert Barouki[1] and Karine Audouze[1,*]

[1]Université de Paris, T3S, Inserm UMR-S1124 and [2]Université de Paris, SPPIN CNRS UMR 8003, Paris F-75006, France and [3]Department of Environmental Medicine, Institute of Public Health, University of Southern Denmark, 5000 Odense C, Denmark

*To whom correspondence should be addressed.

Associate Editor: Jonathan Wren

Received on January 26, 2020; revised on May 18, 2020; editorial decision on May 19, 2020; accepted on May 23, 2020

Abstract

Motivation: Exposure to pesticides may lead to adverse health effects in human populations, in particular vulnerable groups. The main long-term health concerns are neurodevelopmental disorders, carcinogenicity as well as endocrine disruption possibly leading to reproductive and metabolic disorders. Adverse outcome pathways (AOP) consist in linear representations of mechanistic perturbations at different levels of the biological organization. Although AOPs are chemical-agnostic, they can provide a better understanding of the Mode of Action of pesticides and can support a rational identification of effect markers.

Results: With the increasing amount of scientific literature and the development of biological databases, investigation of putative links between pesticides, from various chemical groups and AOPs using the biological events present in the AOP-Wiki database is now feasible. To identify co-occurrence between a specific pesticide and a biological event in scientific abstracts from the PubMed database, we used an updated version of the artificial intelligence-based AOP-helpFinder tool. This allowed us to decipher multiple links between the studied substances and molecular initiating events, key events and adverse outcomes. These results were collected, structured and presented in a web application named AOP4EUpest that can support regulatory assessment of the prioritized pesticides and trigger new epidemiological and experimental studies.

Contact: karine.audouze@u-paris.fr

Availability and implementation: http://www.biomedicale.parisdescartes.fr/aop4EUpest/home.php.

Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

Pesticides are widely used chemical substances that target insects, fungi, weeds, rodents and other unwanted organisms. As most pesticides are not selective, they may affect non-target organisms including humans and impact their health. The general population is exposed to pesticide residues mainly in food and drinking water. Several adverse health effects are suspected to be caused by pesticides, including endocrine disruption and reproductive disorders (Andersen *et al.*, 2008; Kongsbak *et al.*, 2014; Wu *et al.*, 2020), developmental neurotoxicity and carcinogenic effects (Andersen *et al.*, 2015; Bailey *et al.*, 2015). Therefore, it is essential to have a better understanding of their modes of action (MoA) to identify their potential impacts on the human population.

Adverse outcome pathways (AOPs) capture existing biological knowledge, starting from a molecular initiating event (MIE) leading to an adverse outcome (AO), through a linear series of key events (KE) at various levels of the biological organization (Ankley *et al.*, 2010). Although AOPs are chemical-agnostic, it is useful to link chemicals to AOP events in order to identify their targets and their putative toxicities.

Novel and innovative approaches have been developed to analyze and identify key information from the very large number of published scientific papers and from high throughput screening studies (HTS). For example, a recent study combined information on toxic effects from manual literature searches with information from the ToxCast and AOP-Wiki databases to establish connections between environmental chemicals and toxic effects (Bajard *et al.*, 2019). Frequent itemset mining has also been used to decipher relevant information by integrating multiple data sources from HTS studies to establish predictive toxicological models (Oki *et al.*, 2016). A new informatics tool based on artificial intelligence, AOP-helpFinder, has been developed to identify reliable associations between events and a chemical of interest (Carvaillo *et al.*, 2019;

Rugard *et al.*, 2020). Yet, additional automated tools are required to screen even larger number of texts and datasets with a higher efficiency and relevance. Here, we have developed an updated version of the AOP-helpFinder tool called AOP-helpFinder 2 and applied it to a set of pesticides that have been prioritized in the human biomonitoring in Europe project (HBM4EU) (i.e. pyrethroids, chlorpyrifos and glyphosate) (https://www.hbm4eu.eu/). Python codes and examples are freely available to the scientific community under GitHub. To improve access to the results obtained for this set of pesticides, an open access web application, called AOP4EUpest, was developed to provide an extensive overview of known MoA for these chemicals (Fig. 1). The web application is complementary to existing ones, such as HExpoChem that allows prediction of potential toxic mechanisms of environmental chemicals through protein complexes and phenotypic outcomes enrichment (Taboureau *et al.*, 2013) (http://www.cbs.dtu.dk/services/HExpoChem-1.0/), sAOP (stressor-AOP) that represent the full mapping of bioactive compounds from the ToxCast database on the AOP space (Aguayo-Orozco *et al.*, 2019) (http://saop.cpr.ku.dk/) and ProTox-II which uses structural information of chemicals for toxicity prediction (Banerjee *et al.*, 2018) (http://tox.charite.de/protox_II/).

2 Materials and methods

To create the AOP4EUpest web application, we first improved the recently developed AOP-helpFinder tool that is based on Natural Language Toolkit (NLTK, version 3.2.5) (https://www.nltk.org/) and on Dijkstra graph theory (Carvaillo *et al.*, 2019). There are two major improvements in the new version. First, whereas the previous version of the tool uses compiled abstracts from specific toxicological databases (e.g. U.S. NIH TOXNET platform), AOP-helpFinder 2 is designed to pre-process and automatically screen scientific abstracts extracted from the PubMed database (https://www.ncbi.nlm.nih.gov/pubmed) to retrieve links between a stressor (e.g. a chemical) and a dictionary of events (e.g. AOPs or other specific events). Thus, this new feature actually increases the scientific knowledge space to explore and to identify additional co-occurrences. Second, since events from the AOP-Wiki database (MIE/KE/AO) are increasingly precise and complex (i.e. an event can be described by several words), the proposed updated version calculates the two scoring systems (i.e. the position score and the weighted score, see below) for all event types. This is different from the previous version in which the weighted score (wS) was calculated only for MIEs and KEs and the position score (pS) only for AOs. The wS identifies the shortest path between the words of an event in an abstract using the Dijkstra graph theory, meaning that abstracts are considered as acyclic graphs as described before (Carvaillo *et al.*, 2019). The pS is used to decipher the position of the event in the abstract, based on both the position of the event term relative to the other words and the total number of words in the abstract (Carvaillo *et al.*, 2019). It is meant to take into consideration whether the AOP event is more likely to be a working hypothesis (at the beginning of an abstract) or a finding/conclusion (at the end).

Fig. 1. Interface of the search page of the web application AOP4EUpest (http://www.biomedicale.parisdescartes.fr/aop4EUpest/home.php)

AOP-helpFinder 2 was applied to a list of pesticides prioritized by HBM4EU partners because of the lack of knowledge on their long-term health effects, and their exposure levels in humans (Supplementary Table S1). For each pesticide, an individual screening of the PubMed database that contains 30 072 881 articles (as of August 21, 2019) was performed. All abstracts related to one pesticide were saved in .xml files. A dictionary of events extracted from the AOP-Wiki ('aop_ke_mie_ao.tsv' file, as of August 26, 2019) was developed and used to carry out the abstracts text mining (Supplementary Table S2). This dictionary contains information regarding 165 MIEs, 704 KEs, 125 AOs, 12 MIE/KE (in the AOP-Wiki database an event can be defined as an MIE in one AOP and as a KE in another AOP) and 18 KE/AO (events defined as AO in some AOPs and as KEs in other AOPs) (Supplementary Table S2). Consequently, only abstracts co-mentioning a pesticide and at least one event were retrieved (Supplementary Table S1). pS and wS were calculated for each data pair as described previously in order to identify the most relevant scientific publications (Carvaillo *et al.*, 2019).

3 Results

For the 37 selected pesticides, a total of 16 346 articles, mentioning at least one of these chemicals, were identified. The pesticide with the fewest number of articles was epsilon-momfluorothrin (a type I synthetic pyrethroid insecticide) with only three publications. The chemical with the highest number of publications was chlorpyrifos (an organophosphate insecticide) with 4957 abstracts. To retrieve abstracts mentioning at least one event, the text mining was performed on each of the selected abstracts using AOP-helpFinder 2, followed by a manual curation. This procedure resulted in 1523 abstracts mentioning at least 1 event along with 1 of the 37 pesticides (Supplementary Tables S1 and S3). Among the 37 pesticides, 6 were not associated to any MIE/KE/AO, i.e. d-tetramethrin, empenthrin, imiprothrin, esbiothrin, zeta-cypermethrin and acrinathrin, reflecting the deficit in scientific knowledge on the health impact of those pesticides. To illustrate our approach, the search for glyphosate allowed for the identification of previously uncharacterized linkages for several events (Fig. 2). Such findings are complementary to those of existing databases such as AOP-Wiki and CompTox Dashboard, which do not include glyphosate-AOP links information (Williams *et al.*, 2017).

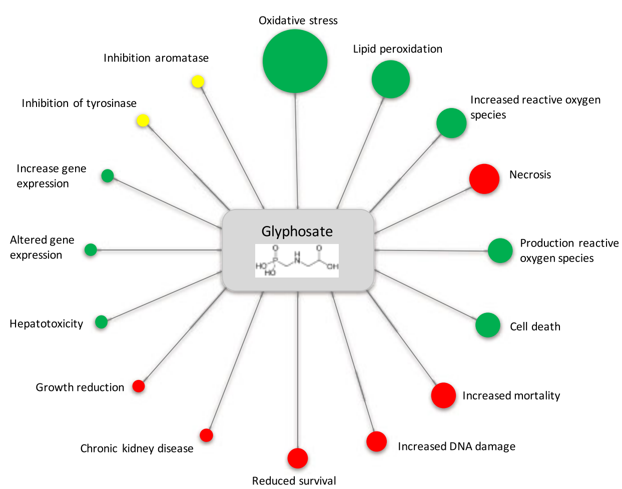

Fig. 2. Example of identified links between glyphosate and events using AOP-helpFinder 2 on the PubMed database. The size of the node reflects the number of abstracts co-mentioning glyphosate and one event. As example, oxidative stress was found in 71 abstracts, cell death in 10 abstracts and growth reduction in 4. Type of events are in yellow (MIE), green (KE) and red (AO). (Color version of this figure is available at *Bioinformatics* online.)

4 Conclusion

The AOP-helpFinder 2 tool explores, in an automatic way, existing published data. This was illustrated by the relevant information that the tool was able to retrieve for a set of highly prioritized pesticides and their linkages to AOP events. The AOP4EUpest web application is an easy and convenient presentation of those linkages and can assist in the identification of pesticides linked to AOPs and in the future, in the development of additional experimental studies. Such computational tools are part of the new approach methodologies, which are alternative methods to animal testing and may be of high interest to regulatory agencies and research related to risk assessment. Indeed, these tools can be applied for hazard identification of single substances, or of mixtures if AOP networks are considered and they can be used for risk assessment, particularly when they are combined with quantitative AOPs which also take exposure parameters into account.

Acknowledgement

The authors would like to acknowledge HBM4EU, Inserm and the Université de Paris for supporting the work.

Funding

This work was supported by the European Union's Horizon 2020 research and innovation programme [HBM4EU, grant agreement no. 733032], (https://www.hbm4eu.eu/).

Conflict of Interest: none declared.

References

Aguayo-Orozco,A. *et al.* (2019) sAOP: linking chemical stressors to adverse outcomes pathway networks. *Bioinformatics*, **35**, 5391–5392.

Andersen,H.R. *et al.* (2008) Impaired reproductive development in sons of women occupationally exposed to pesticides during pregnancy. *Environ. Health Perspect.*, **116**, 566–572.

Andersen,H.R. *et al.* (2015) Occupational pesticide exposure in early pregnancy associated with sex-specific neurobehavioral deficits in the children at school age. *Neurotoxicol. Teratol.*, **47**, 1–9.

Ankley,G.T. *et al.* (2010) Adverse outcome pathways: a conceptual framework to support ecotoxicology research and risk assessment. *Environ. Toxicol. Chem.*, **29**, 730–741.

Bailey,H.D. *et al.* (2015) Home pesticide exposures and risk of childhood leukemia: findings from the childhood leukemia international consortium. *Int. J. Cancer*, **137**, 2644–2663.

Bajard,L. *et al.* (2019) Prioritization of hazards of novel flame retardants using the mechanistic toxicology information from ToxCast and adverse outcome pathways. *Environ. Sci. Eur.*, **31**, 14.

Banerjee,P. *et al.* (2018) ProTox-II: a webserver for the prediction of toxicity of chemicals. *Nucleic Acids Res.*, **46**, W257–W263.

Carvaillo,J.-C. *et al.* (2019) Linking bisphenol S to adverse outcome pathways using a combined text mining and systems biology approach. *Environ. Health Perspect.*, **127**, 47005.

Kongsbak,K. *et al.* (2014) A computational approach to mechanistic and predictive toxicology of pesticides. *ALTEX*, **31**, 11–22.

Oki,N.O. *et al.* (2016) Accelerating adverse outcome pathway development using publicly available data sources. *Curr. Environ. Health Rep.*, **3**, 53–63.

Rugard,M. *et al.* (2020) Deciphering adverse outcome pathway network linked to bisphenol F using text mining and systems toxicology approaches. *Toxicol. Sci.*, **173**, 32–40.

Taboureau,O. *et al.* (2013) HExpoChem: a systems biology resource to explore human exposure to chemicals. *Bioinformatics*, **29**, 1231–1232.

Williams,A.J. *et al.* (2017) The CompTox Chemistry Dashboard: a community data resource for environmental chemistry. *J. Cheminform.*, **9**, 61.

Wu,Q. *et al.* (2020) Computational systems biology as an animal-free approach to characterize toxicological effects of persistent organic pollutants. *ALTEX*, **37**, 287–299.