Ref. Ares(2024)9121363 - 19/12/2024

ENDOMIX

Understanding how endocrine disruptors and chemical mixtures of concern target the immune system to trigger or perpetuate disease

HORIZON-HLTH-2023-ENVHLTH-02-03

Deliverable 2.1

MIXTURE PRIORITISATION

Lead beneficiary	UFZ
Author(s)	Georg Braun, Beate I. Escher
Dissemination level	PU
Type	DATA
Delivery date	31.12.2024

ENDOMIX is funded by the European Union's Horizon Europe Framework under Grant Agreement 101136566.

Funded by the European Union

Funded by the European Union

Table of Contents

Disclaimer

Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Health and Digital Executive Agency (HADEA). Neither the European Union nor the granting authority can be held responsible for them.

Funded by the European Union

Executive summary

Immunomodulation as a toxicity endpoint is currently lacking experimental data and hazard information for several groups of chemicals. The aim of Deliverable D2.1 was to identify chemicals that possess known or likely immunomodulatory properties and humans are likely exposed to, and what yet unknown immunomodulators could look like (Figure 1). Existing bioactivity data for 11 endocrine-disruptive and immunomodulatory endpoints from online repositories as well as from literature were combined with exposure predictions of 83,693 chemicals from physiologically-based toxicokinetic modelling in virtual populations. Mixture effects were calculated under the assumption of concentration additivity. The identified set of 7,874 chemicals allows for general scoring of priority and potential relevance of chemicals. The results of this prioritisation constitute the basis to decide which chemicals to include in mixture experiments in the ENDOMIX project WP2, 3 and 4. In the next step, we will investigate in WP1 if the predicted immunomodulators and mixture effect are actually present in human samples by evaluating existing cohort data and reanalysing non-target screening data. The identified chemicals and realistic mixtures will be tested for their bioactivity and mixture toxicity in established bioassay test batteries and novel effect-based methods. All tables that contain the collected and generated data are collated in Annex A, which is accessible on Zenodo (DOI: 10.5281/zenodo.14499300).

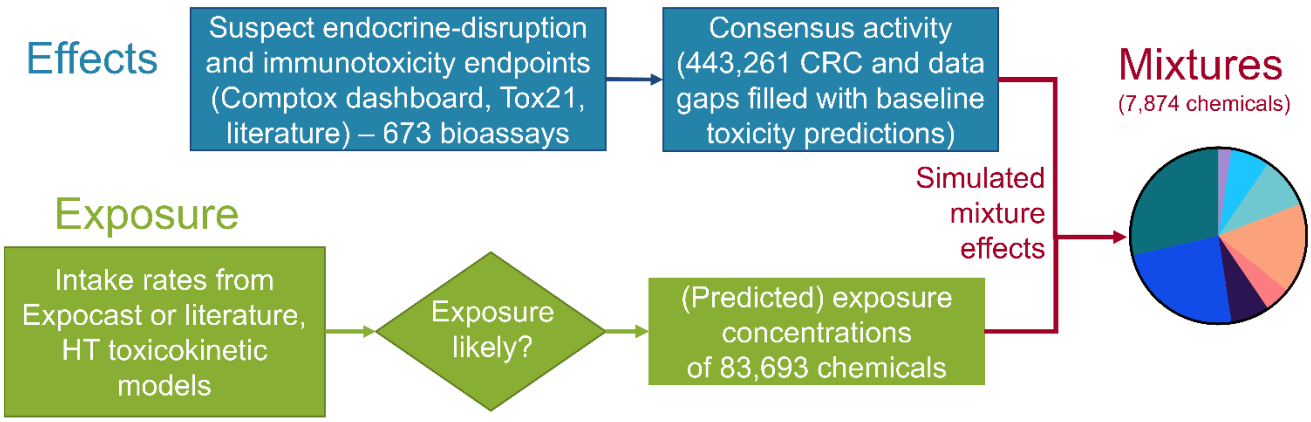

Figure 1: Outline of D2.1 and number of identified and prioritized chemicals in simulated realistic mixtures. CRC = concentration-response curves.

Funded by the European Union

1. Introduction

The human chemical exposome is a complex cocktail of chemicals with diverse properties and from various sources of exposure (Escher et al., 2020b; Vermeulen et al., 2020). We only have information on a small fraction of the chemical space that encompasses this chemical part of the exposome (Samanipour et al., 2024). As it is practically impossible to unravel the totality of the underlying chemicals, it is necessary to prioritise which chemical features seem to be of highest concern for health. Most often this results in a selection of chemical features based on exposure data such as frequent detection, or high signal intensities or concentrations. Another focus lies on compounds that were successfully annotated and show a high potency or bioactivity in test systems of selected toxicity endpoints.

We used a high-throughput *in silico* workflow, which is outlined in Figure 2 that combines measures of exposure and bioactivity to generate a list of chemicals that are likely to be present in human blood samples and contribute to mixture effects for endocrine-disruptive and immunomodulatory endpoints.

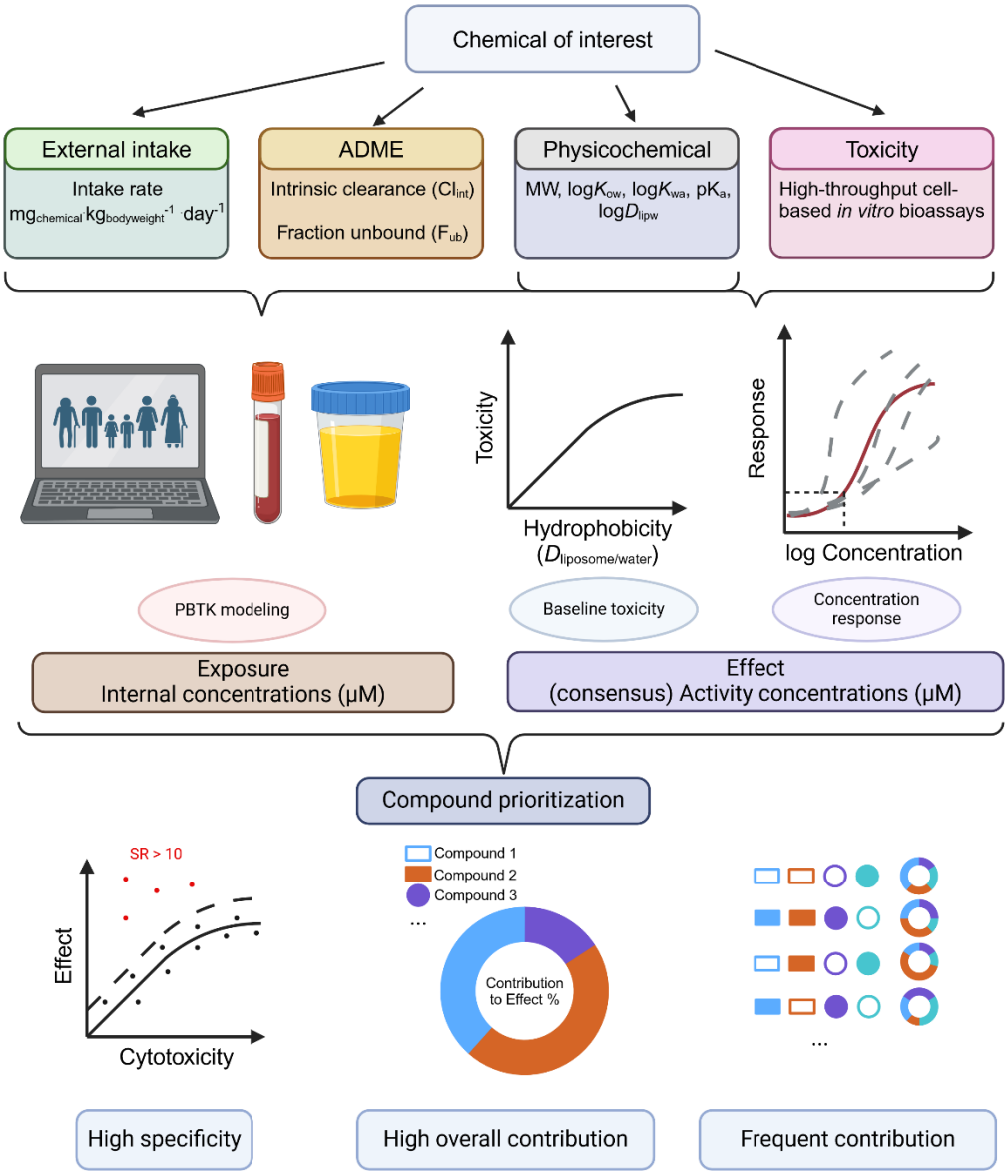

Figure 2: Approach for prioritization of chemicals and mixtures for testing in ENDOMIX. ADME = Absorption, distribution, metabolism, and elimination.

Funded by the European Union

2. Data and models used

The data was collected from public databases, or generated using models and source code from open-access repositories. The workflow was adapted from Braun and Escher (2023). The data processing was performed in R version 4.1.3 and Microsoft Excel.

2.1 Compound selection and properties

The selection of the compounds was mainly focused on lists of chemicals that were both reported endocrine disruptors or in general reported to be found in human blood. One major source were chemicals included in lists related to endocrine disruption and human exposure of the CompTox dashboard (Williams et al., 2017), namely “BISPHENOLS”, “COMPARA”, “EDSP21LIST1”, “EDSP21LIST2”, “EDSPUoC”, “EUCOSMETICS”, “HSDB2019”, “IRIS”, “PFASMASTERLISTV2”, “REACH2017”, “SUSDAT”, “SWISSPHARMA”, “TOXCAST_E1K”, “TOXCAST_INVITRODB_v4_1” (Williams et al., 2017). Further, we included reports of chemicals found in human blood from the curated database ExposomeExplorer (Neveu et al., 2016; Neveu et al., 2019), the Human Metabolome Database (Wishart et al., 2021), as well as the Blood Exposome Database (Barupal & Fiehn, 2019). The databases and repositories were accessed on 2024/02/21.

Identifiers such as CAS number, SMILES, InChIKey, and DTXSID were assigned as reported on the CompTox dashboard (Williams et al., 2017). The SMILES of the chemicals were transformed into SDF format using the open-access tool OpenBabel (O'Boyle et al., 2011). The generated SDF files were used as input for quantitative structure-activity relationship (QSAR) models included in OPERA version 2.9 (https://github.com/kmansouri/OPERA/releases/tag/v2.9.1). Predicted were physicochemical properties such as molecular weight, acidity constants (pK_a), fraction of the neutral form at pH 7.4, the Henry constant, water solubility, and melting point. Absorption, distribution, metabolism, and elimination (ADME) properties such as fraction unbound in plasma or human intrinsic hepatic clearance were also predicted using OPERA 2.9. The ChemmineR R package version 3.46.0 was utilized to calculate octanol-water partition coefficients K_{ow} (Cao et al., 2008).

Chemical exposure in doses of $mg_{chemical}$ kg^{-1} day^{-1} was predicted using models developed within the Systematic Empirical Evaluation of Models (SEEM) approach of the U.S. EPA (Wambaugh et al., 2018). If possible, the exposure was predicted via the consensus model as included in the SEEM3Predictor R package (https://github.com/HumanExposure/SEEM3RPackage/tree/main/SEEM3Predictor). This meta model predicted exposure based on four models and datasets, which were established using characterized chemicals reported in the National Health and Nutrition Examination Survey (NHANES). This means exposure for chemicals from diet, from consumer products, pesticides, or industrial chemicals (Ring et al., 2019). This required molecular fingerprints in the 729-bit toxprint format, which were generated from the SDF files using the software ChemoTyper version 1.3 (https://github.com/mn-am/chemotyper). For chemicals that could not be predicted with the consensus model, a simpler heuristic model based on molecular weight as included in the SEEM2Predictor R package (https://github.com/HumanExposure/SEEM3RPackage/tree/main/SEEM2Predictor) was applied (Wambaugh et al., 2014).

In total 83,693 chemicals were collected as dataset for the prioritization process (Annex A Table 1).

2.2 Bioassay data and processing

The toxicity endpoints considered for the prioritization was mainly focused on cell-based assays that are included in the 10k library of Tox21 and were accessible on the CurveSurfer interface of the Integrated Chemical Environment (Bell et al., 2017).

Funded by the European Union

In total 673 cell-based assays were selected (Annex A Table 2, Figure 3). This included the following 10 endocrine disruption endpoints:

- estrogenicity (116 assays),
- steroidogenesis (109 assays),
- androgenicity (74 assays),
- thyroid signalling (70 assays),
- activation of peroxisome proliferator-activated receptors (18 assays),
- activation of retinoid-x receptors (10 assays),
- activation of retinoic acid receptors (9 assays),
- activation of progesterone receptors (6 assays),
- activation of glucocorticoid receptors (5 assays),
- activation of the aryl hydrocarbon receptor (1 assay).

Immunomodulatory assays were all combined as one endpoint, which encompassed 255 assays.

In total 443,261 concentration response curves were collected (Annex A Table 3) and processed with an openly accessible R script (https://git.ufz.de/braung/automatedbioassayscreening). The evaluation was based on the tcpl R package version 3.1.0 (Filer et al., 2017), yet enforced the hill model for all active hitcalls. The output was the µM effect concentration that results in 10% absolute effect (EC_{10}) or a 1.5-fold induction ($EC_{IR1.5}$). For simplicity, all effect concentrations were eventually referred to as $EC_{benchmark}$. A confidence of the fit and derived effect concentrations was calculated which was the product of the coefficient of determination R^2 and $1-\left|\frac{\text{standard error } EC_{10}/EC_{IR1.5}}{EC_{10}/EC_{IR1.5}}\right|$. A confidence threshold of 0.7 was used to consider a fit as valid. If a chemical was evaluated as active in several assays per effect endpoint, the median consensus effect concentration was derived (Annex A Table 4).

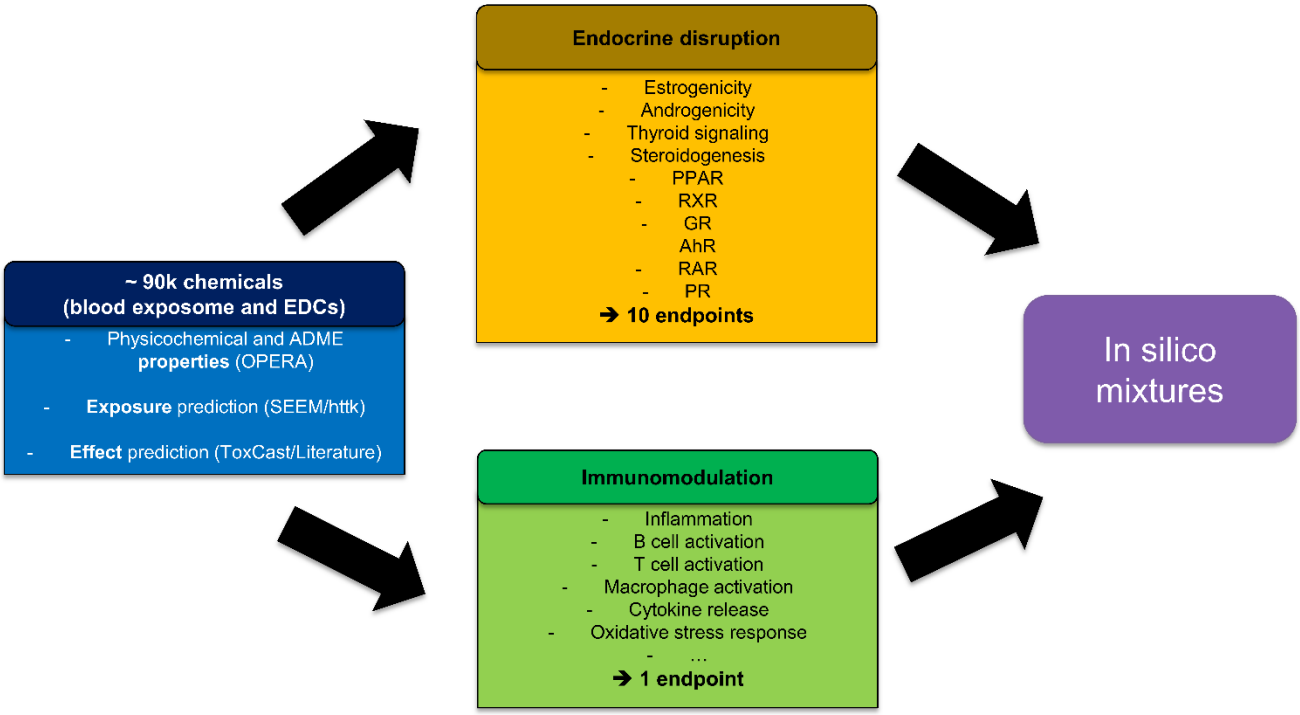

Figure 3: Bioassays included in the mixture prioritization. EDC = Endocrine-disrupting chemicals. ADME = Absorption, distribution, metabolism, and elimination.

Funded by the European Union

2.3 Baseline toxicity for data gap-filling

Baseline toxicity is the minimum toxicity a chemical expressed due to unspecific interaction with cellular membranes and other cell constituents. Baseline cytotoxicity, e.g. the 10% cytotoxicity concentration IC_{10} can be predicted with quantitative structure-activity relationships from liposome-water distribution ratios, $D_{lip/w}$ (Lee et al., 2021). This baseline toxicity was predicted for all 83,693 chemicals as a surrogate for potency (Annex A Table 1).

Structural read-across was utilized to fill data gaps in combination with predicted baseline toxicity. The toxprint fingerprints of all chemicals were used to calculate structural similarity to sets of confidently active chemicals per endpoint (Annex A Table 4), based on the Tanimoto coefficient (Willett et al., 1998).

There were three groups of fingerprints, namely perfect, very good, and suitable fingerprints. Perfect fingerprints show a 95% structural similarity, while molar mass deviated within two folds and $D_{lip/w}$ deviated at maximum by a factor of ten. Very good fingerprints had a 90% structural similarity, molar mass deviated within two folds and $D_{lip/w}$ was allowed to deviate within two orders of magnitude (factor of 100). Suitable fingerprints showed 80% similarity and molar mass deviated within three folds. The specificity ratios (SR) of the highest tiered fingerprints were used to predict effect concentrations of chemicals without experimental data by dividing the predicted baseline toxicity IC_{10} by the predicted SR. This meant first perfect fingerprints, second very good fingerprints, third suitable fingerprints. If more than one fingerprint was available per group, the specificity ratios were averaged. The respective predicted effect concentrations were labelled as based on “fingerprint similarity”.

2.4 Exposure modelling using high-throughput toxicokinetics

Potential plasma concentrations were calculated based on the available physicochemical and ADME properties via the httk R package version 2.2.2 (Pearce et al., 2017). Per chemical the 0%, 10%, 50%, and 90% steady-state micromolar plasma concentrations were approximated via Monte Carlo simulation utilizing physiological data of 20,000 individuals included in the httk virtual population (Annex A Table 1). The calculations were performed under the assumption of the well-stirred model and that clearance and bioavailability is restricted to free or unbound concentrations.

Per mixture design, randomly generated mixtures were considered. For each random mixture the concentration per chemical was drawn from the quantiles with different likelihoods: 80% probability of selecting the 0% quantile, 5% probability of selecting the 10% quantile, 10% probability of selecting the 50% quantile, 5% probability of selecting the 90% quantile. In total, 100,000 different mixtures were simulated per mixture design. The concentration ratios of the simulated mixtures were kept constant, while the absolute concentrations were scored to 10% absolute effect. Effects of each mixture were calculated assuming a linear concentration response for low effect levels and concentration addition as discussed in Escher et al. (2020a).

The mixture designs encompassed A) a simulation of all chemicals, B) chemicals that had $\log D_{lip/w}$ within the applicability domain of the baseline toxicity QSAR model ($0 \leq \log D_{lip/w} \leq 8$), C) chemicals that were reported in the human metabolome database, D) chemicals that only had valid experimental bioactivity data. All mixture designs were then united and underwent prioritization and scoring.

3. Prioritisation and scoring strategies

Chemicals that were modelled to be present in human blood and also had assigned bioactivity data, were considered for further prioritization and scoring.

3.1 Scoring and generation of final list

For each of the 11 bioactivity endpoints, all chemicals that contributed at least once to 90% of the cumulative mixture effect within the random mixtures were included in the final prioritization list. The final

Funded by the European Union

score was a product of A) the frequency factor, meaning the ratio of all the times the chemical contributed to 90% of a mixture effect and the absolute number of occurrence, B) an exposure score which is the mean of the binary inclusion (1 = present, 0 = not present) in lists as ExposomeExplorer, Blood Exposome Database, the U.S. EPA Integrated Risk Information System (IRIS), the U.S. EPA Chemical and Products Database (CPDat), and NHANES, C) a property score which was the mean of binary assignment of 1 for $\log K_{ow} \leq 8$, $\log K_{wa} \geq 4$, and $\log K_{wa} > -4$. Further, the total number of PubMed articles as well as number of articles that included both the name of the chemical and keywords for the bioactivity endpoint were included. PubMed was accessed via the pubmedR R package version 0.0.3 (https://github.com/massimoaria/pubmedR). An activity ratio and the respective number of considered experimental assays was also reported, which represented the ratio of positive hitcalls and all tested bioassays per bioactivity endpoint. This resulted in a set of 7,874 chemicals that were prioritized across all 11 bioactivity endpoints (Annex A Table 5).

4. Design and selection of mixtures

The set of 7,874 chemicals (Annex A Table 5) will be used in the ENDOMIX project to help defining mixtures of endocrine disruptors and immunomodulators for experimental testing. As a first step, two main groups of mixtures were defined (Section 4.1 and 4.2) that are now going forward to mixture formulation and testing bioassays of diverse complexity, starting with the high-throughput screening assays in WP2. In the future, additional mixtures will be identified by matching non-target screening data with the prioritised potential mixture effect drivers (section 4.3).

4.1 Mixtures of broadly analysed endocrine disruptors across Europe

To define mixtures that best represent chemical exposure across Europe, eight different studies, including studies from ENDOMIX partners, or databases were considered. This included reports from the INMA cohort (Montazeri et al., 2023), the HELIX project (Haug et al., 2018), the E3N cohort (Frenoy et al., 2024), the PELAGIE cohort (Warembourg et al., 2015), the ExposomeExplorer database (Neveu et al., 2019) which was accessed on 2024/08/28, and the HBM4EU dashboard (https://hbm.vito.be/eu-hbm-dashboard) which was accessed on 2024/08/26.

All concentrations were transformed to molar blood concentrations. Concentrations reported in ng/g_{lipid} were transformed to molar concentrations assuming a lipid content of 0.4% in human plasma or serum with a density of 1.025 g/mL. Urine concentrations in $\mu g/g_{creatinine}$ were transformed to daily intake rates in $mg_{chemical}/kg_{bodyweight}$/day via Bayesian inference under the assumption of a daily urine production of 1.4 L and an average creatinine level of 122.6 mg/DL as included in the bayesmarker R package version 0.0.0.9000 (Stanfield et al., 2022). The respective daily intake was then used to calculate quantiles of plasma concentrations via Monte Carlo simulation in httk as discussed in section 3.4.

132 chemicals were selected using this approach (Annex A Table 6). Based on use and manufacturing information available from PubChem, chemicals were sorted into use groups and if deemed necessary use and structural subgroups. Low, mean or median, and high molar concentrations were collected for each study, depending on the values reported. Overall, the lowest and highest as well as the median concentration across all studies per chemical were calculated and are listed in Annex A Table 6.

Criteria for selection were 1) high numbers of predicted relevance among the 11 bioactivity endpoints, 2) representation across many studies with a minimum of two independent studies considered, 3) structural diversity, 4) diversity in concentration levels.

30 chemicals from the five main assessment groups including phthalates (PHT), per- and polyfluoroalkyl substances (PFAS), polybrominated diphenyl ethers and polychlorinated biphenyls (BDEPCB), pesticides (PEST), phenols (PHEN), and polycyclic aromatic hydrocarbons (PAHs) were selected (Annex A Table 7). For chemicals that were reported as metabolites, the most likely precursors were selected at the same or median concentration level instead. The intention was to be more representative of mixtures found in blood rather than urinary metabolites.

4.2 Complex mixtures on the level of individuals

For the testing of complex mixtures with more than ten components, the exposure data of 294 up to 473 chemicals as reported in Braun et al. (2024) will be utilized. These complex mixtures and their single chemical components will be tested in cell-based high-throughput test systems to test for confirmation of mixture hypotheses such as concentration addition and the “Something from Nothing” phenomenon (Silva et al., 2002). The “Something from Nothing” phenomenon is based on the observation that chemicals that occur below individual effect thresholds, still contribute to mixture toxicity and can cause cumulative effects. The “Something from Nothing” is commonly observed for complex environmental mixtures , but was also confirmed to be relevant in human samples (Braun et al., 2024). The priority list will be used as a criterion for chemical selection.

4.3 Mixtures of newly identified and potential immunomodulators

The 7,874 chemicals reported in the priority list will be part of a ground-truthing process. They will be used as suspects for annotation of data acquired from liquid and gas chromatography coupled to high-resolution mass spectrometry. The samples and extracts utilized for the annotation originate from the cohorts of the projects and studies EPIC (Riboli, 1992), AIRWAVE (Elliott et al., 2014), PELAGIE (Warembourg et al., 2015), EXPANSE (Vlaanderen et al., 2021), and ATHLETE (Vrijheid et al., 2021). Successfully confirmed analytes will then be tested as single chemicals as well as mixtures in realistic concentration ratios in the established biotest batteries that target immunomodulation, to verify their predicted relevance and bioactivity.

5. Conclusion

A high-throughput *in silico* workflow was utilized to generate a list of potential immunomodulatory endocrine disruptors, which are expected to be relevant for human exposure. The list of 7,874 chemicals was generated considering 443,261 concentration-response data of 673 cell-based assays targeting endocrine disruption and immunomodulation. The list was and will be used for the prioritisation and selection of chemicals for experimental testing and design of mixtures in the ENDOMIX project.

Funded by the European Union

6. References

Barupal, D. K., & Fiehn, O. (2019). Generating the Blood Exposome Database Using a Comprehensive Text Mining and Database Fusion Approach. *Environmental Health Perspectives*, *127*(9), 097008.

Bell, S. M., Phillips, J., Sedykh, A., Tandon, A., Sprankle, C., Morefield, S. Q., Shapiro, A., Allen, D., Shah, R., Maull, E. A., Casey, W. M., & Kleinstreuer, N. C. (2017). An Integrated Chemical Environment to Support 21st-Century Toxicology. *Environmental Health Perspectives*, *125*(5), 054501.

Braun, G., & Escher, B. I. (2023). Prioritization of mixtures of neurotoxic chemicals for biomonitoring using high-throughput toxicokinetics and mixture toxicity modeling. *Environment International*, *171*, 107680.

Braun, G., Herberth, G., Krauss, M., König, M., Wojtysiak, N., Zenclussen, A. C., & Escher, B. I. (2024). Neurotoxic mixture effects of chemicals extracted from blood of pregnant women. *Science*, *386*(6719), 301-309.

Cao, Y., Charisi, A., Cheng, L.-C., Jiang, T., & Girke, T. (2008). ChemmineR: a compound mining framework for R. *Bioinformatics*, *24*(15), 1733-1734.

Elliott, P., Vergnaud, A.-C., Singh, D., Neasham, D., Spear, J., & Heard, A. (2014). The Airwave Health Monitoring Study of police officers and staff in Great Britain: Rationale, design and methods. *Environmental Research*, *134*, 280-285.

Escher, B., Braun, G., & Zarfl, C. (2020a). Exploring the Concepts of Concentration Addition and Independent Action Using a Linear Low-Effect Mixture Model. *Environmental Toxicology and Chemistry*, *39*(12), 2552-2559.

Escher, B., Stapleton, H. M., & Schymanski, E. L. (2020b). Tracking complex mixtures of chemicals in our changing environment. *Science*, *367*(6476), 388-392.

Filer, D. L., Kothiya, P., Setzer, R. W., Judson, R. S., & Martin, M. T. (2017). tcpl: the ToxCast pipeline for high-throughput screening data. *Bioinformatics*, *33*(4), 618-620.

Frenoy, P., Cano-Sancho, G., Antignac, J.-P., Marchand, P., Marques, C., Ren, X., Severi, G., Perduca, V., & Mancini, F. R. (2024). Blood levels of persistent organic pollutants among women in France in the 90's: Main profiles and individual determinants. *Environmental Research*, *258*, 119468.

Haug, L. S., Sakhi, A. K., Cequier, E., Casas, M., Maitre, L., Basagana, X., Andrusaityte, S., Chalkiadaki, G., Chatzi, L., Coen, M., de Bont, J., Dedele, A., Ferrand, J., Grazuleviciene, R., Gonzalez, J. R., Gutzkow, K. B., Keun, H., McEachan, R., Meltzer, H. M., Petraviciene, I., Robinson, O., Saulnier, P.-J., Slama, R., Sunyer, J., Urquiza, J., Vafeiadi, M., Wright, J., Vrijheid, M., & Thomsen, C. (2018). In-utero and childhood chemical exposome in six European mother-child cohorts. *Environment International*, *121*, 751-763.

Lee, J., Braun, G., Henneberger, L., Konig, M., Schlichting, R., Scholz, S., & Escher, B. I. (2021). Critical Membrane Concentration and Mass-Balance Model to Identify Baseline Cytotoxicity of Hydrophobic and Ionizable Organic Chemicals in Mammalian Cell Lines. *Chemical Research in Toxicology*, *34*(9), 2100-2109.

Montazeri, P., Güil-Oumrait, N., Marquez, S., Cirugeda, L., Beneito, A., Guxens, M., Lertxundi, A., Lopez-Espinosa, M.-J., Santa-Marina, L., Sunyer, J., Casas, M., & Vrijheid, M. (2023). Prenatal Exposure to Multiple Endocrine-Disrupting Chemicals and Childhood BMI Trajectories in the INMA Cohort Study. *Environmental Health Perspectives*, *131*(10), 107006.

Neveu, V., Moussy, A., Rouaix, H., Wedekind, R., Pon, A., Knox, C., Wishart, D. S., & Scalbert, A. (2016). Exposome-Explorer: a manually-curated database on biomarkers of exposure to dietary and environmental factors. *Nucleic Acids Research*, *45*(D1), D979-D984.

Neveu, V., Nicolas, G., Salek, R. M., Wishart, D. S., & Scalbert, A. (2019). Exposome-Explorer 2.0: an update incorporating candidate dietary biomarkers and dietary associations with cancer risk. *Nucleic Acids Research*, *48*(D1), D908-D912.

O'Boyle, N. M., Banck, M., James, C. A., Morley, C., Vandermeersch, T., & Hutchison, G. R. (2011). Open Babel: An open chemical toolbox. *Journal of Cheminformatics*, *3*(1), 33.

Pearce, R. G., Setzer, R. W., Strope, C. L., Wambaugh, J. F., & Sipes, N. S. (2017). httk: R Package for High-Throughput Toxicokinetics. *J Stat Softw*, *79*(4), 1-26.

Funded by the European Union

Riboli, E. (1992). Nutrition and cancer: Background and rationale of the European Prospective Investigation into Cancer and Nutrition (EPIC). *Annals of Oncology*, *3*(10), 783-791.

Ring, C. L., Arnot, J. A., Bennett, D. H., Egeghy, P. P., Fantke, P., Huang, L., Isaacs, K. K., Jolliet, O., Phillips, K. A., Price, P. S., Shin, H. M., Westgate, J. N., Setzer, R. W., & Wambaugh, J. F. (2019). Consensus Modeling of Median Chemical Intake for the U.S. Population Based on Predictions of Exposure Pathways. *Environ Sci Technol*, *53*(2), 719-732.

Samanipour, S., Barron, L. P., van Herwerden, D., Praetorius, A., Thomas, K. V., & O'Brien, J. W. (2024). Exploring the Chemical Space of the Exposome: How Far Have We Gone? *JACS Au*, *4*(7), 2412-2425.

Silva, E., Rajapakse, N., & Kortenkamp, A. (2002). Something from "Nothing" − Eight Weak Estrogenic Chemicals Combined at Concentrations below NOECs Produce Significant Mixture Effects. *Environmental Science & Technology*, *36*(8), 1751-1756.

Stanfield, Z., Setzer, R. W., Hull, V., Sayre, R. R., Isaacs, K. K., & Wambaugh, J. F. (2022). Bayesian inference of chemical exposures from NHANES urine biomonitoring data. *Journal of exposure science & environmental epidemiology*, *32*(6), 833-846.

Vermeulen, R., Schymanski, E. L., Barabási, A. L., & Miller, G. W. (2020). The exposome and health: Where chemistry meets biology. *Science*, *367*(6476), 392-396.

Vlaanderen, J., de Hoogh, K., Hoek, G., Peters, A., Probst-Hensch, N., Scalbert, A., Melén, E., Tonne, C., de Wit, G. A., Chadeau-Hyam, M., Katsouyanni, K., Esko, T., Jongsma, K. R., & Vermeulen, R. (2021). Developing the building blocks to elucidate the impact of the urban exposome on cardiometabolic-pulmonary disease: The EU EXPANSE project. *Environmental Epidemiology*, *5*(4).

Vrijheid, M., Basagaña, X., Gonzalez, J. R., Jaddoe, V. W. V., Jensen, G., Keun, H. C., McEachan, R. R. C., Porcel, J., Siroux, V., Swertz, M. A., Thomsen, C., Aasvang, G. M., Andrušaitytė, S., Angeli, K., Avraam, D., Ballester, F., Burton, P., Bustamante, M., Casas, M., Chatzi, L., Chevrier, C., Cingotti, N., Conti, D., Crépet, A., Dadvand, P., Duijts, L., van Enckevort, E., Esplugues, A., Fossati, S., Garlantezec, R., Gómez Roig, M. D., Grazuleviciene, R., Gützkow, K. B., Guxens, M., Haakma, S., Hessel, E. V. S., Hoyles, L., Hyde, E., Klanova, J., van Klaveren, J. D., Kortenkamp, A., Le Brusquet, L., Leenen, I., Lertxundi, A., Lertxundi, N., Lionis, C., Llop, S., Lopez-Espinosa, M.-J., Lyon-Caen, S., Maitre, L., Mason, D., Mathy, S., Mazarico, E., Nawrot, T., Nieuwenhuijsen, M., Ortiz, R., Pedersen, M., Perelló, J., Pérez-Cruz, M., Philippat, C., Piler, P., Pizzi, C., Quentin, J., Richiardi, L., Rodriguez, A., Roumeliotaki, T., Sabin Capote, J. M., Santiago, L., Santos, S., Siskos, A. P., Strandberg-Larsen, K., Stratakis, N., Sunyer, J., Tenenhaus, A., Vafeiadi, M., Wilson, R. C., Wright, J., Yang, T., & Slama, R. (2021). Advancing tools for human early lifecourse exposome research and translation (ATHLETE): Project overview. *Environmental Epidemiology*, *5*(5), e166.

Wambaugh, J., Isaacs, K., Phillips, K., Ring, C. L., Arnot, J., Bennett, D. H., Egeghy, P. P., Fantke, P., Huang, L., Jolliet, O., Shin, H.-M., Westgate, J. N., & Setzer, R. W. (2018). Systematic Empirical Evaluation of Models for Risk Prioritization. *ISEE Conference Abstracts*, *2018*(1).

Wambaugh, J. F., Wang, A., Dionisio, K. L., Frame, A., Egeghy, P., Judson, R., & Setzer, R. W. (2014). High throughput heuristics for prioritizing human exposure to environmental chemicals. *Environ Sci Technol*, *48*(21), 12760-12767.

Warembourg, C., Debost-Legrand, A., Bonvallot, N., Massart, C., Garlantézec, R., Monfort, C., Gaudreau, E., Chevrier, C., & Cordier, S. (2015). Exposure of pregnant women to persistent organic pollutants and cord sex hormone levels. *Human Reproduction*, *31*(1), 190-198.

Willett, P., Barnard, J. M., & Downs, G. M. (1998). Chemical Similarity Searching. *Journal of Chemical Information and Computer Sciences*, *38*(6), 983-996.

Williams, A. J., Grulke, C. M., Edwards, J., McEachran, A. D., Mansouri, K., Baker, N. C., Patlewicz, G., Shah, I., Wambaugh, J. F., Judson, R. S., & Richard, A. M. (2017). The CompTox Chemistry Dashboard: a community data resource for environmental chemistry. *Journal of Cheminformatics*, *9*(1), 61.

Wishart, D. S., Guo, A., Oler, E., Wang, F., Anjum, A., Peters, H., Dizon, R., Sayeeda, Z., Tian, S., Lee, Brian L., Berjanskii, M., Mah, R., Yamamoto, M., Jovel, J., Torres-Calzada, C., Hiebert-

Funded by the European Union

Giesbrecht, M., Lui, Vicki W., Varshavi, D., Varshavi, D., Allen, D., Arndt, D., Khetarpal, N., Sivakumaran, A., Harford, K., Sanford, S., Yee, K., Cao, X., Budinski, Z., Liigand, J., Zhang, L., Zheng, J., Mandal, R., Karu, N., Dambrova, M., Schiöth, Helgi B., Greiner, R., & Gautam, V. (2021). HMDB 5.0: the Human Metabolome Database for 2022. *Nucleic Acids Research*, *50*(D1), D622-D631.

Annex A

Accessible on Zenodo (DOI: 10.5281/zenodo.14499300)

Funded by
the European Union