# Cat-App:[1] learnings from a multi-year research programme on alternatives to animal testing

Enquiries: hans.ketelslegers@concawe.eu

One of the aims of the REACH regulation is to promote alternative methods for the hazard assessment of substances in order to reduce the number of tests on animals. However, in practice it has proven to be challenging to obtain regulatory acceptance for the application of such alternatives alongside existing toxicology data to minimise or replace the standard required animal tests. At the same time, the alternative options currently proposed under REACH are not practically applicable to petroleum substances. Concawe's Cat-App project aims to address these challenges through ongoing research which helps to ensure that excessive animal testing is avoided and the opportunities provided by REACH to innovate the conservative toxicology testing paradigm can eventually be put into practice.

## A 'twin challenge'

Because of the increasing and extensive need for animal testing as a default requirement to fulfil human health endpoints in the REACH dossiers, there is pressure to minimise laboratory animal use when complying with the REACH regulation. However:

1. it is challenging to justify the application of available alternative methods to avoid unnecessary animal testing under the regulation; while at the same time
2. the currently available alternative methods are not practically applicable to petroleum substances due to their inherent chemical complexity.

A rough calculation conducted by Concawe estimates that strict compliance with all data requirements in the petroleum substance REACH dossiers would incur a worst-case testing cost of more than €400 million for the current 168 actively registered substances, together with more than 25 years of testing and a need for around 1 million animals.[2, [1]] This is clearly undesirable from both animal welfare and cost perspectives. Furthermore, in terms of innovation, with an industry in transition, as well as from a timing perspective (it would take decades for these animal test programmes—and, therefore, the regulatory assessments—to be concluded), such an approach is not sustainable. In addition, the benefit that these extensive programmes will bring to better protect human health from the potential risks of exposure to petroleum substances is questionable. Both the hazards and risks are already assessed and carefully managed, based on a conservative application of available toxicological data on categories of petroleum substances from decades of testing under regulatory schemes, extensive research programmes and continuously growing expertise.

To address the needs of both regulators and the industry, there are opportunities under the regulation to avoid unnecessary animal testing and speed up the regulatory assessment. The two main approaches are the concept of data sharing (e.g. joint chemical dossier submissions through a consortium of companies registering the same chemical), and the application of alternative methods and approaches. The latter are described in Annex XI of the REACH regulation,[2] and the main tool described therein is the use of grouping and 'read-across'. The idea is that substances with similar molecular structures can be grouped together, and data on one substance can be applied via read-across to another one with a similar molecular structure for which no data are available.

This is a straightforward concept, until your substances contain thousands to millions of molecules ...

1 New technologies to underpin CATegory APProaches and read-across in regulatory programmes.

2 Figures previously published in Concawe (2019)[1] have been updated for this article based on the most recent number of registered substances and cost estimates (Spring 2021).

## The aim of Cat-App

How can one apply grouping and read-across of data, when the similarity between the group members cannot be exhaustively proven—at least not by structural data, which is the main regulatory requirement? Petroleum substances are so-called UVCBs: substances which are of partly unknown or variable composition, complex reaction products, or of biological origin. In other words, these are substances that are challenging to assess, as they may contain thousands to millions of molecules and are variable in nature; for example, crude oil composition varies between fields and with production time, as well as due to the physical chemistry of boiling crude oil. This means that a substance will never be 100% the same if sampled from one day to another, and its composition can never be described with 100% accuracy. Having said that, petroleum substances are only made of hydrocarbons within a defined range of carbon chain length, and possess various chemical characteristics (aliphatics, naphthenics, aromatics) which constitute the hydrocarbon space of each substance. The constituents that matter from a hazard point of view, such as benzene and polycyclic aromatic hydrocarbons, can be identified and quantified. In addition, the compositions will not vary endlessly as they will need to meet product specifications which limit their boundary compositions. Nevertheless, from a regulatory perspective, it is not yet certain whether these minimal requirements — and the remaining uncertainty — are acceptable for describing UVCB substances.

The question is whether it matters when we do not exhaustively know the composition of the substance. What matters most, for regulatory purposes, is the confidence that we do not underestimate the potential hazards and risk of any substance and all of its constituents, i.e. describing and understanding the full chemical composition (the chemical space or, for a petroleum substance, its hydrocarbon space) is necessary for hazard assessment. Elaborated analytical research is enabling an increased understanding of the hydrocarbon space of petroleum substances. Because we know that the analytical composition of a substance drives its biological response, we can hypothesise that a group of complex petroleum substances within a globally similar hydrocarbon space will have a similar (global) biological response. The 'hydrocarbon space' now becomes a 'hydrocarbon-biological space', adding additional confidence in the grouping of substances with multi-dimensional data, to ultimately tackle the challenges described earlier with the application of read-across of data on petroleum substance UVCBs while not underestimating the potential hazards.

Such a framework, which Cat-App aims to achieve, will enable the most optimal use of the available toxicological information on petroleum substances by chemical-biological read-across, and will help to target additional animal testing in an informed way and only where really needed as a last resort, instead of blindly testing all substances where a data gap exists.

> ***Cat-App is based on the concept of chemical-biological read across***
>
> *The chemical space of a group of petroleum substances is defined by their hydrocarbon constituents with a specific range of carbon chain length and chemical characteristics. These constituents drive the biological responses of these substances, i.e. they define their biological space. Substances can therefore be grouped in both dimensions of the chemical-biological space, facilitating read-across supported by chemical and biological parameters.*

This general concept might sound straightforward, but we still find ourselves on an exciting bumpy road of defining and generating the necessary scientific evidence for this framework while not losing sight of the practical relevance in a regulatory context. The remainder of this article explains how the first phase of this journey, which started in 2015, was completed. It presents a selection of the data that best reflect the main findings so far, explaining why we are not yet at the journey's end, and describing the opportunities that exist to enable progress in the years ahead.

## A multi-year international research consortium

In 2016, an article in a previous edition of the Concawe *Review* (Vol. 25, No. 2) described how the field of toxicological sciences was changing. In particular, the article described how a vision for toxicology testing in the 21st century, published by the US National Research Council[3] and known as Tox21c, *'has fuelled the discussion and changed the perspective on conservative animal-based toxicology studies, driven by animal welfare considerations and the revolutionary advances made in the field of biotechnology over the past decades. The main aim of Tox21c is to take advantage of these technological breakthroughs and move away from a regulatory testing paradigm that is currently still based on vertebrate animal models, following the '3R'*[4] *principle in toxicology testing: Refinement, Reduction and eventual Replacement of animal studies for research purposes'*.[5] Therefore, instead of studying observable outcomes in response to chemical exposures in an animal, such as the formation of a tumour 'in vivo',[3] one would eventually predict such an effect by studying the cellular or molecular mechanisms in the initiation and formation of a tumour 'in vitro'.[4] Unfortunately, as will be explained later, in-vitro assays are currently still not sufficient to fully *Replace* an animal test in order to predict toxicity in a human and meet regulatory requirements. Nevertheless, such types of mechanistic, or biological, responses observed after exposure of a cellular system to a substance provide highly valuable knowledge which can be smartly applied as supporting information to further *Refine* and *Reduce* required animal testing. These are the types of biological response data that are obtained from in-vitro assays, which are at the heart of the Cat-App framework.

> *At the heart of the Cat-App framework are in-vitro data: mechanistic, or biological, responses observed after exposure of a (human) cell system to a substance which can be smartly applied as supporting information to further refine and reduce required animal testing.*

[3] In vivo: Latin for 'within the living', i.e. testing in a whole living animal.

[4] In vitro: Latin for 'in glass', i.e. testing the components of an organism isolated from their normal biological context (organs, cells, subcellular components, molecules such as DNA, etc.).

To develop this approach, Concawe established a research consortium in 2015 with multiple partners and scientific advisers from the US, UK, Canada, Asia and Europe. Participating organisations and their roles are shown in Figure 1.

Figure 1: Overview of the Cat-App research consortium


* Induced pluripotent stem cells

The work began in 2016, and was divided into five different work packages (WPs). WP1 collected samples, as well as any relevant existing data (e.g. phys-chem and analytical), from 141 petroleum substances. By extracting the biologically active fraction of the substance using dimethylsulphoxide (DMSO), this WP also coordinated the generation of petroleum substance extracts (PS-E[5]) which ensure that the lipophilic substances can be introduced into the aqueous environment of the in-vitro assays.

[5] In this article, the test samples are referred to as 'PS-E' to indicate the distinction from full petroleum substance UVCBs.

# Cat-App: learnings from a multi-year research programme on alternatives to animal testing

This DMSO extraction approach is an established methodology, and the PS-E obtained using this method are used routinely for safety testing (e.g. mutagenicity) and chemical characterisation of the refinery streams.[6] Overall, the 141 PS-E tested in Cat-App represent the entire continuum of active petroleum substance registrations under REACH, from diverse manufacturing process categories. For statistical visualisation purposes, some categories were merged together, which lead to the 16 Cat-App-specific PS-E categories shown in Table 1.

*Cap-App Work Package 1 was responsible for coordinating the generation of petroleum substance extracts (PS-E) which ensure that the biologically active fraction of the lipophilic petroleum substances can be introduced into the aqueous environment of the in-vitro assays.*

Table 1: Cat-App-specific petroleum substance categories used in this project

| Category | Abbreviation* | Number of samples in category |
|---|---|---|
| Petrolatums | P.LAT | 3 |
| Paraffin and hydrocarbon waxes/slack waxes | WAX (2) | 10 |
| Low boiling point naphthas (gasolines) | NAPHTHA | 10 |
| Other lubricant base oils/highly refined base oils | BO (2) | 33 |
| Kerosines/MK1 diesel fuel | KER (2) | 10 |
| Foots oils | FO | 3 |
| Other gas oils | OGO | 4 |
| Bitumens/oxidised asphalt | BIT (2) | 5 |
| Residual aromatic extracts | RAE | 2 |
| Treated distillate aromatic extracts | TDAE | 2 |
| Heavy fuel oil components | HFO | 27 |
| Unrefined/acid treated oils | UATO | 4 |
| Cracked gas oils | CGO | 8 |
| Vacuum gas oils, hydrocracked gas oils and distillate fuels | VHGO | 10 |
| Straight-run gas oils | SRGO | 6 |
| Untreated distillate aromatic extracts | UDAE | 4 |

* The number in brackets represents the number of Concawe categories that were analysed together in Cat-App, which in total makes 20 categories.

Notes:

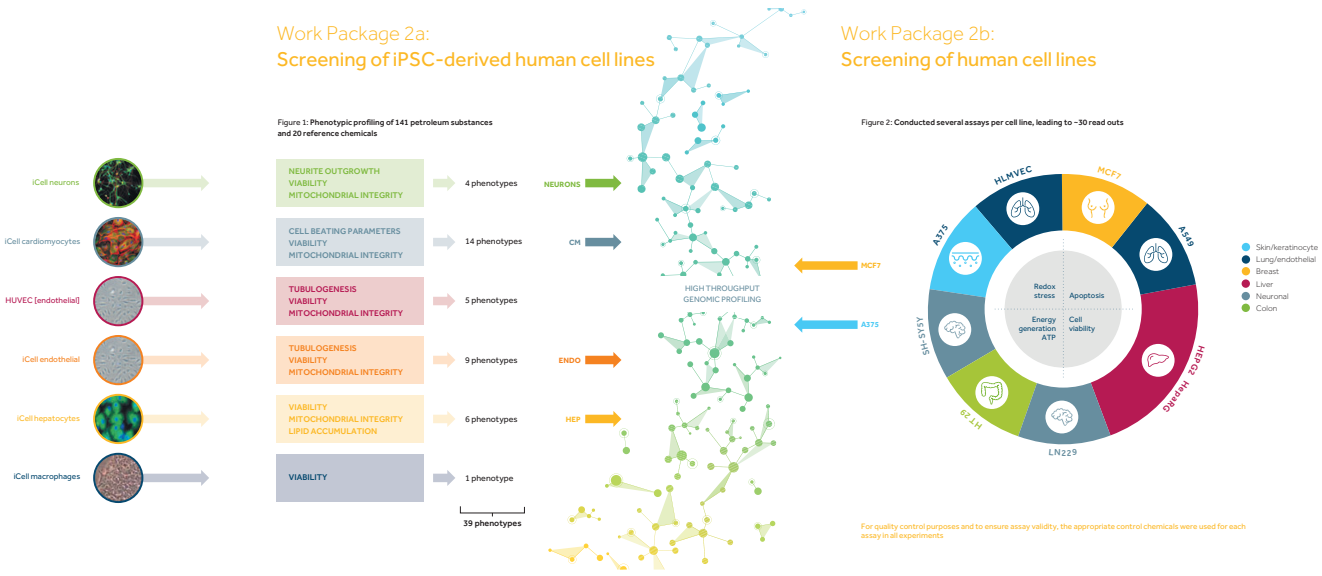
In some cases, closely related substances (PS-E) from different Concawe categories were grouped together solely for statistical and display purposes:

- HRBO was combined into OLBO (BO);
- slack waxes and paraffinic waxes were combined (WAX);
- bitumens were combined with the single substance oxidised asphalt (BIT); and
- a single MK1 was grouped with kerosine (KER).

Subsequently, in WP2, selected cellular systems were exposed to the PS-E of all 141 substances to measure their biological responses, or bioactivity, in these cell models. Cell type and vendor selections were based on the following considerations. Cells had to be of human origin and represent diverse organs/tissues. A number of more conventional, established cell models were used, as well as 'primary' cells, and so-called induced pluripotent stem cell (iPSC)-derived cells which are novel cell models that are more biologically active. All in-vitro models had to be reproducible (i.e. a particular cell/donor can be obtained from a commercial source) and suitable for the evaluation of both 'functional' and 'cytotoxicity' endpoints to enable an assessment of the specificity of the effects of test compounds. It was considered to be more important to have a strong screening assay which delivers consistent and reproducible responses regardless of its toxicological functionality, rather than have a model with a strong biological relevance, as the aim is to support grouping of substances based on a consistent response and not to predict the toxicological effects of a substance. Figure 2 provides an overview of all cell models used in Cat-App, and their in-vitro assays from which the biological response data were generated, i.e. bioactivity monitoring.

*In WP2, selected human cell systems were exposed to the PS-E of all 141 substances to measure their biological responses, or bioactivity.*

**Figure 2: Human cell lines used for Cat-App bioactivity monitoring**



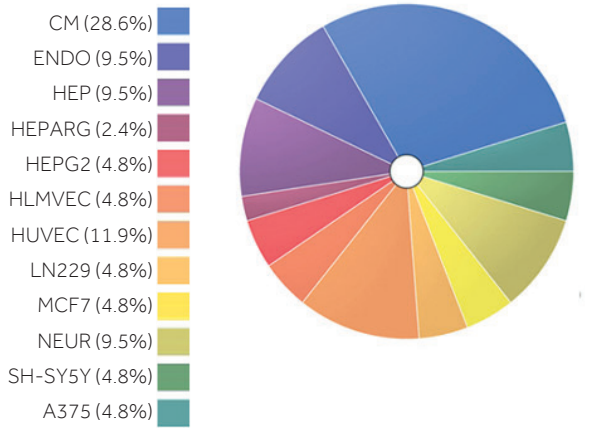

As can be seen in Figure 2, a number of assays which performed best in the biological monitoring experiments were subsequently selected for 'high content genome profiling'. This gene expression profiling, conducted in WP3, investigates the activity of genes, i.e. which genes are turned on or off in these cell systems in response to chemical exposure. It generates further mechanistic understanding behind the biological activity observed from the in-vitro assays. As will be further clarified later in the article, this additional mechanistic information is important for building further confidence in the generated in-vitro data, and as additional evidence in building grouping and read-across hypotheses.

The eventual application of chemical-biological grouping and read-across will require an integrated analysis of all the generated data. For this purpose, the statistical and visualisation tool called ToxPi[7] (Toxicological Prioritisation Index) was used. After running a quality control and uncertainty analysis on all data,[6] WP4 proceeded to run the tool on all assays which passed this control step. The principle behind ToxPi is that all assays (i.e. all biological activity measurements) in one cell model are grouped together in one slice of a pie chart, with weighting in proportion to the number of assays conducted per cell model, as shown in Figure 3.

*The gene expression profiling, conducted in Work Package 3, investigates the activity of genes, i.e. which genes are turned on or off in these cell systems in response to chemical exposure, which adds further mechanistic insights into the observed biological responses.*

**Figure 3: ToxPi construction with bioactivity data**

Note: acronyms refer to the cell types shown in Figure 2 on page 71.




Data from 43 assays across 12 cell lines were used to construct a ToxPi for each tested substance. The relative contribution of each cell type is shown on the pie chart.

The result is that each tested PS-E obtains its own bioactivity profile in the form of a ToxPi, integrating all data types (i.e. all cell models and assays). In addition to this overall integrated analysis, which thus compares overall bioactivity from all cell models across all petroleum substance categories, analyses of bioactivity profiles per cell model were also compared between the categories. In this case, the ToxPis for each substance tested were constructed with the pie reflecting a cell type, and each slice reflecting one assay type conducted on that specific cell model (no example is shown here but see, for example, Figure 5 in the next section).

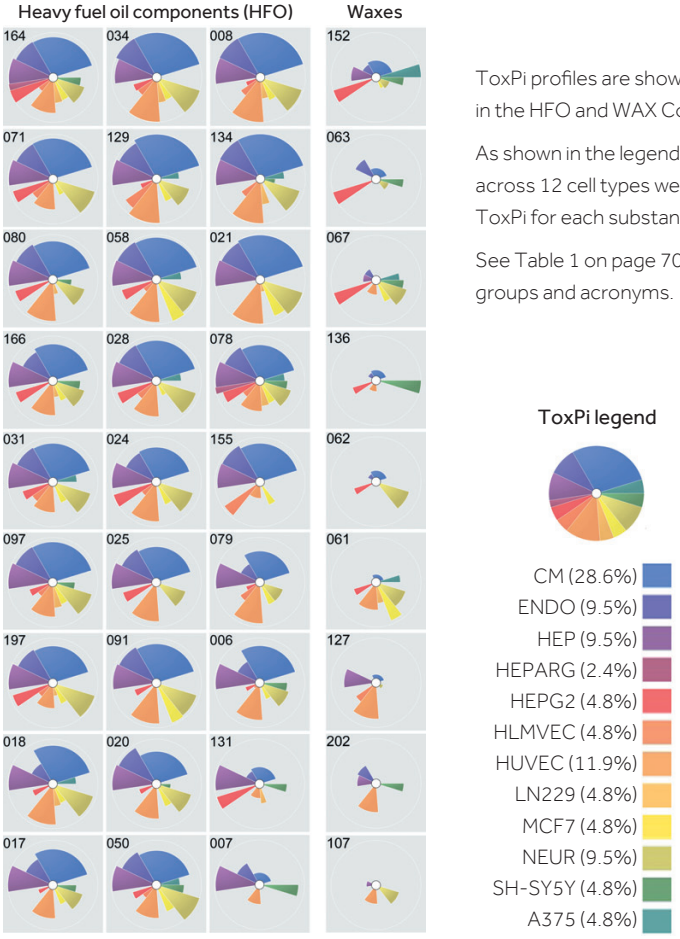
*Each tested PS-E obtains its own bioactivity profile in the form of a ToxPi, integrating all data types. Substance-specific ToxPis are scored and ranked into the biological (Fig. 5) and chemical-biological (Fig. 6) spaces.*

[6] Results of the quality control are not shown here, but can be found in the Cat-App report[8] which is available online at: www.concawe.eu/cat-app. In brief, 13 of the 15 cell lines used in the experiments were deemed to be of acceptable quality for further analyses. Data from 43 assays conducted in these cell lines were used in further analyses.

## Bioactivity-based grouping of petroleum substances

Once every tested substance, or PS-E, has its own bioactivity profile, these can be compared across the different categories of petroleum substances. The hypothesis here is that, globally, substances within one category will have similar bioactivity profiles, as they are chemically similar, while they will be different between different categories of petroleum substances. Figure 4 shows the bioactivity profiles of the individual substances tested in the heavy fuel oils (HFO) and waxes (WAX) categories as an example.

**Figure 4: Supervised grouping of petroleum UVCBs based on the bioactivity profiling data (i.e. bioactivity data were grouped based on the existing Cat-APP categories)**




ToxPi profiles are shown on the left for substances in the HFO and WAX Concawe categories.

As shown in the legend below, data from 43 assays across 12 cell types were used to construct a ToxPi for each substance tested.

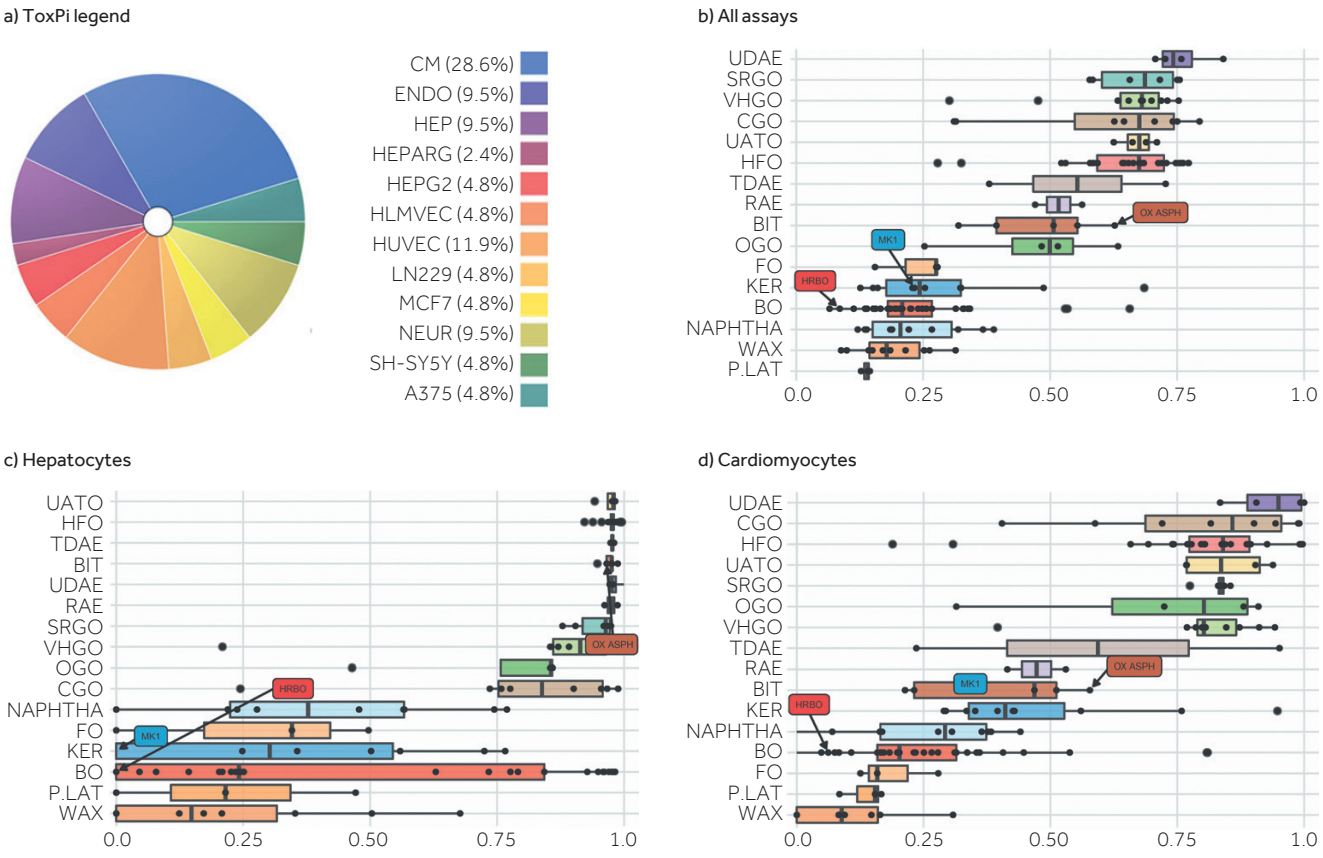
See Table 1 on page 70 for Cat-App-specific groups and acronyms.




In this example, it can be concluded visually that, for the HFO category, most of the PS-E exhibited very similar ToxPi profiles across all cell types, indicating an overall similarity in bioactivity (the left panel on Figure 4). Very different ToxPi profiles from those observed in the HFO category are apparent for the WAX PS-E (right panel on Figure 4). However, some variability among substances in each of the two categories displayed is also apparent. For example, two PS-E in the HFO category (bottom right) are quite different in the observations on cardiomyocytes (blue slice) and other cell types.

To further investigate this observed variability within a category, and to compare the bioactivity-based groups between each other, each ToxPi is scored, ranging from low (0) to high (1) bioactivity. Based on this scoring, all substances can then be ranked and compared to each other, as shown in Figure 5.

Figure 5: ToxPi analysis of the bioactivity data




Notes:

a) Data from 43 assays across 12 cell types were used to construct a ToxPi for each tested substance. The relative contribution of each cell type is shown here.

b) ToxPi scores based on all data for each Concawe category are shown as a box-and-whiskers plot, ranked from low to high bioactivity.

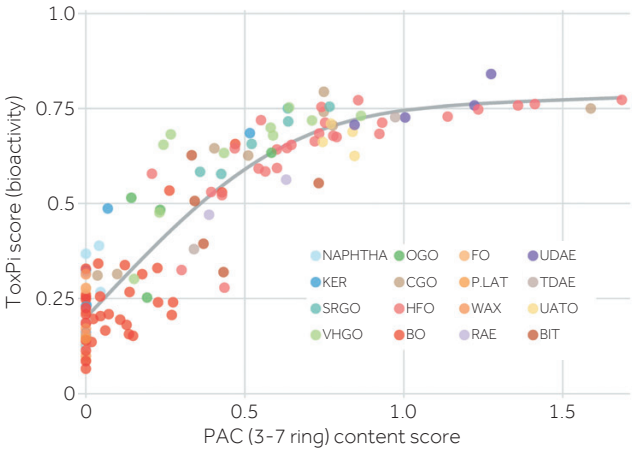
c–d) Separate ToxPi analyses were performed on the data from hepatocytes (i.e. liver cells; 5 assays) and cardiomyocytes (i.e. heart cells; 12 assays). See Table 1 on page 70 for an explanation of the acronyms and Cat-App-specific groupings. Individual substance data are presented in the supplemental material to House *et al.*, 2021.[9]

It is immediately visible that a clear gradient can be observed among the petroleum substance categories (Figure 5b). Overall bioactivities are higher for substances in the HFO, gas oils and aromatic extract categories, compared to, for example, petrolatums and waxes. This trend is well aligned with what is known from decades of toxicology testing on petroleum substances in our industry; the categories that exhibit higher bioactivity, now observed on current samples of petroleum substances, are the ones that are classified for specific hazards based on historical animal test data. The other side of the range can be explained similarly: categories which show almost no bioactivity in the current test samples are the ones that are not classified based on existing toxicological data. This separation is even more apparent when we zoom in to specific cell types: liver cells, shown in Figure 5c, are able to strongly separate the categories of substances into two broad bioactivity regions, whereas the cardiomyocytes (Figure 5d) show a gradient among the categories at the lower bioactivity spectrum. It is clear that, while a gradient of bioactivity exists between the Cat-App categories, there is also an appreciable degree of variability in bioactivity within each category. One explanation is that certain categories of petroleum substances were grouped together for statistical visualisation purposes (see Table 1 on page 70). An example is the merging of highly refined base oils (HRBO) with lubricant base oils (LBO). The HRBO is highlighted in the little red box in Figure 5, and it is obvious that these substances are at the very low end of the bioactivity spectrum. In addition, due to the inherent nature of petroleum substances (they are UVCBs) and due to the physical chemistry of refining, it is expected that these substances will form a continuum in the hydrocarbon space, i.e. they cannot be strictly separated by analytical boundaries. Based on this chemical overlap between categories, and the chemical variation within them, the observed overlap and variability in bioactivity can also be explained.

The observed overall trend is an important finding, and adds new current data to the weight of evidence and historical knowledge on petroleum substances. In addition to providing further confidence in these historical data, another hypothesis can be tested: it is known from the existing toxicology data on petroleum substances that observed effects are mostly driven by the levels of specific constituents in these substances — namely 3-7 ring polycyclic aromatic compounds (PACs). Can the observed variation in bioactivity be explained by the variability in 3-7 ring PAC content of each substance? To examine this, the relationship between bioactivity of the substances and the 3-7 ring PAC content in each substance was evaluated — see Figure 6 on page 76.

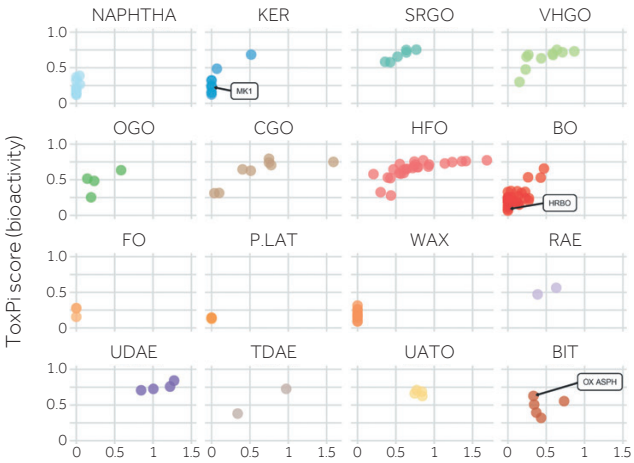
**Figure 6: Relationships between the bioactivity-based ToxPi scores of PS-E and the PAH 3-7 ring content score of the petroleum UVCB used in Cat-App**

**a) Overall correlation plot**




a) The chart shows the overall correlation plot with all substances included. The X-axis is the 3-7 ring PAC content score that was calculated by taking the sum of aromatic ring content (for 3 ring-through 7 ring-containing constituents) times the percent total weight of DMSO-extractable PACs determined by the PAC-2 Method. The Y-axis is the cumulative ToxPi score of each substance based on the bioactivity in 13 cell lines. Each substance is marked by a colour that corresponds to Concawe Cat-App-specific categories.

**b) Separate plots for each Cat-App-specific category**




b) The charts show the same information as above, but each plot contains the substances for a Cat-App-specific category. Note: for statistical visualisation reasons the Concawe categories were merged into 16 classes shown here by 16 colors. Subcategories are noted in the white boxes: MK1 (in KER), HRBO (in BO) and oxidised asphalt (in BIT). For further explanation of Cat-App-specific acronyms and groupings refer to Table 1 on page 70. See the supplemental material in House *et al.*, 2021[9] for cell-specific correlations.

Consistent with the hypothesis, as can be clearly seen in Figure 6, the overall fit for the ToxPi scores based on the bioactivity data from all 13 cell types showed a strong positive correlation (Spearman rho=0.89) with the 3-7 ring PAC content of each substance. The strong overall trend observed is that the higher the level of these constituents is in a substance, the higher the overall bioactivity; e.g. high 3-7 ring PAC-containing substances such as HFO or untreated distillate aromatic extracts (uDAE) show high bioactivity overall. On the other hand, these trends were not observed for the higher-refined substances, which contain low to negligible levels of 3-7 ring PACs, such as HRBO (highlighted in the LBO category), petrolatums, foots oils and waxes. In addition, clear trends can be observed not only overall but also within categories, but these data add further evidence that this can be explained by the variation in chemical composition of the substances even within petroleum substance categories. Overall, the results presented in this Figure corroborate the known relationship between the content of PACs, especially of the 3-7 ring type, in the petroleum refining products with their potential health hazard.

One important note is that these data cannot be interpreted as a quantitative indicator of the human health hazards of a substance. The ToxPi scores and ranking are helpful indicators of observed trends in the global biological response of substances between and within categories, in strong correlation with their (variable) analytical composition. This will not predict a human health hazard endpoint directly, but will help in underpinning the grouping of substances and read-across assessment in an overall integrative testing strategy, which maximises the efficient use of animals needed for toxicological assessments of petroleum UVCBs and reduces the overall time to complete the regulatory assessment of petroleum substances. Grouping and read-across hypotheses are needed to facilitate this, and a further understanding of the biological mechanisms underpinning the global bioactivity trends observed so far can help to build these.

## Mechanistic underpinning of the bioactivity-based grouping of petroleum substances

To obtain this mechanistic information, gene expression changes were investigated in selected cell models which were exposed to the PS-E. Gene expression analysis, also called 'genomic analysis' or 'transcriptomics', investigates the activity of the genome (genes) in cells in response to chemical exposure, i.e. which genes are turned on or off and how strongly; this provides insights into how the biology works, and how it leads to the specific effects observed in earlier experiments. Of all cell models used in the bioactivity experiments, five were selected for genomic analysis based on the following criteria:

i. cells that have passed quality control analyses for bioactivity;
ii. cells that represent a diverse set of human tissues and/or organs; and
iii. priority was given to human iPS cells as these are (proven to be) biologically more active than the conventional cell models. As shown in Figure 2 on page 71, this led to the selection of four i-cell and two human cell line models for genomic analysis.

To get an initial idea of the gene expression activity across the different categories of petroleum substances, the transcriptomic data from all cell models were combined per substance and compared between substance categories. No obvious group-specific effects could be observed (data not shown here; see Concawe Report 24/20[10]). One explanation for this is that even the PS-E tested here contain a large number of constituents which all trigger the expression of various genes, and global genomic activity alone (i.e. just the number of affected genes in response to PS-E exposure) will not, therefore, be a good discriminator. However, when the category comparisons were conducted per cell model, more pronounced separation between categories could be observed in liver cells (Concawe, 2020). Since this effect was again similar to the effects observed in the bioactivity experiments, and correlated strongly with 3-7 ring PAC content, this was a first indicator of the biological mechanisms that will likely be key.

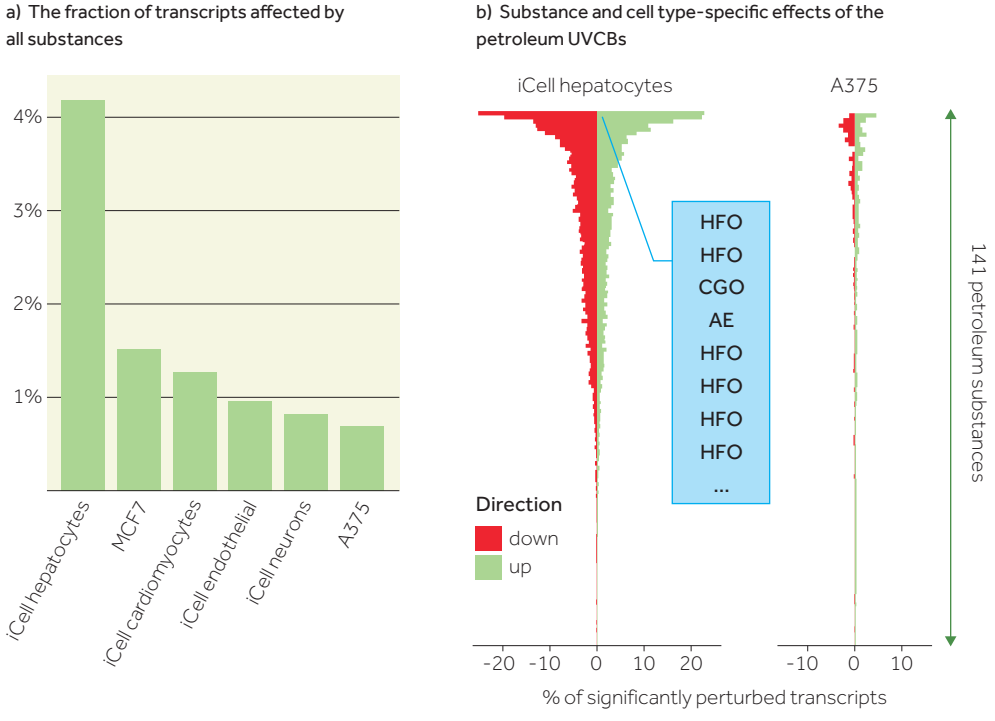
# Cat-App: learnings from a multi-year research programme on alternatives to animal testing

To further investigate this, the gene expression activity was compared between cell models to compare tissue and substance-specific effects. In addition, a more detailed analysis was conducted to understand which specific genes are affected and in which biological pathways (mechanisms) they are involved. As can be seen in the transcriptomic data shown in Figure 7, liver tissue is most responsive to exposure with PS-E.

The most probable explanation for this is that liver tissue is more metabolically competent than other tissue. This is confirmed by the fact that PS-E with the highest PAH 3-7 ring content have elicited the most pronounced effects on gene expression. In addition, the functions of the genes and the biological mechanistic pathways in which they are involved all relate to metabolic processes and, specifically, to the metabolism of PACs.

Taken together, this information further underpins the PAC hypothesis for petroleum substances, namely that the level and type of polycyclic aromatic hydrocarbons drive the observed biological responses in human tissue.

**Figure 7: Transcriptional effects of DMSO extracts of petroleum UVCBs (PS-E) on gene expression in six cell types**




Notes:

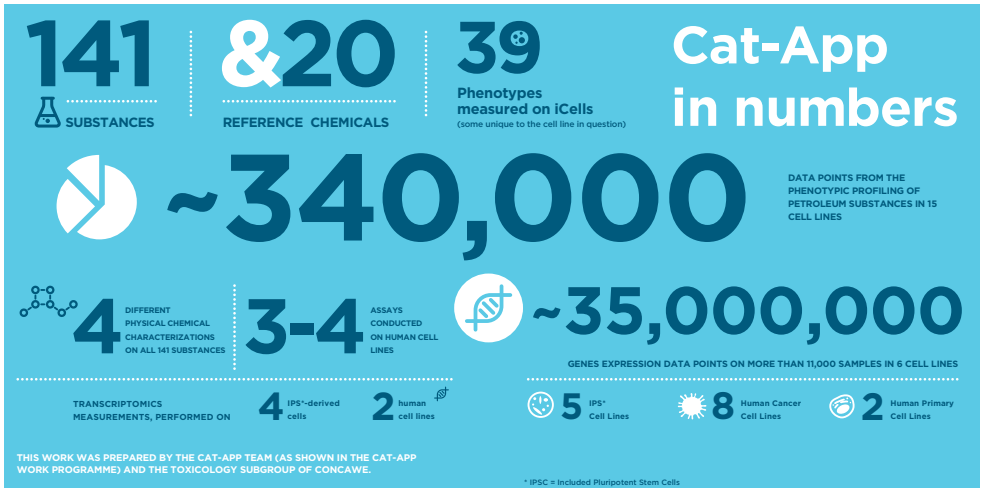
a) Near right: for each cell type, ~3,000 transcripts were evaluated across all 141 substances. For example, 4% represents the proportion of differentially expressed genes in the i-cell hepatocytes.

b) Far right: the i-cell hepatocytes and A375 cells, which represent the cell types that had the most and least pronounced UVCB-induced transcriptional effects, are shown as examples where substances are ranked by the total number of transcripts significantly affected by treatment. Colours represent the directionality of change. The top eight substances (indicated by their Concawe category) are shown in the insert for hepatocytes.

## Addressing the 'twin-challenge'

After five years on this journey, Concawe has generated an enormous database with valuable biological information on its petroleum substances (Figure 8).

Figure 8: Infographic of the Cat-App work programme output




All data were published in Concawe Report number 24/20[10] in December 2020, and the initial part of the data was published in a peer-reviewed paper that made headline news in a renowned journal on alternatives to animal testing.[9] To present the scientific progress along the way, the project team organised annual meetings and multiple successful workshops, and presented the work in numerous lectures at international fora. The feedback on the programme, from the scientific community and key stakeholders, has been positive and constructive without exception. However, Concawe regrets that this scientific approach has not yet been accepted by the regulators, which undermines its impact and does not allow the potential benefits for human health testing and animal welfare that the work could bring.

The main aim of this project was to develop a framework which would be directly applicable to address the twin challenge: firstly, the need for animal testing is increasing as this testing strategy is still the default under REACH and, partly because of this, even the solid scientific evidence justifying the use of alternatives to animal data under REACH is proving difficult to get accepted. Secondly, the available alternatives under the regulation are not practically applicable to UVCBs, and this provides an additional challenge for petroleum substances.

It is therefore clear that petroleum substances, and UVCBs in general, warrant an additional approach addressing the specific need to underpin grouping and read-across, which moves away from the standard requirements based on molecular constituents. The main opportunity to make read-across work for complex substance such as UVCBs is to move towards an approach that is group-based and more holistic.

# Cat-App: learnings from a multi-year research programme on alternatives to animal testing

The following key learnings can be derived from the Cat-App programme:

1. The first learning from Cat-App is that, with these data, we can indeed add a biological component to the similarity argument to facilitate the grouping of similar substances. This is key, as manufacturing data, refining history or phys-chem/analytical parameters alone will not be sufficient to prove this grouping concept in the context of addressing the needs for hazard assessment of these substances. It also shows, now from a biological perspective, that petroleum substances form a continuum of substances without hard boundaries between groups.
2. Secondly, the bioactivity strongly correlates with the analytical composition of the substances, in particular the level and types of polycyclic aromatic hydrocarbons, which further underpins the PAC hypothesis for petroleum substances. This learning helps to add further granularity into the grouping exercise, helps to build read-across hypotheses and aids the selection of substances to be tested in the required animal studies.

Overall, it now shows, in multiple dimensions, that petroleum substances form a continuum in the hydrocarbon-biological space, while at the same time chemical-biological trends can be observed across and within the different categories of petroleum substances. This can help to address particular issues with UVCBs regarding unknown constituents and variability of substances, as further animal testing across a wider hydrocarbon space (holistic vs substance by substance approach) can be better targeted and prioritised. On the other hand, it could also help to justify where additional testing might not be immediately prioritised, and can perhaps eventually be addressed by other means in a weight-of-evidence approach applying all relevant available in-vivo and in-vitro data. This has the potential to significantly reduce animal testing, while not underestimating potential hazards, by adding additional biological information into the assessment.

But this is not the end of the journey. The main open issue is that much of this analysis is built around the PAC hypothesis, and from early interactions with authorities on these data, the question is being raised as to how we prove that we are indeed assessing the (biologically) relevant fraction of the substance, and that the remainder of the substance is not relevant in the eventual hazard assessment context. Within that hazard context, it has been deemed from a regulatory perspective that a predictive aspect in the analysis remains absent. However, it should be stressed again that the aim of the project is not to develop an alternative method to replace animal testing at this stage; in other words, Concawe is not aiming to predict toxicity (hazard). These points might not necessarily be a problem when considered purely from a grouping perspective. Nevertheless, they should be addressed when building read-across hypotheses to prove that any potential hazards are not overlooked. Concawe is currently working on a similarity approach to ensure that the entire chemical space of a particular group or groups of substances is assessed. At the same time, work is ongoing to provide further evidence that, for the PS-E, the relevant parts of the substances are tested in the in-vitro assays.

Concawe now has a major opportunity, as part of the organisation's human health strategy, to run Cat-App and other new approach methodologies (NAM) data on samples collected from the animal testing programme. This is being undertaken now, and will provide in-vitro data alongside the animal data from the same experiments, and can therefore lay the foundation for the development of true alternative screening assays. In addition, Concawe has completed, and is initiating, other NAM projects with different academic partners, including non-animal based in-vitro models that target specific hazard endpoints. Many data are already available, hence this is a critical point for Concawe to continue in its journey to make the best use of all of these data. Integration is key, as the combined data will help to further address the issues raised above, as well as helping to further develop these NAM approaches as acceptable alternatives under REACH.

Part of the reason why Concawe still finds itself on a bumpy road to the next stage in this project is that, at the moment, the data have not yet been formally evaluated in our REACH dossiers. As indicated earlier, the authorities have only seen part of the programme and have not yet reviewed the full published data. It is vital that the authorities and industry familiarise themselves with these types of data in a regulatory setting, to enable further development and progress towards the full replacement of animal testing in the longer term. The early applications described above should allow this and, in principle, ECHA should support this paradigm, being that one of the three main aims of the REACH legislation is to 'promote alternative methods for the hazard assessment of substances in order to reduce the number of tests on animals.[11] Hence, it is now of critical importance to have the outcome of this project included into the dossiers.

There have been valuable learnings from this project so far, but the journey continues. It is vital that Concawe continues to push for the opportunities provided by REACH and, by extension, by the EU's Chemical Strategy for Sustainability (CSS), to put into practice the vision for toxicology in the 21st century and ensure that the necessary progress is made.

*The animal testing programme that Concawe is conducting as part of its human health strategy for REACH compliance provides an outstanding opportunity to further develop in-vitro assays, as well as the Cat-App framework, which should eventually lead to a more sustainable testing and assessment paradigm.*

*Special thanks go to the various teams at Concawe, especially the Toxicology Subgroup, as well as Concawe's partners involved in the research consortium (see Figure 1 on page 69), for their innovative science and excellent insights over the past years.*

## References


1. Concawe (2019). *REACH: Roadmap for Petroleum substances*. https://www.concawe.eu/wp-content/uploads/REACH-Roadmap-for-Petroleum-Substances-2019-update.pdf
   Note that the figures published in this document have been updated for the current article based on the most recent number of registered substances and cost estimates (Spring 2021).
2. Regulation (EC) No 1907/2006 of the European Parliament and of the Council of 18 December 2006 concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH), establishing a European Chemicals Agency. https://osha.europa.eu/en/legislation/directives/regulation-ec-no-1907-2006-of-the-european-parliament-and-of-the-council
3. NRC (2007). *Toxicity testing in the 21st century: A vision and a strategy*. Committee on Toxicity Testing and assessment of Environmental Agents, National Research Council. Washington, DC. National Academies Press. www.nap.edu/catalog/11970.html
4. Russell, W. M. S. and Burch, R. L. (1959). *The Principles of Humane Experimental Technique*. Methuen, London. https://www.worldcat.org/title/principles-of-humane-experimental-technique/oclc/595267154
5. Concawe (2016). Concawe *Review*, Volume 25, Number 2. https://www.concawe.eu/publication/concawe-review-25-2/
6. ASTM (2010). Standard Test Method for Determining Carcinogenic Potential of Virgin Base Oils in Metalworking Fluids. E1687-10. ASTM International, West Conshohocken, PA, 2010. https://www.astm.org/DATABASE.CART/HISTORICAL/E1687-10.htm
7. Marvel, S. W., *et al.* (2018). 'ToxPi Graphical User Interface 2.0: Dynamic exploration, visualization, and sharing of integrated data models.' In BMC Bioinformatics, Vol. 19, Article number 80. https://bmcbioinformatics.biomedcentral.com/articles/10.1186/s12859-018-2089-2
8. Concawe (2021). 'Cat-App Project' (website). https://www.concawe.eu/cat-app/
9. House, J. S. *et al.* (2021). 'Grouping of UVCB substances with new approach methodologies (NAMs) data.' In *ALTEX*, Vol. 38, No. 1. https://www.altex.org/index.php/altex/article/view/1994
10. Concawe (2020). *Cat-App: New Technologies to Underpin Category Approaches and Read-across in Regulatory Programmes*. Concawe Report 24/20. https://www.concawe.eu/wp-content/uploads/Rpt_20-24.pdf
11. ECHA (2021). 'Understanding REACH' (website). European Chemicals Agency. https://echa.europa.eu/regulations/reach/understanding-reach