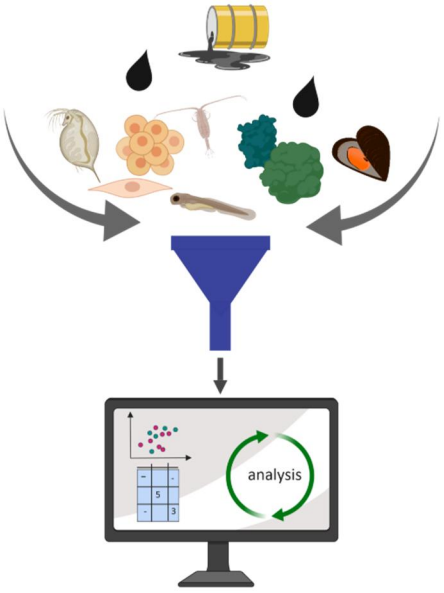
GRACE grant no 679266

Fingerprinting effect-based toolbox

Definition of an effect-based toolbox (bioassay battery) for high-throughput cost-effective investigation and fingerprinting of oil contamination and environmental risk assessment

D3.9

WP3: Oil impacts on biota using biomarkers and ecological risks assessment



Prepared under contract from the European Commission
Contract n° 679266
Research and Innovation Action
Innovation and Networks Executive Agency
Horizon 2020 BG-2014-2015/BG2015-2

Project acronym: GRACE
Project full title: Integrated oil spill response actions and environmental effects
Start of the project: 01 March 2016
Duration: 42 months
Project coordinator: Finnish Environment Institute (SYKE)
Project website: http://www.grace-oil-project.eu

Deliverable title: Fingerprinting effect-based toolbox
Deliverable n°: D3.9
Nature of the deliverable: Report
Dissemination level: Public

WP responsible: WP3
Lead beneficiary: RWTH

Due date of deliverable: 31.07.2019
Actual submission date: 31.07.2019

Deliverable status:

Version	Status	Date	Author	Approved by
1.0	draft	16.07.19	Sarah Johann, Thomas Benjamin Seiler	WP3 members 23.07.2019
2.0	final	23.07.2019	Sarah Johann, Mira Goßen, Henner Hollert, Richard Ottermanns, Xabier Lecube, Ionan Marigómez, Laura de Miguel Jiménez, Alberto Katsumiti, Tamer Hafez, Aino Ahvo, Kari Lehtonen, Tomasz Maciej Ciesielski, Björn Munro Jensen and Thomas-Benjamin Seiler	Steering group 31.07.2019

Table of Content

Executive Summary....5

I Introduction....7

II Sample background and WAF preparation....8

II.1 Sample background....8

II.2 Preparation of water-accommodated fractions....9

III Material and Methods of individual bioassays....10

III.1 Acute toxicity in selected aquatic vertebrates and invertebrates....10

III.2 Mechanism-specific toxicity using *in vitro* based methods....11

III.2.1 Cell Viability examination using MTT bioassay....11

III.2.2 Oxidative stress response in U2OS cells using Nrf2-CALUX® assay....11

III.2.3 Genotoxicity using Micronucleus assay....13

III.2.4 Mutagenicity using Ames fluctuation assay....14

III.2.5 Endocrine disruption using ERα-CALUX® assay....15

III.2.6 Endocrine disruption using A-YES® assay....17

IV. Material and Methods of toxicity profiling....18

IV.1 Classification system....19

IV.2 Principal component analysis (PCA)....23

IV.3 Stochastic resampling for PCA....23

V. Results of individual bioassays....24

V.1 Acute toxicity in selected aquatic vertebrates and invertebrates....24

V.2 Mechanism-specific toxicity....24

V.2.1 Cell Viability examination using MTT bioassay....24

V.2.2 Oxidative stress response in U2OS cells using Nrf2-CALUX® assay....25

V.2.3 Genotoxicity using Micronucleus assays....25

V.2.4 Mutagenicity using Ames fluctuation assay....26

V.2.5 Endocrine disruption using ERα-CALUX® assay....26

V.2.6 Endocrine disruption using A-YES® assay....26

VI. Toxicity profiling....28

VI.1 Toxicity profile of Naphthenic North Sea crude oil....28
VI.1.1 Toxicity profile of untreated and chemically dispersed NNA WAFs....28
VI.1.1 Dispersant influence on toxicity....30
VI.1.3 Validation of PCA for toxicity profiling by stochastic resampling....33
VI.2 Toxicity profiles of refined petroleum products....33
VI.3 Comparison of different oil types....33
VII Suggestion of a sensitive, cost- and time efficient bioassay battery....34
VIII Critical considerations for data interpretation....36
IX Conclusion and outlook....38
X References....39
XI Appendix....45

Executive Summary

In WP3 of the GRACE project a comprehensive battery of bioassays has been applied to investigate the toxicity of crude and refined petroleum products (NNA crude oil, MGO, IFO180). Additionally, two third-generation chemical dispersants (Finasol OSR 51 and 52) were included to evaluate their influence on the petroleum products' toxicity. Within this the bioanalytical tools covered different biological organization levels: on the one hand the acute toxicity on whole organisms including invertebrate and vertebrate laboratory model and regional relevant species was investigated. On the other hand, also mechanism-specific *in vitro* based methods on different modes of action were included. Within this framework of the GRACE project a set of methods standardized in ISO or OECD guidelines were optimized for oil toxicity testing. Bioassays are useful screening tools that complement chemical analysis for water quality assessment as they do not require *a priori* information on the identity and physical-chemical properties of contaminants and cover mixture toxicity of complex environmental samples. The major goal of the present report was to suggest a petroleum product-specific assay battery that provides cost- and time-efficient bioanalytical tools being sensitive endpoints for the risk assessment of oil contamination. Furthermore, as each oil is a unique and complex sample the study addressed the question whether the exposure to different petroleum product types leads to specific distinguishable toxicity profiles.

Most biological effect data were available for the NNA crude oil with > 20 different endpoints further extended with data on different salinity conditions, temperature regimes, concentration ranges and time windows of exposure. Additionally, the refined petroleum products MGO and IFO180 were investigated in a set of corresponding endpoints, depending on suitability of the endpoint, and available resources within the project. For the comprehensive analysis effect data were first divided in categories from acute toxicity over genotoxicity, endocrine disruption, neurotoxicity or biotransformation activity to more general stress markers including cytotoxicity and oxidative stress response. Effect data were then transformed in a shared classification system to rescale the different responses into one common scale. A principal component analysis (PCA) was used to statistically evaluate the multivariate approach combining the different bioassays. The analysis was able to distinguish between more and less sensitive bioassays on oil toxicity testing.

Based on available data, no individual toxicity profiles were found for the different oil types, as they seem to have identical modes of actions. However, additional data on the refined petroleum product should be included to evaluate the big picture of a toxicity profile.

A sensitive petroleum product toxicity profiling toolbox is recommended to contain endpoints on both acute and mechanism specific toxicity. In detail, the regionally relevant copepod *C. finmarchicus* as well as early life stages of fish (*D. rerio*) were sensitive towards the WAF

exposure. To investigate the genotoxic potential of WAFs the micronucleus induction in fish cells should be preferred over the mutagenic potential in bacterial strains. Interestingly, also the endocrine disruptive potential was found to be a relevant endpoint for untreated and dispersed petroleum products investigated in cost and time efficient small-scale receptor mediated *in vitro* assays. Furthermore, also typical biomarkers for hydrocarbon contamination as for example the CYP activity (e.g. in fish embryos) were highly sensitive and therefore suggested for an oil-specific bioassay battery. The oxidative stress induction investigated either via the high throughput *in vitro* based method (Nrf2-CALUX®) or the low-throughput biomarker responses in mussel tissues (CAT, LPO) was an equally sensitive and relevant endpoint category.

Further research should focus on additional endpoints, for example on biomarker responses in the highly sensitive and relevant species *C. finmarchicus* in order to contribute to an oil toxicity profiling battery. Endpoints presented for NNA should be finalized for the refined petroleum products and additional experiments (e.g. with adult zebrafish, stickleback and medaka embryos) could contribute to the big picture of the toxicity profile.

I Introduction

Chemical analysis is often described as the primary method to assess the hazard potential of environmental samples by verifying chemical compounds. Verifying a complex sample presupposes knowledge about the presence of contaminants and their toxicity. Additionally, observed adverse effects on biota often cannot be explained by a simple addition of single compounds toxicity due to mixture toxicity covering synergistic or antagonistic interactions as well. Bioassays do not require *a priori* information on the composition of a complex environmental sample and physical-chemical properties of the individual compounds. Depending on the organization level of the bioassays they furthermore provide mechanistic insight into the mode of action of the observed toxicity. Bioanalytical tools can therefore complement chemical analysis for water quality assessment (Escher et al. 2013). In light of refinement, reduction and replacement of animal testing (3R principle), a large variety of time and cost-efficient cell-based (*in vitro*) bioassays have been optimized. However, cellular responses will not always imply higher-level effects and should not be interpreted as such. Environmental samples can cause toxic effects on different biological levels, where interrelation and connection of these levels is very complex but simultaneously important for the understanding of environmental reaction. Hence, cell-based bioassays function as a first tool in the characterization of the toxic action of chemicals and environmental samples (Fent 2001).

In view of the above explained advantages of bioanalytical tools, a set of methods of which some are standardized in DIN EN ISO norms or OECD guidelines and included in the risk assessment of chemicals or environmental samples were adapted to oil toxicity testing. Besides *in vitro* cell based methods also whole-organism tests with laboratory model and study region relevant species were used.

Within the GRACE project three different oil types varying in their degree of processing from crude to distillate, and hence also varying in their physical-chemical properties were selected. As each oil is a unique and complex sample the resulting toxicity to the exposed biota could also deviate for different oil types. Thus, for risk assessment of an oil spill or contamination it is important to define unique toxicity profiles. The importance of individual profiles is emphasized even further by requirements for rapid decisions on the best response actions in case of an oil spill. In order to limit the environmental impact, decision makers need scientific input not only about the affected region but also about the oil toxicity for processing a net environmental benefit analysis. In a literature review on the aquatic toxicology of petroleum oil it was recommended to compare the toxicity of different oil products and, in case they show unique toxicity, to determine the mechanisms of effects (Dupuis and Ucán-Marín 2015).

Besides different oil types also two third-generation chemical dispersants were integrated in the GRACE project to evaluate their influence on the petroleum products toxicity. Third generation dispersants are reported to be less toxic and more efficient compared to earlier generations developed during the 1960s to late 1980s (Grote et al. 2016). Dispersants have been used on a

large scale in many oil spill response actions in the last years (Prince 2015). They lower the interfacial tension between the oil and the water resulting in breaking down the oil slick on the surface and transporting oil micelles in the water column. However, the benefit of the application is controversially discussed.

The present report addresses the question whether the selected three different oil types with and without the application of dispersants lead to distinguishable toxicity profile. Within this a major benefit is that the toxicity of different petroleum products was evaluated using comparable WAF preparation procedures and the same bioassay protocols. The difficulty of comparing different oil toxicity studies based on various WAF preparation procedures or bioassay protocols is a huge shortcoming in respect to the interpretation of petroleum product toxicity data (Redman and Parkerton 2015). Furthermore, the data of the present report on the acute and mechanism-specific toxicity of the individual oil types can provide useful information for the Environment and oil spill tool (EOS) developed in WP 5 of the GRACE project.

Against the background of an extended set of comparable biological effect data one major challenge is to interpret all the different endpoints in one common context. Hence, another focus of the present report is the suggestion and the critical discussion of a bioassay-specific classification already including expert judgement on observed responses in early steps of data processing.

Based on this context the major goal of the present report is to suggest a petroleum product-specific, useful bioassay battery that provides simultaneously cost- and time-efficient bioanalytical tools being sensitive endpoints for the risk assessment of oil contamination.

II Sample background and WAF preparation

II.1 Sample background

A naphthenic North Sea crude oil (NNA) was selected as the crude and untreated petroleum sample. It is a light crude oil with low viscosity and characterized by a high proportion of low molecular weight saturates and aromatics. A detailed chemical profile of the naphthenic North Sea crude oil was generated in the GRACE project.

In addition to the described crude oil a marine gas oil (MGO), a distillate formed during the fractional distillation of a crude oil, was used in this study. This MGO is supplemented with the green dye Dyeguard Green MC25 produced by John Hogg Technical Solutions.

Furthermore, an intermediate fuel oil (IFO 180) which is a blend of heavy fuel oil and gas oil was selected as an intermediate stage of petroleum product purity between crude and marine gas oil. The intermediate fuel oil is characterized by a high viscosity (maximum viscosity = 180 centistokes) and a sulphur content of less than 3.5 %.

To investigate the influence of dispersants on petroleum product toxicity Finasol OSR 51 and Finasol OSR 52 produced by Total were included. The selected chemical dispersants are complex mixtures of anionic and neutral surfactants and hydrocarbon solvents. Both dispersants are relevant in the study region of the GRACE project and the treatment of the selected oil types. The selected third-generation dispersants have slightly different chemical composition (see Table 1). Finasol OSR 51 contains 15-30 % non-ionic and 0.2-0.5 % anionic surfactants, while Finasol OSR 52 contains >30 % non-ionic surfactants and 15-30 % anionic surfactants. Finasol OSR 52 is compliant with all the three regulations on the market (EPA, MMO, CEDRE), while Finasol OSR 51 is compliant with two of them (MMO, CEDRE) (Total 2019). The crude oil was subject to a variety of different investigations also in other tasks of WP3. Consequently, for this oil type an expanded dataset was available for the analysis in this report. In general, while MGO and IFO 180 samples were exclusively combined with the dispersant Finasol OSR 52, the naphthenic North Sea crude oil was combined with both dispersants used in this report.

Table 1 Information on ingredients and composition of the dispersants Finasol OSR 51 and Finasol OSR 52 (Total®). *All information is based on the safety data sheets (Total 2012a, b).*

Finasol OSR 51		**Finasol OSR 52**	
chemical name	**weight %**	**chemical name**	**weight %**
hydrocarbons, C11-C14, n-alkanes, isoalkanes, cyclics, <2% aromatics	60 - 70	hydrocarbons, C11-C14, n-alkanes, isoalkanes, cyclics, <2% aromatics	15 - 20
docusate sodium	0.2 - 5	docusate sodium	20 - 25
		(2-methoxymethylethoxy)propanol	15 - 20
		carboxylic acids, di, C6-12 cmpds, with ethanolamine, boric acid cmpd with ethanolamine	0 - 2
		ethanolamine	0 - 1
non-ionic surfactants	15 - 30 %	non-ionic surfactants	> 30 %
anionic surfactants	0.2 - 5 %	anionic surfactants	15 - 30 %

II.2 Preparation of water-accommodated fractions

In general, all different water-accommodated fractions were prepared according to Singer et al. (2000). Modifications of WAF preparation in different laboratories of the GRACE project were mainly related to stock oil loading for low energy water-accommodated fractions (LEWAF) exposure (1:40, 1:50 or 1:200, see below) and dimensions of stock preparation (eg. 300 mL, 20 L)

leading to usage of different types of aspirator bottles with varying head space. Details on laboratory specific WAF preparations can be found in Table 4 (Appendix).

In the GRACE project LEWAF for oil exposure only, chemically enhanced water-accommodated fractions (CEWAF) as well as dispersed low-energy water-accommodated fractions (LEWAF+D) for the combination of oil and dispersant exposure, and high energy water-accommodated fractions (HEWAF) for dispersant exposure only were used.

Details on the different WAF preparation approaches for acute toxicity investigations in aquatic vertebrates and invertebrates can be found in the corresponding deliverables D 3.12 (AOL in zebrafish), D 3.14 (Effects on zooplankton), D 3.16 (Toxic impacts of oil spills).

As additional results on mechanism-specific toxicity endpoints (produced at RWTH University) are presented in the current report the preparation of WAFs in these small-scale experiments are described in detail in the following section:

In small-scale experiments at RWTH WAFs were prepared in aspirator glass flasks (500 mL) by application of oil or a dispersant-oil mixture (1:10) on the surface of 300 mL artificial medium (fish embryo exposure) or ultrapure water (cell-based bioassays) at 10 °C at an oil-to-water (w:v) ratio of 1:50 (LEWAF) or 1:200 (CEWAF), respectively.

The LEWAF setup was carefully stirred with low energy avoiding a vortex in the water phase while the CEWAF was stirred at higher stirring speeds to create a 25 % vortex in the water phase. The HEWAF (dispersant alone) was prepared as described for the CEWAF stock solution with dispersant loadings corresponding to the amounts added for the CEWAF production due to ensure the comparability of the resulting stock solution. LEWAFs, CEWAFs and HEWAFs were incubated stirring at 10 °C for 40 h and followed by 1 h settling time. Afterwards, water fractions were carefully drained off.

After draining off the water phase dilution series prepared from the 100 % stock solutions (1:50 LEWAF, 1:200 CEWAF, 1:200 HEWAF) were prepared for each bioassay. In general, the WAF samples were treated as water samples in the different small-scale *in vitro* bioassays with specific modifications due to oil toxicity testing.

III Material and Methods of individual bioassays

III.1 Acute toxicity in selected aquatic vertebrates and invertebrates

The acute toxicity of crude oil and refined petroleum products in selected study region relevant and laboratory model species has been addressed in submitted reports including D 3.2 (Test conditions for zebrafish), D 3.12 (AOL in zebrafish), D 3.14 (Effects on zooplankton) and D 3.16 (Toxic impacts of oil spills). Hence, methodological details can be found there in corresponding sections.

III.2 Mechanism-specific toxicity using *in vitro* based methods

Several results of petroleum product-induced mechanism-specific toxicity small scale *in vitro* assays have not been integrated in previous submitted reports in GRACE. Hence the methods are described in detail in the following sections.

III.2.1 Cell Viability examination using MTT bioassay

The MTT (3-(4,5-Dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide) assay, first described by Mossmann et al. (1983), is a commonly used method in ecotoxicology to investigate the viability of cells that are exposed to chemicals or polluted environmental samples. Investigating a general endpoint in cell toxicity it is also an important pretest for other mechanism-specific assays, in which avoiding non-specific cytotoxic effects exposure concentrations is crucial to exclude false negative results. The MTT assay is reported as a very sensitive method (Fotakis and Timbrell 2006). The underlying principle is based on the uptake of the yellow tetrazolium salt via the plasma membrane potential due to its net positive charge (Berridge et al. 2005), followed by the reduction to the insoluble purple product formazan, which accumulates in crystals in viable cells (Gonzalez and Tarloff 2001). After lysing cells and formazan crystals, absorbance measurement quantifies the formazan content.

The MTT assay procedure of seeding, incubation and exposure regime was performed according to optimized SOPs for the individual cell lines used in the Nrf2 CALUX®, ERα-CALUX® and the Micronucleus assay.

Briefly, after cells were seeded and exposed to a 1:2 dilution series of WAF concentrations with 100 % of stock as the highest concentration the approaches were incubated at cell line specific temperatures for cell line specific timer intervals. Afterwards, the exposure medium was removed, cells were washed with PBS and yellow MTT salt (500 µg/mL) was added. After an incubation for 30 min at 37 °C the MTT solution was replaced by DMSO to dissolve the cell membranes and the formed formazan crystals on an incubation shaker for 15 min. The absorbance at 492 nm was measured using an Infinite M200 plate reader (Tecan Group AG, Männedorf, Schweiz). The intensity of absorption is proportional to the amount of viable cells. Cell viability was calculated by correcting the absorbance values first for the response of the blank values. A control absorbance of untreated cells was defined as 100 % cell viability while all sample dilutions were calculated relative to this viability.

For cell exposure in all mechanism-specific bioassays only sample concentrations resulting in at least 80 % cell viability were used.

III.2.2 Oxidative stress response in U2OS cells using Nrf2-CALUX® assay

a) human osteosarcoma U2OS cells transfected with transcription factor Nrf2

In this assay the osteosarcoma cells U2OS purchased from BioDetection Systems BV (BDS, Amsterdam, The Netherlands) have been used. Human osteoblastic U2OS cells have been stably transfected with the transcription factor Nrf2, which is involved in the activation of genes containing anti-oxidant responsive elements contributing to the oxidative stress response (van der Linden et al. 2014). Cells were cultured in a mixture of Dulbecco's modified Eagle's medium and F12 medium (1:1), which was supplemented with 7.5 % fetal calf serum (Biowest, France), non-essential amino acids and a penicillin-streptomycin solution as described in the SOP P-BDS-076 from BDS (b.v. 2017). Cells were cultivated at 37 °C with an atmosphere containing 5 % CO2. Periodically, cells were passaged when reaching 90 % confluence.

For the Nrf2-CALUX® assay a 3x-concentrated assay medium was prepared from cell culture medium powder (Sigma Aldrich, D2902), which was finally supplemented with FCS (chrcoal stripped, Biowest, France), non-essential amino acids and penicillin-streptomycin as described in ISO guideline no 19040-3 (2014). 1x-concentrated assay medium was prepared by diluting the 3x-concentrated assay medium with sterile ultrapure water.

b) Nrf2-CALUX® assay procedure

The chemical activated luciferase gene expression (CALUX®) bioassays are mechanism-specific in vitro reportergene assays to detect a wide range of different (eco)toxicological effects. The Nrf2-CALUX® assay is used to detect chemicals or complex mixtures potentially inducing oxidative stress. The nuclear factor erythroid 2 (NFE2)-related factor 2 (Nrf2) is a transcription factor with different functions in oxidative stress response and xenobiotic biotransformation. The Nrf2 mediates the induction of xenobiotic metabolization enzymes such as Glutathione-S-transferase. In its activated form the transcription factor induces the expression of several anti-oxidant response element dependent genes but it also directly affects the homeostasis of reactive oxygen and nitrogen species (ROS and RNS) (Ma 2013, van der Linden et al. 2014). The activation of the Nrf2 pathway indicating an oxidative stress response can be a first indicator for genotoxicity as reactive radicals can interact with cellular macromolecules possibly leading to DNA damage (Van der Oost et al. 2003). Hence, the Nrf2-CALUX® provides insight into the mode of action behind a genotoxic potential.

The Nrf2-CALUX® assay was performed according to the SOP provided by BioDection Systems (BDS) Amsterdam. In general, all specific modifications in respect to crude oil samples that are described in detail in the ERα-CALUX® assay procedure (see chapter III.2.5 below) have been applied to this assay. Only sample concentrations of LEWAF, CEWAF and HEWAFs not affecting U2OS cell viability have been used (see results MTT assay).

Briefly, cells were seeded in a 96-well plate at a density of 1 x 10^5/mL in 1x-concentrated assay medium. 24 h after seeding, cells were exposed to dilutions series of the standard reference

compound curcumin ($1*10^{-8}$ - $1*10^{-4}$ M) and LEWAF, CEWAF or HEWAF samples. For all WAF samples a serial dilution (1:2) of the 100 % stock solution was prepared in sterile ultrapure water. Each dilution was complemented with 3x concentrated assay medium (1:3) to guarantee equally nutrient supply comparable to normal culturing conditions.

After 24 h of exposure the medium was removed, cells were lysed and luciferase activity was measured with the application of luciferin substrate mixture in a luminescence reader (Glomax 96-microplate reader, Promega, Madison, USA). The intensity of the luminescence signal correlated with the activation of the transcription factor Nrf2.

As the Curcumin calibration series does not have maximum response, it is not possible to use the typical relative induction evaluation (see ERα-CALUX®). Instead, the induction factors (IF) were calculated to quantify the response of a sample as recommended by BDS. The IF was calculated for each dilution step of the curcumin standard and the samples by normalizing each luminescence value to the luminescence of the background of the standard. IF values were used to establish a concentration-response fit (4 parameters non-linear regression with variable slope) using the GraphPad Prism 6. Cytotoxic concentrations of the reference compound leading to IF values of 0 were excluded for the fit. Within this fit the concentration of curcumin and the sample that results in an IF=1.5 was calculated because the curcumin standard gives a non-cytotoxic stable response in this range. At an IF of 1.5 the specific activity of each sample was calculated by dividing the sample concentration by the standard concentration and finally expressed in ng Curcumin/µL sample.

III.2.3 Genotoxicity using Micronucleus assay

a) zebrafish liver cells (ZF-L)

The permanent ZF-L cell line was cultured in L15 medium (Leibovitz, with L-glutamine, Sigma Aldrich, L4386), supplemented with 10 % fetal calf serum (Biowest, France) in 75cm² flasks at 28 °C. Cells were passaged regularly when reaching 90 % of confluence. A 3x-concentrated assay medium was prepared from L15-powder, which was finally supplemented with FCS (charcoal stripped, Biowest, France), and penicillin-streptomycin. 1x-concentrated assay medium was prepared by diluting the 3x-concentrated assay medium with sterile ultrapure water.

b) Micronucleus assay with ZF-L cells

To investigate the genotoxic potential of the different oil types the micronucleus assay, a common method to detect structural or numerical chromosomal aberrations, was performed. Micronuclei are defined as acentric condensed chromosomal fragments or whole chromosomes that are not

located in the main nuclei but in the cytoplasm surrounded by a nucleus membrane (Countryman and Heddle 1976, Reifferscheid et al. 2007). The origin of micronuclei is either a chromosome breakage (clastogenic effect) or an inhibition or damage of the spindle apparatus (aneugenic effect) leading to chromosomal fragments not transported to the cell poles during mitosis.

The assay was performed according to the ISO guideline 21427-2 (2004). However, major modifications in respect to assay optimization based on the selected cell line and the sample types mainly influencing the incubation periods and exposure materials were established in pretests. In accordance with other bioassays of this report, the viability of ZF-L cells exposed to dilution series of all WAF samples was evaluated in the MTT assay. The two highest non-cytotoxic sample concentrations of each treatment resulting in a viability of at least 80 % were selected as exposure dilutions for micronucleus assay.

Briefly, cell suspension (in 1x concentrated assay medium) at a density of $5*10^4$ was seeded in sterile small glass petri dishes (40 mm, VWR, Germany) containing sterile cover slips (20 x 20 mm, VWR, Germany), on which cells attach after settling. 24 h after seeding, the attached cells were exposed to dilution series of LEWAF, CEWAF and HEWAF samples in duplicates. To evaluate the test validity a negative control (assay medium only), a positive control (4-Nitroquinoline 1-oxide, exposure concentration $6.22 * 10^{-8}$ M, stock in DMSO) and a solvent control (0.1 % DMSO) were included. After 48 h of exposure the cover slips with attached cell layers were fixed with MeOH: acetic acid (4:1, each for 5 min), air dried and finally sticked onto glass slides using Aquatex adhesive (Merck KGaA, Darmstadt, Germany). Microscopy slides were stained using acridine Orange dye. An Eclipse 50i epifluorescence microscope (Nikon Instruments, Düsseldorf, Germany) with 40 x magnification was used to generate pictures in which micronucleated cells were identified according to the following criteria (ISO 21427-2): a) the maximum size of a micronucleus was one-third of the main nucleus, b) micronuclei had the same staining intensity as normal nuclei and c) micronuclei were clearly separated from the nucleus, d) only cells with clear plasmatic outlines were observed. A total number of 2000 cells per treatment were evaluated for micronuclei formation. Validity criteria were met when in negative and solvent controls not more than 3 % of counted cells were micronucleated and positive control induced a significant increase in micronucleated cells. Statistical analysis was done by Chi^2 test with Yates correction using the program SigmaStat 12.5 (Systat Software, 2007) for each replicate.

III.2.4 Mutagenicity using Ames fluctuation assay

The Salmonella mutation (Ames) assay, developed by Bruce Ames in 1973, is an *in vitro* method to determine mutagenicity of pure substances, mixtures and complex environmental samples (Ames et al. 1973, Maron and Ames 1983). This test uses the gram-negative bacterial strain *Salmonella typhimurium* containing a mutation in the histidine operon (auxotrophic mutants). Thus, these bacteria cannot synthesize the amino acid and are unable to grow on histidine-free culture

medium. Spontaneously, or, to a greater extent, in the presence of mutagens some bacteria mutate to wild-types, and thus are able to synthesize histidine again (prototroph revertants). The fluctuation assay is a microplate format modification of the classical plate incorporation assay with advantages like smaller sample volume, less material consumption and faster evaluation (Kamber et al. 2009).

The Ames fluctuation assay was performed according to international standard operation guideline 11350 (ISO 2012) with water samples described in Reifferscheid et al. (2012). In general, the two tester strains *Salmonella typhimurium* TA98 and TA100 were used indicating the potential to induce frame shift and base pair exchanges mutations, respectively. The samples were tested with and without the metabolic activation system S9 (obtained from rat liver and induced with β-naphthoflavone/phenobarbital, Envigo, Germany) to detect a possible pre-mutagenic character which could be activated by liver enzymes. The non-cytotoxic concentration range of each sample was detected within in the normal fluctuation assay procedure using the Salmonella strain TA 98 by measuring the cell density before and after the 100 min incubation period as cell growth is an indicator for toxicity. Concentrations resulting in cytotoxicity of 50 % or higher were excluded for the mutagenicity detection.

Overnight-cultures of both strains were incubated at 37 °C and 150 rpm for 9.75 h in an Innova-40 incubation shaker (New Brunswick, Scientific, New Yersey, USA). After overnight-cultures were adjusted to a certain cell density (1800 FAU for TA98, 450 FAU for TA100), bacteria solution in exposure medium and, if needed, supplemented S9 fraction were added to the sample dilution in a sterile 24-well glass plate. Additionally, negative and strain-specific positive controls were tested in order to identify the validity of each experiment. Approaches were incubated for 100 min under the same conditions as described above. The cell suspension was then transferred to a 384 well-plate preparing 16 replicated wells per treatment concentration. The plates were incubated at 37 °C for 48 h. Wells with revertant bacteria changed the color from purple to yellow in the pH-sensitive reversion indicator medium. Validity criteria were met when in negative control no more than 10 positive wells and in positive control more than 25 positive wells were observed. For statistical evaluation the software ToxRat (ToxRat Solutions GmbH, Alsdorf, Germany) was used. After verifying the variance homogeneity and normal distribution, Williams multiple t-test was used to determine significant differences from control.

III.2.5 Endocrine disruption using ERα-CALUX® assay

a) human osteosarcoma U2OS cells transfected with human ERα

In this assay the osteosarcoma cells U2OS provided by BDS have been used. The ERα cells are stably transfected doubly with the human estrogen receptor α (ERα) and a reporter gene construct expressing a luciferase gene. The reporter gene is expressed under the control of responsive

elements for the activated estrogen receptor(Sonneveld et al. 2004). The culturing conditions for this cell line are described in chapter III.2.2.a.

b) ERα-CALUX® assay procedure

The ERα-CALUX® bioassay is used to determine the receptor mediated estrogenic potential of chemicals, complex mixtures and environmental samples. The mechanism is based on the activation of the human estrogen receptor α (ERα). In principle, exposure to xeno-estrogens results in the uptake of chemicals through the cytoplasmic membrane due to the lipophilic properties of most endocrine disruptors (Legler et al. 2002, Sonneveld et al. 2005). The chemicals bind and thus activate the endogenous ERα. The ligand-receptor complex is translocated into the nucleus and binds to hormonal responsive elements (estrogen responsive elements, ERE) in the promotor region of the luciferase genes. The DNA-bound receptor then induces the expression of the luciferase genes, leading to presence of the enzyme in the cell. Luciferase expression can finally be measured by lysing the cells, adding the substrate luciferin and measuring light photon production.

The ERα-CALUX® was performed according to the ISO/DIS 19040-3 (2014) for water quality assessment and the standard operation procedures (SOP) of BDS (2017). Methodical and material adaptions concerning the crude oil testing were elaborated in pretests. Major adaptions include the avoidance of any plastic material in order to exclude adsorption and reduce evaporation of hydrophobic sample compounds in the WAFs. In particular, glass-coated 96-well plates (WebSeal Plate+, VWR, Germany) and glass plates that covered the well plates were used.

Briefly, cells were seeded in a 96-well plate at a density of 1 x 10^5/mL in 1x-concentrated assay medium.

24 h after seeding, cells were exposed to dilutions series of the standard reference compound 17-ß estradiol (E2, 0.1 pM - 330 pM) and LEWAF, CEWAF or HEWAF samples. For all WAF samples a serial dilution (1:2) of the 100 % stock solution was prepared in sterile ultrapure water. Each dilution was complemented with 3x concentrated assay medium (1:3) to guarantee equally nutrient supply comparable to normal culturing conditions. To avoid non-specific cytotoxic effects, the highest test concentrations for each WAF were determined in the MTT cell viability assay.

After 24 h of exposure the medium was removed, cells were lysed and luciferase activity per well was measured in relative light untis (RLU) with the application of luciferin substrate mixture in a luminescence reader (Glomax 96-microplate reader, Promega, Madison, USA). The intensity of the luminescence signal correlated with the activation of the ERα. All data were checked for validity criteria developed by BDS focusing on selected criteria of the E2 calibration curve like goodness of fit (> 0.98) or range of resulting E2 EC50. Relative luminescence was expressed as percentage of maximal E2 activity. Afterwards, concentration-response curves were fitted for the E2 standard and the WAF dilutions using 4-parameter non-linear regression with variable slope in GraphPad Prism 6. Additionally, 17ß-estradiol equivalents (EEQs) were calculated for each WAF dilution using the

E2 curve fit. Importantly, resulting EEQ values in ng E2/L sample were already corrected for the sample dilution and hence represent the EEQ value of one sample. In case more than the highest test concentration resulted in calculable EEQ values, the EEQ of the lowest concentration was used for further analysis.

III.2.6 Endocrine disruption using A-YES® assay

A second assay on endocrine disruptive potential was integrated in the bioassay battery, as the A-YES assay® is able to investigate the receptor-mediated estrogenicity of water samples spanning a broad range of salinities between freshwater and marine conditions.

a) Arxula adeninivorans cells

The yeast A. adeninivorans has been stably transfected with the human ERα. Deviating from the aforementioned U2OS cells the receptor activates the reporter gene phyK, which encodes the enzyme phytase (Hettwer et al. 2018). The A-YES® test kit was purchased from New Diagnostics GmbH, Dresden, Germany. In the kits, the yeast is delivered as freeze-dried aliquots.

b) A-YES® assay procedure

The A-YES® was performed according to the ISO/FDIS draft 19040-2 (2018) for water quality assessment and the SOP by New Diagnostics GmbH with minor modifications regarding crude oil toxicity testing. Methodical and material adaptions concerning the crude oil testing were elaborated in pretests. As described previously in the present report, plastic materials were avoided with the usage of glass coated 96-deep well plates (WebSeal Plate+, VWR, Germany) in particular.

Briefly, freeze dried yeast cells were reactivated by incubating the washed and re-dissolved yeast at 30 °C and 450 rpm for 1 h. In the meantime, exposure concentrations of the standard reference compound 17ß-estradiol (1 - 80 ng/L) and the different WAF samples were prepared in ultrapure water (low salinity conditions) or artificially re-mineralized dilution water (elevated salinity conditions at 6 ‰), respectively. For the WAF samples a 1:2 dilution series of the 100 % stock solution was prepared. All WAF samples were tested with the undiluted stock solution as the highest test concentration as no cytotoxic effects have been observed in pretests.

After reactivation the yeast suspension was split up to prepare separate inoculation media for the E2 concentration series and the WAF samples. Based on the salinity of the WAF samples (low or elevated) sterile medium mixes for inoculation (low and high salinity) of the test kit were mixed to obtain similar salinity conditions for the E2 and the WAF. Sample dilutions and corresponding E2 calibration series were tested on one plate in triplicates (WAF dilutions) or duplicates (E2), respectively.

Exposure solutions and individual yeast inoculation media were combined in the deep-well plate, which was then covered by a sterile absorptive foil and incubated at 31 °C and 900 rpm for 22 h.

Afterwards, the cells were separated from the exposure medium by centrifugation (10 min, 700 g) and supernatant was transferred to a new 96-well plate (CytoOne®, StarLab International GmbH, Hamburg, Germany). The substrate p-nitrophenyl phosphate was added to the supernatant and the phytase activity was photometrically detected (405 nm) using a microplate reader (Infinite® M 200, Tecan Group, Switzerland). Additionally, the growth rate of the yeast was photometrically detected (630 nm) in re-dissolved yeast pellet.

IV. Material and Methods of toxicity profiling

In general, a complex set of biological effect data was available for the toxicity profiling. Small scale *in vitro*-based methods as well as methods using whole organisms of a broad range of developmental stages (from early embryo to adults) were used to investigate the acute toxic, genotoxic, endocrine disruptive, neurotoxic or biotransformation activating potential of crude oils and refined petroleum products. An overview of biological endpoints with the modes of action categorization that is used throughout the present report can be found in **Figure 1**. Most biological effect data were available for the NNA crude oil with > 20 different endpoints further extended with data on different salinity conditions, temperature regimes, concentration ranges and time windows of exposure. The refined petroleum products (MGO, IFO180) were investigated in around half of the assays used for NNA toxicity profiling.

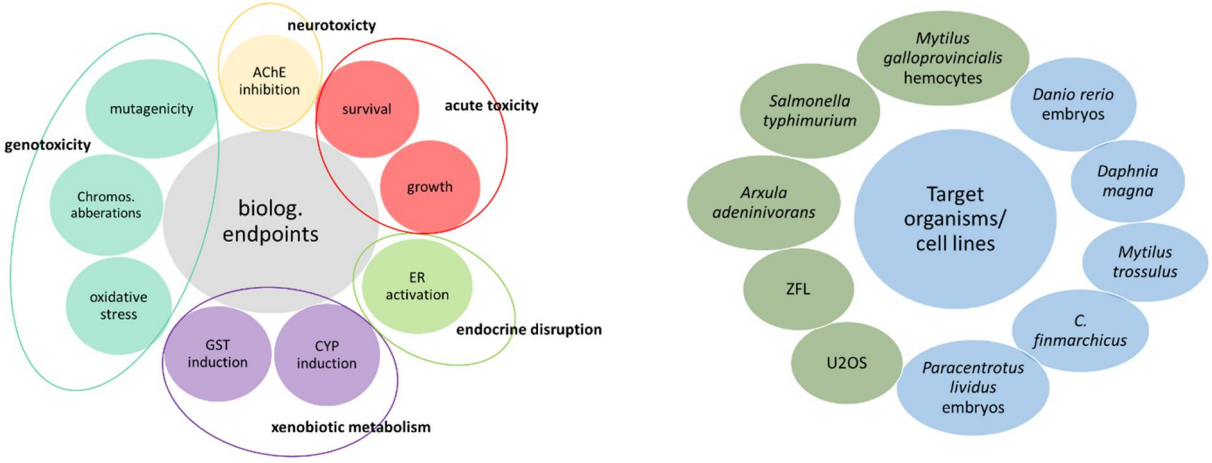


Figure 1 Overview of biological endpoints and target organisms/cell lines used for toxicity profiling in WP3 of the GRACE project. *The biological endpoints investigated in a set of acute and mechanism-specific bioassays were further categorized into the main classes of modes of actions (acute toxicity, endocrine disruption, neurotoxicity, genotoxicity and xenobiotic metabolism). Target organisms can be separated in in vitro based methods using permanent cell lines or strains (green) and methods using whole organisms (blue).*

The data processing workflow to establish a fingerprinting toolbox included first the selection of data suitable for the fingerprinting toolbox followed by the classification of each individual response

within the biological endpoints. Finally, the combined data sharing a common scale were interpreted using biological expert judgement and statistical analysis in order to give a recommendation on a sensitive, time and cost efficient bioassay battery. Details on the classification system and statistical approaches can be found in the following sections.

In order to allow for a conclusion about the oil toxicity the goal of this workflow and simultaneously the main challenge was to combine toxicity data on various levels, some focusing on highly specific molecular interactions others giving a combined response of multiple processes.

IV.1 Classification system

The effect data of all biological endpoints were individually transformed into effect classes between 1 and 5 (see **Figure 2**). The aim of the data transformation in a shared classification system was to rescale the different responses into one common scale.

Methods combining a huge set of biological effect data for a comprehensive toxicity evaluation have been established in previous studies. For example, in order to summarize and simplify biomarker responses in monitoring programs the integrated biomarker response (IBR) was suggested (Beliaeff and Burgeot 2002). However, the transformation of the data often includes a normalization to a maximum induction or the induction of a reference site, which is in fact often not available in monitoring studies. Additionally, within this the resulting factor of normalized data does only marginally or not at all reflect the dimension of the biological relevance. While for some endpoints slight changes compared to control or maximum induction imply a significant change in the biological response other endpoints have a much broader spectrum of biological responses.

Hence, the classification system of the present study integrated biological expert judgement already in the first step of data transformation. The classification of individual bioassays was established based on criteria involving

a) limits of detection and/or - quantification
b) baseline activities
c) effect-based thresholds
d) knowledge on biological responses occurred before (with focus on petroleum components)
e) knowledge on extrapolation towards higher biological organization levels

In general, it has to be considered that not all criteria could be included in each individual bioassay classification due to lack of information and relevance.

An overview showing the selected endpoints and classification can be found in **Table 2**. Detailed references on established classification steps can be found in Table 5 (Appendix). As receptor-mediated *in vitro* bioassay share the main advantages of response quantification and an increasing integration in risk management and monitoring programs for surface and drinking water quality

assessment the present classification includes bioassay-specific recommendations of trigger values elaborated previously below which no adverse effect is expected. Other endpoints like biomarker responses and acute toxicity data lack a discussion about general trigger values and hence those classes were established mainly on knowledge on biological responses due to known effects of well-described chemical groups (e.g. PAHs).

Radar plots of the classified effect data were plotted in RStudio (Version 1.2.1335) with the package “tidyverse” according to Wickham et al. (2017).

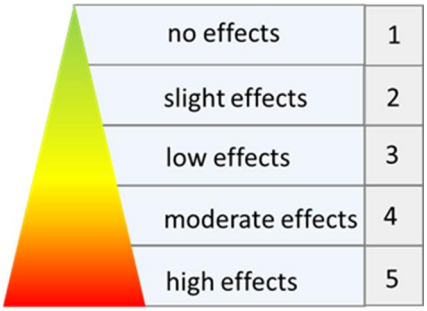


Figure 2 Classification system for biological effect data. *All effect based responses were transformed into the classes 1 to 5 based on classification systems individually elaborated for each bioassay.*

Table 2 Effect classes of individual bioassays. *Background information for individual classification can be found in Table 5.*

	Target organism/cell line	Effect category	Endpoint	Effect classes				
				1	**2**	**3**	**4**	**5**
acute toxicity	*D. rerio, D. magna, C. finmarchicus, P. lividus, M. trossulus*	acute toxicity	LC50 [% of stock]	x ≥ 80	80 > x ≥ 55	55 > x ≥ 30	30 > x ≥ 5	x < 5
ERα-CALUX®	U2-OS	endocrine disruption	EEQ [ng E2/L]	x ≤ 0.1	0.1 < x ≤ 0.3	0.3 < x ≤ 1	1 < x ≤ 10	x > 10
A-YES®	*A. adeninivorans*	endocrine disruption	EEQ [ng E2/L]	x ≤ 0.56	0.56 < x ≤ 1.68	1.68 < x ≤ 5.6	5.6 < x ≤ 56	x > 56
Nrf2-CALUX®	U2-OS	oxidative stress	specific curcumin activity [µgCurc/L]	x ≤ 21	21 < x ≤ 63	63 < x ≤ 210	210 < x ≤ 2100	x > 2100
MTT assay	U2-OS/ ZFL/ *M. galloprovincialis hemocytes*	cell viability	reaching 80 % viability [% of stock]	x ≥ 66	66 > x ≥ 49.5	49.5 > x ≥ 33	33 > x ≥ 16.5	x < 16.5
Micronucleus assay	ZF-L	genotoxicity	IF	x ≤ 1	1 < x ≤ 1.8	1.8 < x ≤ 2.6	2.6 < x ≤ 3.5	x > 3.5
Ames fluctuation assay TA 98 +S9	*S. typhimurium*	genotoxicity	# revertant wells	0 ≤ x < 1.6	1.6 ≤ x < 15	15 ≤ x < 28.4	28.4 ≤ x < 41.8	x ≥ 41.8
Ames fluctuation assay TA 98 -S9	*S. typhimurium*	genotoxicity	# revertant wells	0 ≤ x < 2.5	2.5 ≤ x < 15.6	15.6 ≤ x < 28.7	28.7 ≤ x < 41.8	x ≥ 41.8
Ames fluctuation assay TA 100 +S9	*S. typhimurium*	genotoxicity	# revertant wells	0 ≤ x < 3.2	3.2 ≤ x < 16.53	16.53 ≤ x<29.87	29.87 ≤ x<43.2	x ≥ 43.2
Ames fluctuation assay TA 100 -S9	*S. typhimurium*	genotoxicity	# revertant wells	0 ≤ x < 4	4 ≤ x < 18	18 ≤ x < 32	32 ≤ x < 46	x 46
EROD activity	*D. rerio*	xenobiotic	IF	x ≤ 1	1 < x ≤ 1.5	1.5 < x ≤ 2	2 < x ≤ 5	x > 5

AChE inhibition	*D. rerio*	metabolism neurotoxicity	IF	$x \geq 1$	$1 > x \geq 0.8$	$0.8 > x \geq 0.6$	$0.6 > x \geq 0.4$	$x < 0.4$
AChE inhibition	*M. trossulus*	neurotoxicity	[specific activity mg^{-1} min^{-1}]	$CT_{mean} \pm 0.5 \geq x$	$CT_{mean} \pm 0.5 < x \leq CT_{mean} \pm 1$	$CT_{mean} \pm 1 < x \leq CT_{mean} \pm 1.5$	$CT_{mean} \pm 1.5 < x \leq CT_{mean} \pm 2$	$x > CT_{mean} \pm 2$
Lipidperoxidation	*M. trossulus*	oxidative stress	[specific activity mg^{-1} min^{-1}]	$CT_{mean} \pm 0.5 \geq x$	$CT_{mean} \pm 0.5 < x \leq CT_{mean} \pm 1$	$CT_{mean} \pm 1 < x \leq CT_{mean} \pm 1.5$	$CT_{mean} \pm 1.5 < x \leq CT_{mean} \pm 2$	$x > CT_{mean} \pm 2$
Catalase activity	*M. trossulus*	oxidative stress	[specific activity mg^{-1} min^{-1}]	$CT_{mean} \pm 0.5 \geq x$	$CT_{mean} \pm 0.5 < x \leq CT_{mean} \pm 1$	$CT_{mean} \pm 1 < x \leq CT_{mean} \pm 1.5$	$CT_{mean} \pm 1.5 < x \leq CT_{mean} \pm 2$	$x > CT_{mean} \pm 2$
Glutathion-S-transferase activity	*M. trossulus*	xenobiotic metabolism	[specific activity mg^{-1} min^{-1}]	$CT_{mean} \pm 0.5 \geq x$	$CT_{mean} \pm 0.5 < x \leq CT_{mean} \pm 1$	$CT_{mean} \pm 1 < x \leq CT_{mean} \pm 1.5$	$CT_{mean} \pm 1.5 < x \leq CT_{mean} \pm 2$	$x > CT_{mean} \pm 2$
Glutathione reductase activity	*M. trossulus*	oxidative stress	[specific activity mg^{-1} min^{-1}]	$CT_{mean} \pm 0.5 \geq x$	$CT_{mean} \pm 0.5 < x \leq CT_{mean} \pm 1$	$CT_{mean} \pm 1 < x \leq CT_{mean} \pm 1.5$	$CT_{mean} \pm 1.5 < x \leq CT_{mean} \pm 2$	$x > CT_{mean} \pm 2$
Condition index	*M. trossulus*	acute toxicity		$CT_{mean} \pm 0.5 \geq x$	$CT_{mean} \pm 0.5 < x \leq CT_{mean} \pm 1$	$CT_{mean} \pm 1 < x \leq CT_{mean} \pm 1.5$	$CT_{mean} \pm 1.5 < x \leq CT_{mean} \pm 2$	$x > CT_{mean} \pm 2$

IV.2 Principal component analysis (PCA)

As the main goal was to establish an oil type-specific toxicity profile over all bioassays, typical univariate statistics would not achieve the target question due to, e.g., alpha error accumulation. Hence, the multivariate Principal Component Analysis (PCA) statistical approach was selected to support biological interpretation of an effect-based bioassay battery. Within the PCA effect-specific bioassays can be found that statistically explain x% of the data variance against the background of all biological responses. Hence, key bioassays that are statistically sensitive for a specific treatment can be distinguished from redundant bioassays without additional value for the profile.

PCA was performed with classified biological effect data using Canoco for Windows (version 4.5, Canoco GmbH). Ordination diagrams were plotted using CanoDraw (Canoco GmbH).

IV.3 Stochastic resampling for PCA

First, all individual bioassay data were investigated for normal distribution (ND, Shapiro Wilk) and equal variance followed by the identification of bioassays leading to significant differences in treatments (LEWAF, CEWAF, HEWAF) using pairwise comparison (T-Test, in case ND and/or VH failed: Man Whitney Rank Sum Test) by means of the software SigmaPlot (Version 12.0, Systat Software, 2007) with a significance level of $p<0.05$.

For normal distributed bioassays with significant differences across treatments, 1000 randomized numbers were generated in R (R core team, Vienna, Austria, 2019).

V. Results of individual bioassays

V.1 Acute toxicity in selected aquatic vertebrates and invertebrates

The results of the acute toxicity test used for the toxicity profiling of the present report can be found in detail in the corresponding deliverables on zebrafish early life stages (D 3.12, D 3.16), copepods (D 3.14, D 3.16) sea urchin embryos (D 3.16) and mussels (D 3.11).

V.2 Mechanism-specific toxicity

V.2.1 Cell Viability examination using MTT bioassay

To exclude false negative results based on cytotoxicity in different mechanism-specific bioassays the concentration ranges of WAF dilutions resulting in normal cell viability had to be defined. Hence, the colorimetric viability MTT bioassay was performed to identify non-cytotoxic dilutions of WAFs for the U2OS cell line used in the CALUX bioassays, and the ZF-L cell line used in the micronucleus assay.

With the exception of the NNA crude oil toxicity on the ZF-L cell line all oil types in both cell lines led to higher toxicity in dispersed oil exposure (CEWAF) compared to oil alone (LEWAF) exposure. While for the exposure to LEWAF already the highest concentrations resulted in a cell viability comparable to the untreated control, the CEWAF exposure induced a decreased viability especially in the highest test concentration, with non-cytotoxic concentration down to below 16.6 % of stock (reaching > 80% viability).

Both cell lines showed slight differences in their sensitivity towards the different oil types with the shared characteristic of the NNA crude oil being the least toxic. For U2OS cells both refined petroleum products resulted in equal cell viability while in ZF-L cells the heavy fuel oil (IFO 180) induced higher cytotoxicity.

To evaluate the influence of the dispersant on the resulting CEWAF toxicity cells were exposed to the dispersant alone. In general, the cytotoxicity of the dispersant exposure approximately correlated with the cytotoxicity of the dispersant content in the CEWAF sample. Hence, it can be concluded that the higher cytotoxicity of the dispersed crude oil (CEWAF) compared to the crude oil only (LEWAF) was caused more likely by the dispersant toxicity then by a higher bioavailability of crude oil compounds in the water phase due to the partitioning kinetics of the dispersion effect.

V.2.2 Oxidative stress response in U2OS cells using Nrf2-CALUX® assay

The naphthenic North Sea crude oil LEWAF did not activate the transcription factor Nrf2, as the induction factor of 3 individual experiments did not differ from the background level. The chemically dispersed crude oil using the dispersant Finasol OSR 51 and the dispersant alone did show a concentration-related increase in the Nrf2 activity. Based on the Curcumin calibration series, the calculated specific activity indicated a slightly reduced response of the dispersant alone (HEWAF) compared to dispersed crude oil leading to the assumption that the dispersant itself contributes to the higher responses in CEWAF compared to LEWAF treatments.

V.2.3 Genotoxicity using Micronucleus assays

The exposure concentrations for the different oil types were selected based on the cytotoxicity data (see chapter V 2.1; MTT assay) to guarantee normal cell growth and defined as the highest exposure concentration reaching at least 80 % viability. Hence, based on the different cytototoxic potential of the three oil types the exposure solutions varied between 25 - 66 % of stock (LEWAF) and 0.5 - 16 % of stock (CEWAF). The validity criteria of a maximum of 3 % micronucleated cells in negative and solvent control and a significant increase in micronuclei formation in positive control were met for all the data included in the present report.

In general, both LEWAF and CEWAF of the naphthenic North Sea crude oil induced significantly increased micronuclei formation compared to the unexposed negative control.

In contrast to the crude oil, both refined petroleum products did not show a genotoxic potential in cells exposed to LEWAFs with micronuclei formation rates comparable to untreated control. However, dispersed marine gas oil exposure (CEWAF MGO) resulted in significantly increased micronuclei formations.

In general, a clear trend in the genotoxic potential of the three oil types was observed. The NNA crude oil showed the highest potential to induce chromosomal aberrations both in LEWAF and CEWAF exposure, followed by the marine gas oil (MGO) in CEWAF exposure and the heavy fuel oil (IFO 180) with no significantly increased micronucleus induction in all treatments.

Additionally, the micronucleus assay was performed with HEWAFs of the dispersant Finasol OSR 51 to evaluate the influence of the dispersant on the elevated micronuclei formation of CEWAF compared to LEWAF exposure. Within this the dispersant and oil proportions in HEWAF complied with the proportions in CEWAF to guarantee a comparability of results. The HEWAF treatment resulted in micronuclei induction slightly but not significantly increased compared to the untreated control.

V.2.4 Mutagenicity using Ames fluctuation assay

The cytotoxicity of all WAF samples was investigated during the normal Ames fluctuation assay procedure using the tester strain TA98 by measuring the optical density before and after the short term incubation as the cell growth rate is an indicator for viability.

All results of the Ames fluctuation assay passed the validity criteria as the number of revertants of the negative control was <10 and of positive controls was >25.

For both tester strains (TA98 and TA100) no significant increase in the revertant formation compared to the negative control was found for any of the different WAF treatments of the NNA crude oil. The same trend is indicated for the refined oil types. Even the application of the S9 fraction obtained from rat livers to detect a pre-mutagenic character which could be activated by the liver enzymes did not convert the WAF components into DNA-intercalating compounds.

The results indicate that the Naphthenic North Sea crude oil does not cause frame shift or base exchange mutations in the Salmonella strains TA 98 and TA 100, respectively. However, concluding a non-mutagenic character of the Naphthenic North Sea crude oil has to be treated carefully as only two tester strains on frame shift and base exchange mutation were investigated. To date, several tester strains with different mutation types exist.

V.2.5 Endocrine disruption using ERα-CALUX® assay

In order to avoid false negative results, only non-cytotoxic sample dilutions as discussed in a previous chapter (MTT bioassay) were investigated.

Results obtained from the ERα-CALUX® assay with the NNA crude oil WAFs indicate that sample compounds of the LEWAF and CEWAF do interact with the estrogen receptor in the highest test concentrations as the luciferase induction increased with increasing sample concentrations. The CEWAF exposure resulted in a stronger receptor mediated estrogenicity.

Again, the role of the dispersant on the estrogenic potential observed for the CEWAF was investigated by the dispersant HEWAF approach. The dispersant alone did not contain compounds activating the ERα with responses below the limits of quantification (LOQ) or even detection (LOD). Hence, the dispersant had no influence on the receptor mediated estrogenicity of the CEWAF.

V.2.6 Endocrine disruption using A-YES® assay

Within the A-YES® assay water samples of a huge range of salinities between freshwater and marine conditions can be investigated for their potential to activate the human estrogen receptor α (ERα) stably transfected in the yeast *A. adeninivorans*. As an additional assay originally not planned for the project the assay was first performed with the NNA crude oil samples in order to evaluate the applicability for this field of research.

Again, out of all WAF treatments the dispersed crude oil (CEWAF) showed the highest endocrine disruptive potential. The dispersant alone (HEWAF Fin51) did activate the ERα in 1 out of 3 or 4 independent replicates in a quantifiable manner. In direct comparison, a trend of a higher receptor-mediated estrogenicity for WAFs prepared in elevated salinity conditions (6 ‰) compared to WAFs prepared in low salinity conditions was observed. This is in contrast to previous expectations as especially for PAHs a reduction of the solubility with increasing salinity was reported (Eganhouse and Calder 1976, Saranjampour et al. 2017, Whitehouse 1984, Xie et al. 1997). However, increased estrogenic activity in elevated salinity conditions might not only be related to dissolved hydrocarbons in the water sample but rather be related to the test system itself. The yeast cells do show slightly divergent characteristics when growing in different media (e.g. morphology, size) which might lead to changes in sensitivity.

In general, this assay was a useful screening tool for receptor-mediated endocrine disruption especially in respect to water samples from brackish and marine environments. Even though a higher variability of resulting estrogenic response compared to the ERα-CALUX® was observed, the sensitivity of both assays were in a comparable range. Importantly, a direct comparison between both assays on ER activation potential is limited due to several differences regarding the test systems including for example cell types (yeast vs. human cell), reporter mechanism (extracellular reporter activity mediated absorption vs. intracellular reporter mediated luminescence) and EEQ calculation procedure with E2 calibration curve.

VI. Toxicity profiling

VI.1 Toxicity profile of Naphthenic North Sea crude oil

VI.1.1 Toxicity profile of untreated and chemically dispersed NNA WAFs

In total, 46 different data endpoints of which 31 were included in the statistical analysis were available for the NNA crude oil toxicity profile. Remaining data focused on additional time windows of exposure or exposure concentrations and hence were not included to maintain a better overview.

Both LEWAFs and CEWAFs of the NNA crude oil induced acute and mechanism-specific toxicity in the selected bioassays. The most sensitive organisms were the invertebrates *Calanus finmarchicus* and *Daphnia magna* followed by the fish embryos of *Danio rerio* for both LEWAF and CEWAF treatments. Additionally, crude and dispersed oil compounds did interact with specific receptors (ERα), activate or inhibit regulatory enzymes (CYP, AChE), induced oxidative stress (e.g. Nrf2, catalase, lipid peroxidation) and further chromosomal aberrations (MNC assay) indicating a set of toxicity mechanisms like endocrine disruption, general cellular stress or genotoxicity.

In order to determine the most sensitive biomarkers for the NNA crude oil the biological interpretation and the statistical evaluation using PCA were combined. In general, the first axis (x-axis) of the ordination diagram already explained 59.4 % of the bioassay data variance (**Figure 3**), highlighting the importance of bioassay vector orientation in relation to this axis. Furthermore, also the length of a vector, the angle between different vectors and the spatial orientation is important for the interpretation.

In a direct comparison a much higher toxicity of the dispersed crude oil compared to untreated crude oil has been observed which was further shown by the accumulated spatial distribution of the bioassays in the corresponding corner of CEWAF replicates in the ordination diagram (**Figure 3**). The acute toxicity towards selected organisms, in particular to *C. finmarchicus* and *D. rerio*, as well as the activity of biotransformation phase II CYP enzymes in zebrafish larvae (120 hpf) were highly sensitive and responsive in CEWAF exposure as indicated by the small angle to the first axis (high explanation of data variance) and the length of the corresponding vectors. Those observations support the current scientific knowledge on crude oil toxicity mode of actions as hydrocarbon compounds such as PAHs have been shown to cause a strong AhR-dependent or independent CYP activation (e.g. Barron et al. 2004, Incardona et al. 2006, Van der Oost et al. 2003). Furthermore, especially early life stages of fish are known to be highly sensitive towards crude oil exposure (e.g. reviewed in Incardona 2017, Johann et al. 2019b, Perrichon et al. 2016, Perrichon et al. 2018). However, also the *in vitro*-based bioassays including the investigation of cytotoxicity

(MTT), endocrine disruption (AYES®) and oxidative stress (Nrf2) were determined to be sensitive towards the exposure as all biomarkers were clustered around the first axis in the ordination diagram.

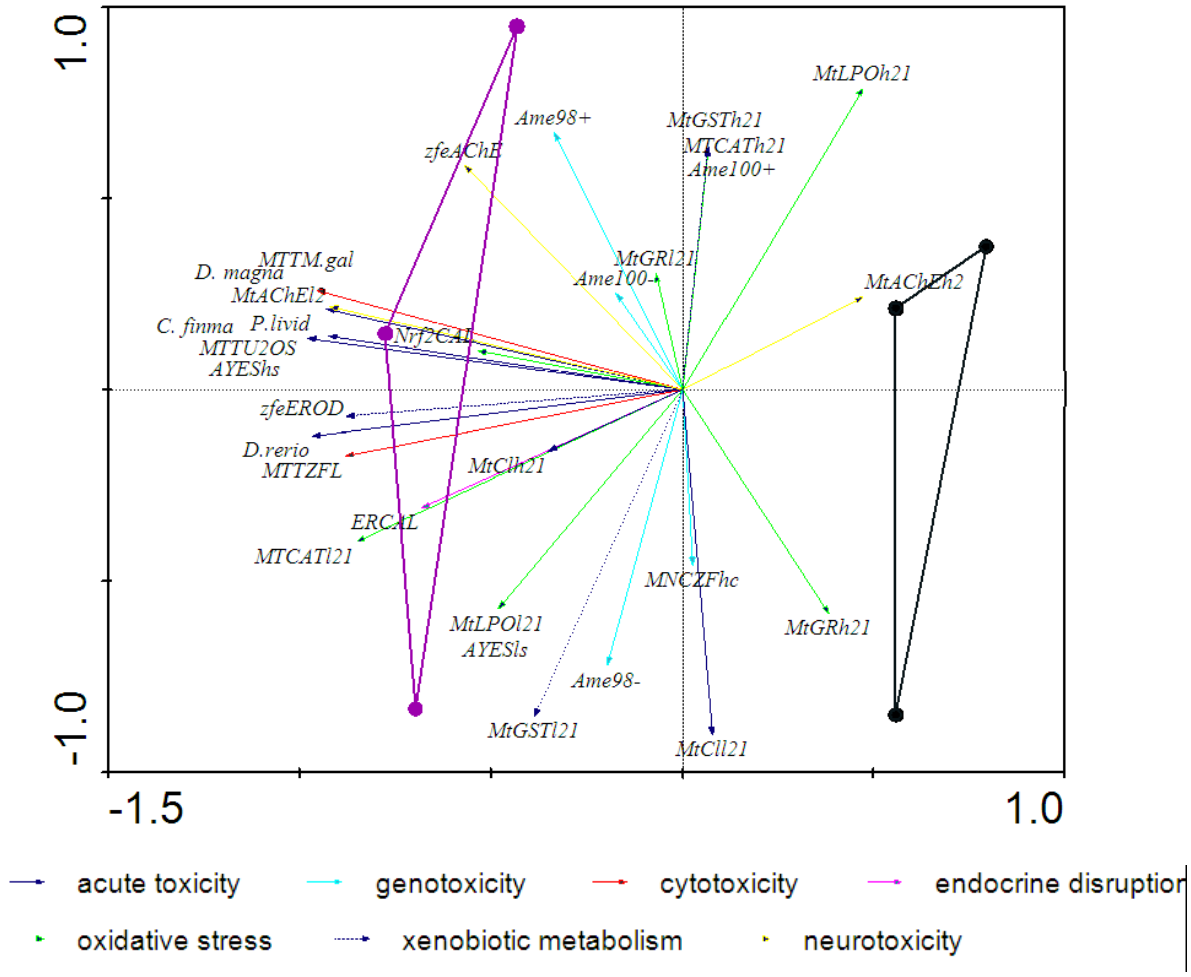


Figure 3 Principal component analysis of classified biological effect data on untreated and dispersed NNA crude oil WAFs. *Arrows denote the individual bioassays while points denote enveloped Individual replicates of LEWAF (black points and envelope) and CEWAF (purple points and envelope) treatments (n=3). Cumulative variance of the first and second order axis were 59.4 % and 77.1 %, respectively. PCA was performed in Canoco for Windows 4.5 with ordination diagram plotted using CanoDraw.*

Ames fluctuation assay on mutagenicity is the one assay that can be clearly identified as not sensitive and therefore not recommended for a bioassay battery for crude oil toxicity testing. Neither LEWAF nor CEWAF did induce base exchange or frame shift mutations in the bacterial strains independent of the application of a mammalian metabolic activation system (S9). However, only two tester strains with specific mutations have been investigated limiting the overall conclusion. For future experiments additional tester strains with different mutation types could be investigated for mutagenic response towards petroleum WAF exposure.

Besides *in vitro*-based bioassays also biomarkers in *Mytilus trossulus,* which are presented in detail in deliverable 3.11 (Lekube et al. 2019), have been included in the present report. Semi-chronic exposure experiments at two salinity conditions (5.6 and 15 psu) have been performed. Interestingly, some biomarkers in *M. trossulus* including the induction of oxidative stress (MtLPOh21, MtGRh21) and neurotoxicity (AChEh2) after 21 day of exposure at elevated salinity conditions (15 psu) were specifically responsive for LEWAF treatment with vectors pointing towards the LEWAF replicates in the ordination diagram. However, those endpoints had relatively larger angles to the first axis limiting the explanatory power. Hence, even though an increased practical complexity and reduced efficiency compared to the small scale *in vitro* methods is implied results of the present study indicate that biomarkers in mussels should not be excluded from a useful bioassay battery.

VI.1.1 Dispersant influence on toxicity

How does the dispersant contribute to the higher toxicity of the CEWAF treatment? The benefit of a chemical dispersant used to combat oil spills at sea is discussed controversially in the scientific community (e.g. Bejarano et al. 2014, Prince 2015). Several studies have concluded that the application of a chemical dispersant serving as an additional source not only for dissolved hydrocarbons but also for particulate oil droplets (Redman and Parkerton 2015) increases oil toxicity towards biota (e.g. Couillard et al. 2005, Dussauze et al. 2015, Hansen et al. 2019, Ramachandran et al. 2004). Apart from that the application reduces surface slicks and partly increases the biodegradation due to the dispersion effect (e.g. Dupuis et al. 2015). However, in contrast also no increase in biodegradation (WP 2 of current project) has been reported and dispersant role has been clearly attributed to allow the dispersion of an oil and not to stimulate the biodegradation (Prince et al 2013).

In general, only a limited number of studies addressed the role of a dispersant in experimentally relevant conditions allowing a direct comparison to CEWAF data. Hence, a set of small-scale bioassays in GRACE were performed using HEWAFs of dispersant alone with corresponding amounts of Finasol used for the CEWAF preparation. **Figure 4** illustrates the reduction of bioassays from the LEWAF/CEWAF approaches (n= 31) to those bioassays implementing the dispersant HEWAF approach to guarantee a direct comparability (n=20). In particular, experiments on biomarker measurements in mussels and mussel hemocytes (oxidative stress, neurotoxicity, cytotoxicity) as well as the acute toxicity towards *D. magna* and *P. lividus* lack the HEWAF application. Data on all three WAFs were available for the *in vitro* based methods using (recombinant) cell lines to investigate the endocrine disruptive, cytotoxic, oxidative stress inducing or genotoxic potential.

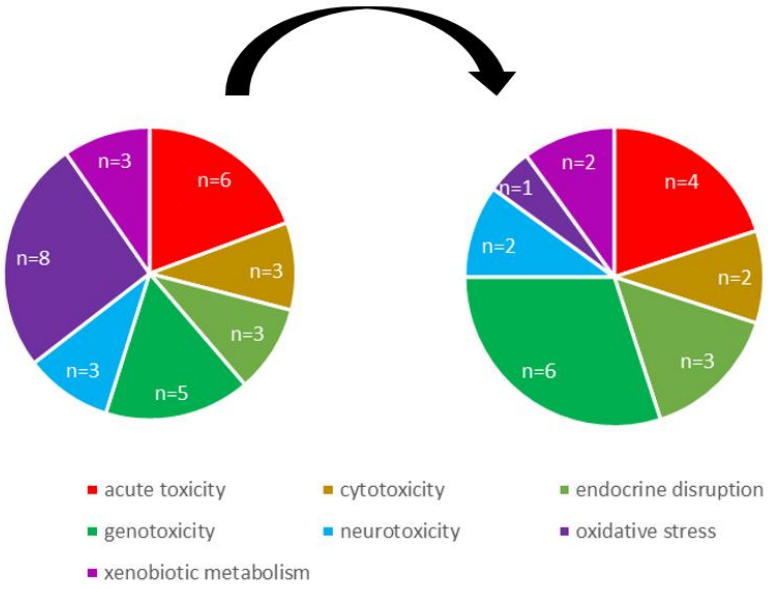


Figure 4 Endpoint-specific data reduction for statisitical analysis based on the implementation of the dispersant HEWAF approach in addition to LEWAF and CEWAF exposure scenarios.

Focusing on a direct comparison between dispersed crude oil and dispersant alone shared and divergent characteristics can be observed. On the one hand a set of biological endpoints including the acute toxicity towards different organisms (*C. finmarchicus, D. rerio*), cytotoxicity (MTT) and oxidative stress (Nrf2-CALUX®) induced comparably strong response in CEWAF and HEWAF. On the other hand, the dispersant showed a reduced potential to induce micronuclei (genotoxicity), to interact with the ERα (ER CALUX®, AYES® low salinity) or to activate biotransformation enzymes (zfEROD120 and 96). Hence, two important hypotheses of dispersant toxicity can be derived: First, the dispersant can contribute to CEWAF toxicity not only based on a higher bioavailability of crude oil compounds (dissolved and particulate) due to the dispersion effect but also based on the toxicity of the dispersant itself. Second, the results obtained from the present study indicate that the dispersant toxicity acts not via specific modes of action including the activation of the ERα or the AhR resulting in CYP activation, but rather induces general stress and damage in cells and organisms.

With the third component of the HEWAF treatment a more complex picture in the ordination diagram can be observed (**Figure 5**). The first axis explains 45.1 % of the bioassay variance, while the second axis explains 26.3 %, indicating the importance of vector orientation especially in relation to the first axis. Again a cluster of bioassays representing acute toxicity for invertebrate species like *C. finmarchicus* and *D. magna* as well as *in vitro*-based assays on e.g. cytotoxicity were correspondingly sensitive for the CEWAF and HEWAF treatment as indicated by their vector orientation close to the first axis and their relative vector length. The CYP induction in zebrafish embryos seems to be specifically indicative for dispersed crude oil exposure as those long vectors

were orientated directly to the CEWAF replicates. Interestingly, the ERα-CALUX® on receptor-mediated estrogenicity, explaining almost the entire variance of the second axis, is statistically able to identify differences between dispersed crude oil and dispersant alone exposure. For the LEWAF treatment a slight responsiveness in micronucleus induction (MNCZFL) and acetylcholinesterase inhibition in zebrafish embryos (zfAChE) can be observed. However, the acetylcholinesterase inhibition in 96 hpf embryos was based on high variance in the datasets. Furthermore, the corresponding results in 120 hpf zebrafish resulted in a decreased AChE inhibition with only slight differences between all three WAF treatments limiting the relevance of this vector in the ordination diagram.

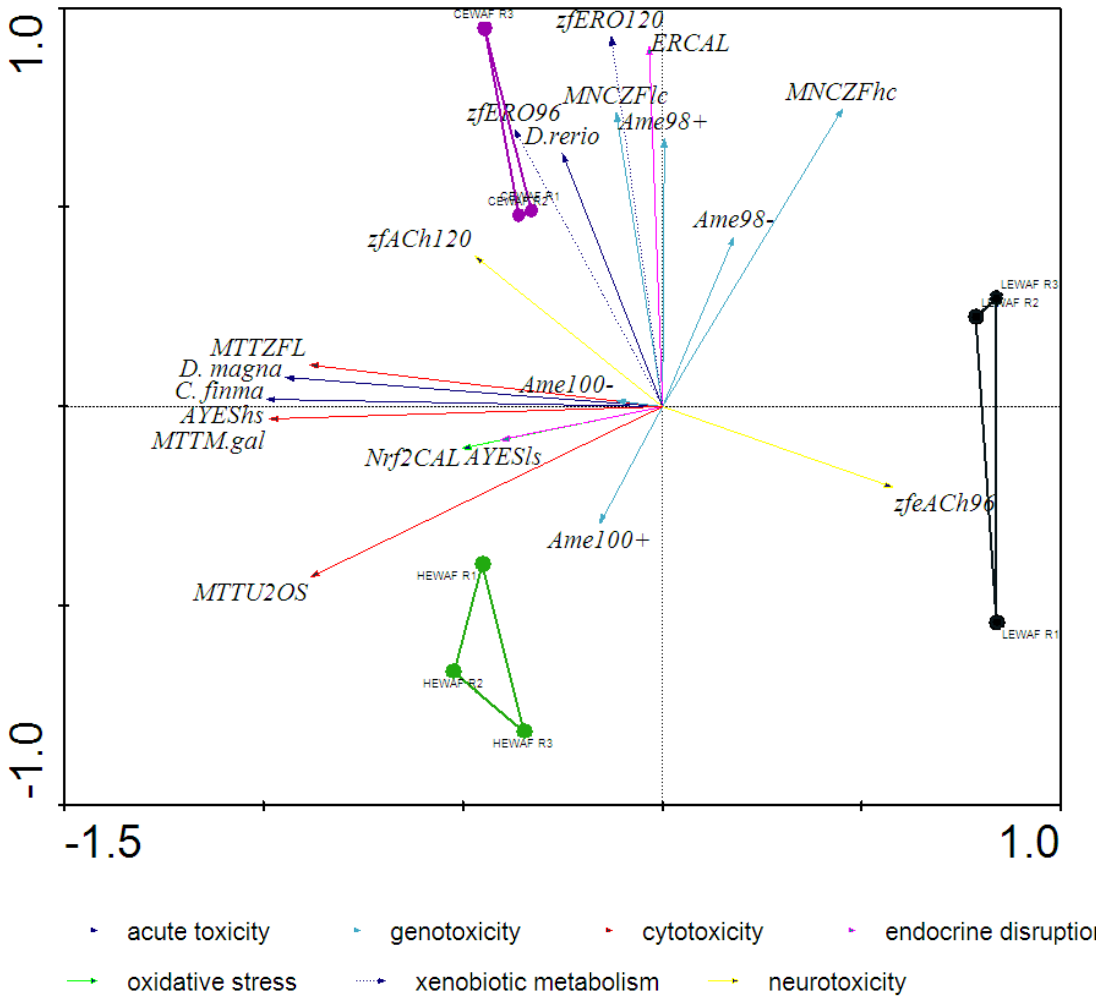


Figure 5 Principal Component Analysis of classified biological effect in WAF dilutions of untreated and dispersed NNA crude oil and dispersant. *Arrows denote the individual bioassays while points denote enveloped Individual replicates of LEWAF (black dots and envelopes), CEWAF (purple) and HEWAF (green) treatments (n=3). Cumulative variance of bioassay data for the first and second axis were 45.1 % and 71.4 %, respectively. PCA was performed in Canoco for Windows 4.5 with ordination diagram plotted using CanoDraw.*

VI.1.3 Validation of PCA for toxicity profiling by stochastic resampling

In the correlation matrix the bioassays cannot be defined as dependent variables, which, from a statistical point of view, limits the implementation of PCA in this data evaluation. However, from a practical point of view performing more than 30 assays with one sample stock at the same day cannot be realized further to the fact that independency of replicates is of high importance for biological effect interpretation.

Hence, additional statistical tests were performed in order to validate the application of PCAs for the present toxicity profiling. In this context, stochastically resampled data (n=1000) based on mean and standard deviation of the original effect data were generated, classified and compared for resulting trends in PCA. Only assays following a normal distribution and showing variance in biological replicates (SD ≠ 0) have been used for the stochastic resampling.

Even though some differences were observed in the ordination diagrams generated from stochastically resampled and real effect data the overall trend of indicative bioassays for the different treatments were consistent. Hence, the results indicate that the PCA was a useful approach to be applied on the current data set.

VI.2 Toxicity profiles of refined petroleum products

LEWAFs and CEWAFs of the refined petroleum products MGO and IFO180 have been investigated in 8 (MGO) or 7 (IFO) individual bioassays focusing on the effect categories acute toxicity, cytotoxicity and genotoxicity. In compliance to the NNA crude oil, the acute and mechanism-specific toxicity increased when a chemical dispersant was added.

For both LEWAF and CEWAF treatments again the acute toxicity tests were the most responsive of the small bioassay battery. Hence, especially the copepods *C. finmarchicus* and fish larvae of *D. rerio* were sensitive towards the exposure of untreated and dispersed oil. While the mechanism-specific endpoints on genotoxicity and general cytotoxicity were only marginally responsive to the LEWAF treatments, highly specific activities were observed in cells exposed to the CEWAFs. In contrast, the Ames fluctuation assay did not indicate a mutagenic potential of the MGO and would not be recommended for a sensitive bioassay battery as already concluded for the NNA crude oil. The cytotoxicity in the human cell line (U2-OS) was the most sensitive *in vitro*-based endpoint followed by the cytotoxicity in the fish liver cell line (ZFL) and the micronucleus induction in the fish cells.

VI.3 Comparison of different oil types

In summary, all oil types were more or less responsive in the different acute toxic and mechanism-specific bioassay. In general, no clear conclusion about the most toxic oil type could be drawn (**Table 3**). Furthermore, no individual toxicity profiles for each oil type could be established as all

WAFs showed comparable modes of action. However, at least the trend of the highest toxicity induced by the MGO and the lowest toxicity induced by the heavy fuel oil IFO 180 can be observed as MGO was the most toxic oil type in 6 and IFO the least toxic in 4 out of 11 treatments.
However, the direct comparison between the oil-induced biological effectiveness has to be treated carefully, as the MGO is supplemented with the green dye Dyeguard Green MC25 produced by John Hogg Technical Solutions. So far, the dye could not be investigated in the different toxicity assays and hence a contribution of the dye to the resulting toxicity cannot be excluded.
Furthermore, more data should be included to evaluate the toxicity ranking. At least for some mechanism-specific endpoint used in the present study, data are processed and will be finalized in the near future.

Table 3 Toxicity ranking of different oil types in acute and mechanism-specific bioassays. *Ranking was performed individually on LEWAF and CEWAF treatments of the naphthenic North Sea crude oil (NNA) and the refinded petroleum products MGO and IFO180 used in the GRACE project.*

acute toxicity in whole organisms:	
Lethality of *C. finmarchicus*	NNA > MGO > IFO180 (LEWAF)
	MGO > NNA ≥ IFO180 (CEWAF)
Lethality of *D. rerio*	MGO > IFO180 > NNA (LEWAF)
	MGO > NNA > IFO180 (CEWAF)
Lethality of *Arctia tonsa**	MGO > IFO180 > NNA (LEWAF)
mechanism-specific endpoints:	
cytotoxicity in permanent cell lines (U2-OS, ZFL)	IFO180 ≥ NNA = MGO (LEWAF)
	MGO > NNA = IFO (CEWAF)
micronucleus induction in ZFL	NNA > MGO = IFO180 (LEWAF)
	NNA ≥ MGO > IFO180 (CEWAF)
CYP activity in embryos of *D. rerio (*120 hpf) **	MGO > IFO180 ≥ NNA (LEWAF)
AChE inhibition in embryos of *D. rerio (*120 hpf) **	NNA > MGO > IFO180 (LEWAF)

* not included in statistical data analysis due to limited independent biological replicates
** only LEWAF data available, CEWAF will be finished in near future

VII Suggestion of a sensitive, cost- and time efficient bioassay battery

From the current state of available data and experience in the GRACE project, endpoints for a useful bioassay battery of petroleum product toxicity profiling were established.
With respect to time and cost efficiency a set of the investigated *in vitro*-based bioassays that are validated in international standards (ISO/ OECD guidelines) has been shown to be sensitive towards water-accommodated fraction exposure. One major advantage of the *in vitro*-based

methods that are methodologically and materially optimized for oil toxicity testing (see Johann et al. 2019a in prep) is that they give a final response in maximum 2 to 4 days of total assay procedure depending on the individual bioassay. Furthermore, those assays provide mechanistic insight into the modes of action of a complex environmental sample, and hence are useful screening tools for water samples. Focusing on a genotoxic potential investigated via the induction of mutations (Ames fluctuation assay) or chromosomal aberrations (micronucleus assay) the latter should be preferred over the mutagenic endpoint as the Ames fluctuation assay was not responsive with any WAF treatment independent of a vertebrate metabolic activation system. Compared to micronucleus induction in, e.g., fish erythrocytes, which is often used in crude oil studies, the micronucleus assay using permanent cell lines (e.g. ZFL) is far more sensitive. However, higher micronuclei induction rates are also based on the lack of complete detoxification mechanisms.

The endocrine disruptive potential of water samples is already established in fresh and drinking water quality assessment and increasingly important also in the field of crude oil toxicity testing, as some studies already found adverse effects in wild fish or marine mammal populations indicating an endocrine disruptive potential (e.g. Jobling et al. 2005, Villanger et al. 2011). Furthermore, produced water discharges from offshore oil and gas production platforms have been found to induce estrogenicity contributing to the pollution of the aquatic environment (Thomas et al. 2009, Thomas et al. 2004, Tollefsen et al. 2007). In this respect the yeast and human cell line based reporter gene assays used in the present project are discussed in detail in Johann et al. (2019a). Both assays are very sensitive for detecting ERα interactions (low ng E2-equivaltent/L) with the advantage of the AYES® assay being applicable for elevated salinity conditions and hence being relevant for the investigation of brackish and marine environmental samples. However, both assays have been performed with the crude oil WAFs only. To determine if the assays are differently useful for specific scenarios in petroleum product toxicity testing, the assays have to be performed with the refined petroleum products.

An additional advantage of receptor-mediated bioassays is that the response is quantified in terms of a reference compound making the results easier to compare across a huge set of data that are available for water samples from, e.g., monitoring programs. In this context, also the receptor-mediated Nrf2 CALUX® on oxidative stress seems to be a sensitive screening tool for general stress induction in cells.

Besides a set of useful mechanism-specific endpoints also the acute toxicity in study region-relevant and laboratory model species should be included in a sensitive bioassay battery. Even though semi-static experimental setups (exposure solution renewal, e.g., every 24 h), higher exposure solution volumes as well as longer experiment durations (e.g. 96 h or 120 h) limit the efficiency, those assays provide information on a higher biological organization level. Especially the North Atlantic copepod *Calanus finmarchicus* was highly sensitive towards the petroleum products WAF exposure and is therefore recommended for a bioassay battery. Even though highly sensitive towards the crude oil exposure, *D. magna* is not recommended to be implemented in a

useful bioassay battery, as more relevant and equally sensitive invertebrates for brackish and marine water conditions (e.g. *C. finmarchicus*) were available.

Early life stages of fish have been shown to be very sensitive towards oil compound exposure in the water column. Sublethal effects occurring in embryo/larval development, commonly referred to as blue sac disease, can have adverse effects on organismic fitness in adults (e.g. (Hicken et al. 2011, Mager et al. 2014). As the phenotypic effects (e.g. edema, craniofacial deformations etc.) are consistent across several species (Incardona 2017), the ecotoxicological model species *D. rerio* can be implemented in an effect-based bioassay battery. However, marine fish species have been reported to be far more sensitive towards the exposure to crude oils (e.g. Perrichon et al. 2016). Hence, ongoing experiments associated to the GRACE project on the three-spined stickleback (*Gasterosteus aculeatus*) and the marine medaka (*Oryzias melastigma*) will complete the picture of toxicity in fish early life stages. Based on the direct comparison the best suitable early life stage fish model will be suggested for a sensitive bioassay battery.

In the direct comparison of NNA LEWAF and CEWAF treatments a huge set of mussel biomarker from a semi-chronic exposure experiment (21 days) at two salinity conditions was included. Even though the mussel experiments are not time efficient and a huge operational effort, some biomarkers seem to be highly sensitive for the LEWAF treatment. Hence, endpoints in mussels should not be excluded from a comprehensive profiling, but treatment scenarios could be reduced compared to those selected for the GRACE project.

In general, results of the PCA supported the biological interpretation. The PCA identified several of the aforementioned biological endpoints as statistically redundant because they were clustered in the ordination diagram indicating no additional information when tested in parallel. However, from a biological point of view those clusters contained bioassays focusing on different mode of actions which should be combined in a bioassay battery.

In summary, with the exception of some insensitive endpoints, bioassays on both acute and mechanism-specific toxicity should always be combined in order to comprehensively assess petroleum product toxicity.

VIII Critical considerations for data interpretation

Some aspects regarding experimental setups or data evaluation steps are of crucial importance with respect to the overall data interpretation and hence are addressed in more detail in this section.

One important question in the context of the present toxicity profiling is whether environmentally realistic exposure concentrations in the WAFs were addressed. The WAF stock solutions (1:40, 1:50, 1:200) that have been used in the different laboratories are much above realistic concentrations occurring after an oil spill, as the initial oil slick quickly breaks down to a film of some µm thickness. Initially concentrations of up to 54 mg/L within the top few meters decline

within minutes to concentrations of ≤ 1mg/L (Bejarano et al. 2014). Hence, in this respect the laboratory stocks represented a worst case scenario. However, several sublethal endpoints (e.g. biomarkers in zebrafish embryos, micronucleus induction, etc.) were investigated at highly diluted stock concentrations which are more representative for environmental scenarios. In this respect the biomarkers in zebrafish embryos were investigated at 12.5 % (LEWAF) and 0.78 % (CEWAF) of the stocks (around EC10) which is corresponding to 1:400 (2.5 g/L) and 1:25600 (40 mg/L) dilutions, respectively. Furthermore, it is almost impossible to reproduce a realistic scenario under laboratory conditions as each oil spill is unique depending on a variety of environmental conditions (water temperature, salinity, natural dispersion due to waves, etc.) and the characteristics of the spilled oil (physical-chemical properties like composition, weathering status, viscosity, etc.). Additionally, the mechanism-specific endpoints reflect acute exposure scenarios up to a maximum of 2 days leading to the assumption that chronic exposure would result in higher effects. However, the extrapolation to higher biological organization levels due to the lack of toxicokinetic implementation in *in vitro* based assays is limited.

In general, establishing a classification system was the most critical aspect of the present study, as it should include expert judgement already in the first step of data processing, and with this avoid over- or underestimation of the detected responses. The data transformation in effect classes between 1 and 5 might mask slight differences between treatment groups that would be visible in the raw data. As an example, the classified acute toxicity in zebrafish embryos was in a comparable range for all oil type LEWAFs but from the raw data a trend of MGO being more toxic towards the embryos could be observed. Hence, a finer graduation of the class scale would further improve the classification system (e.g. 10 classes). However, for some bioassays a finer graduation would artificially inflate the responses. As an example, micronucleus induction rates of ≥ 3.5 are already in the range of the response of positive control substances that are known for their high genotoxic potential. Establishing, e.g., 10 classes between IF 1 and 3.5 would hence result in biologically less relevant nuances.

Importantly, the classification was established individually for each bioassay taking into account several assay-specific aspects including baseline activity, thresholds and reference effects in order to generate data of one common scale. One major problem of including expert knowledge on adverse effects induced by other petroleum products was the limited comparability to previous studies. As reviewed recently, oil toxicity tests showed a high diversity of reported test methods that affect composition, stability and toxicity of exposure solutions leading to a high variance in adverse effects (Hodson et al. 2019). Furthermore, not only the comparison to previous studies but also the comparison between results obtained in the present project was limited due to different experimental setups. Partially, differences could not be avoided due to for example divergent dimensions of stock solution needed or endpoint specific requirements. XI Appendix

Table 4 (appendix) shows the variability of experimental setups used in the different laboratories. Varying stock solutions (oil:water proportion) might result in WAFs with different dissolved hydrocarbon contents as the equilibrium might or might not be reached after the 40 h incubation. Also different medium salinities, aspirator bottles with different head space, different dispersants (Finasol OSR 51 and Finasol OSR 52) and dispersed crude oil mixing energies (with/without 25 % vortex) have been used in the laboratories. However, main aspects of WAF preparation procedures like the water temperature, dispersant:oil ratio (DOR), incubation and settling time were consistent across the laboratories. Importantly, temperature has been shown to have a much greater impact on the solubility of, e.g., PAHs compared to salinity (Eganhouse and Calder 1976, Saranjampour et al. 2017, Whitehouse 1984, Xie et al. 1997). Furthermore, some differences could not be avoided due to experimental requirements (e.g., exposure solution volumes, salinity conditions due to target organisms). Additionally, first results indicate no major differences for dissolved fractions of PAHs across the different laboratories.

Besides inter- and intra-laboratory limitations of comparability also the class limit definition was critical for selected bioassays. While some bioassay classifications included mathematical aspects (e.g. receptor-mediated *in vitro* assays: 1* trigger value, 3* trigger value, 10* trigger value) others were mainly based on expert knowledge (e.g. micronucleus induction in ZF-L).

Importantly, it has to be considered that especially the results available for the refined petroleum products are limited and more data should be included to guarantee a better comparability.

Taking all the limitations and discussed aspects into account, the present classification is a first suggestion that can simply be modified due to optimization steps. In summary, already at this state the suggested bioassay battery is highly sensitive and relevant when interpreted within the discussed limitations.

IX Conclusion and outlook

As shown in the present study several bioassays that have been optimized for the complex mixture of petroleum product WAFs are useful screening tools for an oil toxicity profiling. The final analysis was able to distinguish between more and less sensitive bioassays. A combination of bioassays on different biological organization levels is highly recommended as chemical analysis or single bioassays cannot display the complex picture of environmental samples comprehensively. Namely, the acute toxic effects towards sensitive laboratory and study region-relevant invertebrate and vertebrate species should be combined with small-scale *in vitro* based bioassays on defined modes of actions. Based on the available dataset no oil type-specific toxicity profiles could be derived which might be related to a relatively consistent composition of dissolved hydrocarbon fractions in the WAFs of which mainly the lower molecular weight PAHs are discussed to be accountable for acute and sub-chronic effects in biota.

The application of dispersants as an oil spill response action is discussed controversially. The complete removal of oil by mechanical recovery systems is difficult if not impossible due to, e.g., physical limitations of the mechanical skimmer equipment (Lee et al. 2011, Prince 2015). Hence, based on net environmental benefit analysis spill response coordinators have to decide whether the benefit of dispersant usage overweighs an additional damage that could be caused by its application. It is a fact that the application introduces even more chemicals into an already impacted environment. The present discussion also supports the general statement that the water column under the freshly dispersed oil slick is significantly more toxic to organisms, probably due to the higher concentration of oil droplets in the water column in combination with a higher bioavailability of oil and dissolved oil constituents for organisms (Dussauze et al. 2015, Prince 2015), but also as a result of toxicity of dispersant components. Consequently, ecotoxicological effect data from individual bioassays of the present report can contribute to oil spill response planning tools such as the EOS (Environment and oil spill response) tool established in the GRACE project.

Further research should focus on additional endpoints, e.g., biomarker responses in the highly sensitive and relevant species *Calanus finmarchicus*, in order to contribute to an oil toxicity profiling battery. In fact, some additional endpoints presented for NNA will be finalized also for the refined petroleum products and further experiments (e.g. with adult zebrafish, stickleback and medaka embryos) are being processed and will finally contribute to the big picture of the toxicity profile. Additionally, the classification system should be critically discussed in the scientific community in order to improve data evaluation.

X References

Ames, B.N., Lee, F.D. and Durston, W.E. (1973) An improved bacterial test system for the detection and classification of mutagens and carcinogens. Proceedings of the National Academy of Sciences 70(3), 782-786.

b.v., B.S. (2017) standard operating procedures P-BDS-050 to P-BDS-085.

Barron, M.G., Heintz, R. and Rice, S.D. (2004) Relative potency of PAHs and heterocycles as aryl hydrocarbon receptor agonists in fish. Marine environmental research 58(2-5), 95-100.

Bejarano, A.C., Clark, J.R. and Coelho, G.M. (2014) Issues and challenges with oil toxicity data and implications for their use in decision making: A quantitative review. Environmental toxicology and chemistry 33(4), 732-742.

Beliaeff, B. and Burgeot, T. (2002) Integrated biomarker response: a useful tool for ecological risk assessment. Environmental Toxicology and Chemistry: An International Journal 21(6), 1316-1322.

Berridge, M.V., Herst, P.M. and Tan, A.S. (2005) Tetrazolium dyes as tools in cell biology: new insights into their cellular reduction. Biotechnology annual review 11, 127-152.

Couillard, C.M., Lee, K., Légaré, B. and King, T.L. (2005) Effect of dispersant on the composition of the water-accommodated fraction of crude oil and its toxicity to larval marine fish. Environmental toxicology and chemistry 24(6), 1496-1504.

Countryman, P.I. and Heddle, J.A. (1976) The production of micronuclei from chromosome aberrations in irradiated cultures of human lymphocytes. Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis 41(2), 321-331.

Dupuis, A. and Ucán-Marín, F. (2015) A literature review on the aquatic toxicology of petroleum oil: An overview of oil properties and effects to aquatic biota, Canadian Science Advisory Secretariat.

Dussauze, M., Pichavant-Rafini, K., Le Floch, S., Lemaire, P. and Theron, M. (2015) Acute toxicity of chemically and mechanically dispersed crude oil to juvenile sea bass (Dicentrarchus labrax): Absence of synergistic effects between oil and dispersants. Environmental toxicology and chemistry 34(7), 1543-1551.

Eganhouse, R.P. and Calder, J.A. (1976) The solubility of medium molecular weight aromatic hydrocarbons and the effects of hydrocarbon co-solutes and salinity. Geochimica et Cosmochimica Acta 40(5), 555-561.

Escher, B.I., Allinson, M., Altenburger, R., Bain, P.A., Balaguer, P., Busch, W., Crago, J., Denslow, N.D., Dopp, E. and Hilscherova, K. (2013) Benchmarking organic micropollutants in wastewater, recycled water and drinking water with in vitro bioassays. Environmental science & technology 48(3), 1940-1956.

Fent, K. (2001) Fish cell lines as versatile tools in ecotoxicology: assessment of cytotoxicity, cytochrome P4501A induction potential and estrogenic activity of chemicals and environmental samples. Toxicology in vitro 15(4-5), 477-488.

Fotakis, G. and Timbrell, J.A. (2006) In vitro cytotoxicity assays: comparison of LDH, neutral red, MTT and protein assay in hepatoma cell lines following exposure to cadmium chloride. Toxicology letters 160(2), 171-177.

Gonzalez, R. and Tarloff, J. (2001) Evaluation of hepatic subcellular fractions for Alamar blue and MTT reductase activity. Toxicology in vitro 15(3), 257-259.

Grote, M., Nagel, A., Nies, H., Rauterberg, J. and Wahrendorf, D.-S. (2016) The use of dispersants to combat oil spills in Germany at sea. Federal Institute for Risk Assessment. Press and Public Relations. ISSN 1614-3795 (Print) 1614-3841 (Online).

Hansen, B.H., Salaberria, I., Read, K.E., Wold, P.A., Hammer, K.M., Olsen, A.J., Altin, D., Øverjordet, I.B., Nordtug, T. and Bardal, T. (2019) Developmental effects in fish embryos exposed to oil dispersions–The impact of crude oil micro-droplets. Marine environmental research, 104753.

Hettwer, K., Jähne, M., Frost, K., Giersberg, M., Kunze, G., Trimborn, M., Reif, M., Türk, J., Gehrmann, L. and Dardenne, F. (2018) Validation of Arxula Yeast Estrogen Screen assay for detection of estrogenic activity in water samples: results of an international interlaboratory study. Science of The Total Environment 621, 612-625.

Hicken, C.E., Linbo, T.L., Baldwin, D.H., Willis, M.L., Myers, M.S., Holland, L., Larsen, M., Stekoll, M.S., Rice, S.D. and Collier, T.K. (2011) Sublethal exposure to crude oil during embryonic development alters cardiac morphology and reduces aerobic capacity in adult fish. Proceedings of the National Academy of Sciences 108(17), 7086-7090.

Hodson, P.V., Adams, J. and Brown, R.S. (2019) Oil toxicity test methods must be improved. Environmental toxicology and chemistry 38(2), 302-311.

Incardona, J.P. (2017) Molecular mechanisms of crude oil developmental toxicity in fish. Archives of environmental contamination and toxicology 73(1), 19-32.

Incardona, J.P., Day, H.L., Collier, T.K. and Scholz, N.L. (2006) Developmental toxicity of 4-ring polycyclic aromatic hydrocarbons in zebrafish is differentially dependent on AH receptor isoforms and hepatic cytochrome P4501A metabolism. Toxicology and applied pharmacology 217(3), 308-321.

ISO (2004) ISO/DIS 21427-2 Water quality - evaluation of the genotoxicity by measurements of the induction of micronuclei.

ISO (2012) ISO 11350 Water quality - determination of genotoxicity of water and waster water using the *Salmonella*/microsome fluctuation test. Ames fluctuation test.

ISO (2014) ISO/CD 19040-3 Water quality- Determination of the estrogenic potential of water and waste water - Part 3: In vitro human cell-based reporter gene assay (ISO/TC 147/SC 5N 864.

ISO (2018) ISO/FDIS 19040-2:2018(E) Draft: Water quality — Determination of the estrogenic potential of water and waste water —Part 2: Yeast estrogen screen (A-YES, *Arxula adeninivorans*).

Jobling, S., Williams, R., Johnson, A., Taylor, A., Gross-Sorokin, M., Nolan, M., Tyler, C.R., van Aerle, R., Santos, E. and Brighty, G. (2005) Predicted exposures to steroid estrogens in UK rivers correlate with widespread sexual disruption in wild fish populations. Environmental health perspectives 114(Suppl 1), 32-39.

Johann, S., Esser, M., Nüßer, L., Hollert, H. and Seiler, T.-B. (2019a) in preparation: Receptor-mediated bioassays for estrogenicity - applications in petroleum product toxicity assessment.

Johann, S., Nüßer, L., Goßen, M., Hollert, H., Seiler, T.-B., Esteban, A. and Orbea, A. (2019b) Adverse outcome links for zebrafish - public report D3.12 of the EU project GRACE (Integrated oil spill response actions and environmental effects) Contract No 679266 Horizon 2020 BG-2014-2015/BG2015-2.

Kamber, M., Flückiger-Isler, S., Engelhardt, G., Jaeckh, R. and Zeiger, E. (2009) Comparison of the Ames II and traditional Ames test responses with respect to mutagenicity, strain specificities, need for metabolism and correlation with rodent carcinogenicity. Mutagenesis 24(4), 359-366.

Lee, K., Nedwed, T. and Prince, R.C. (2011) Lab tests on the biodegradation rates of chemically dispersed oil must consider natural dilution, p. abs245, American Petroleum Institute.

Legler, J., Dennekamp, M., Vethaak, A.D., Brouwer, A., Koeman, J.H., Van Der Burg, B. and Murk, A.J. (2002) Detection of estrogenic activity in sediment-associated compounds using in vitro reporter gene assays. Science of The Total Environment 293(1-3), 69-83.

Lekube, X., Aguirre-Rubí, J., Izagirre, U., Soto, M., Bilbao, D., Etxebarria, N., Marigómez, a.I., Ahvo, A., Turja, R. and (SYKE), K.L. (2019) Report, baseline and database on influence of latitude and season on adverse outcome links in northern Atlantic and Baltic mussels after oil spills and oil spill responses - public report D3.11 of the EU project GRACE (Integrated oil spill response actions and environmental effects) Contract No 679266 Horizon 2020 BG-2014-2015/BG2015-2

Ma, Q. (2013) Role of nrf2 in oxidative stress and toxicity. Annual review of pharmacology and toxicology 53, 401-426.

Mager, E.M., Esbaugh, A.J., Stieglitz, J.D., Hoenig, R., Bodinier, C., Incardona, J.P., Scholz, N.L., Benetti, D.D. and Grosell, M. (2014) Acute embryonic or juvenile exposure to Deepwater Horizon

crude oil impairs the swimming performance of mahi-mahi (Coryphaena hippurus). Environmental science & technology 48(12), 7053-7061.

Maron, D.M. and Ames, B.N. (1983) Revised methods for the Salmonella mutagenicity test. Mutation Research/Environmental Mutagenesis and Related Subjects 113(3-4), 173-215.

Mosmann, T. (1983) Rapid colorimetric assay for cellular growth and survival: application to proliferation and cytotoxicity assays. Journal of immunological methods 65(1-2), 55-63.

Perrichon, P., Le Menach, K., Akcha, F., Cachot, J., Budzinski, H. and Bustamante, P. (2016) Toxicity assessment of water-accommodated fractions from two different oils using a zebrafish (Danio rerio) embryo-larval bioassay with a multilevel approach. Science of The Total Environment 568, 952-966.

Perrichon, P., Mager, E.M., Pasparakis, C., Stieglitz, J.D., Benetti, D.D., Grosell, M. and Burggren, W.W. (2018) Combined effects of elevated temperature and Deepwater Horizon oil exposure on the cardiac performance of larval mahi-mahi, Coryphaena hippurus. PloS one 13(10), e0203949.

Prince, R.C. (2015) Oil spill dispersants: boon or bane? Environmental science & technology 49(11), 6376-6384.

Prince, R.C., McFarlin, K.M., Butler, J.D., Febbo, E.J., Wang, F.C. and Nedwed, T.J. (2013) The primary biodegradation of dispersed crude oil in the sea. Chemosphere 90(2), 521-526.

R core team (2019) R: A language and environment for statisticical computing. R Foundation for Statistical Computing. Vienna, Austria.

Ramachandran, S.D., Hodson, P.V., Khan, C.W. and Lee, K. (2004) Oil dispersant increases PAH uptake by fish exposed to crude oil. Ecotoxicology and environmental safety 59(3), 300-308.

Redman, A.D. and Parkerton, T.F. (2015) Guidance for improving comparability and relevance of oil toxicity tests. Marine pollution bulletin 98(1-2), 156-170.

Reifferscheid, G., Dill, F., Fieblinger, D., Gminski, R., Grummt, H.-J., Hafner, C., Hollert, H., Kunz, S., Rodrigo, G. and Stopper, H. (2007) Untersuchung von Abwasserproben auf Gentoxizität. Umweltwissenschaften und Schadstoff-Forschung 19(1), 7-16.

Reifferscheid, G., Maes, H., Allner, B., Badurova, J., Belkin, S., Bluhm, K., Brauer, F., Bressling, J., Domeneghetti, S. and Elad, T. (2012) International round-robin study on the Ames fluctuation test. Environmental and molecular mutagenesis 53(3), 185-197.

Saranjampour, P., Vebrosky, E.N. and Armbrust, K.L. (2017) Salinity impacts on water solubility and n-octanol/water partition coefficients of selected pesticides and oil constituents. Environmental toxicology and chemistry 36(9), 2274-2280.

Singer, M., Aurand, D., Bragin, G., Clark, J., Coelho, G., Sowby, M. and Tjeerdema, R. (2000) Standardization of the preparation and quantitation of water-accommodated fractions of petroleum for toxicity testing. Marine pollution bulletin 40(11), 1007-1016.

Sonneveld, E., Jansen, H.J., Riteco, J.A., Brouwer, A. and van der Burg, B. (2004) Development of androgen-and estrogen-responsive bioassays, members of a panel of human cell line-based highly selective steroid-responsive bioassays. Toxicological Sciences 83(1), 136-148.

Sonneveld, E., Riteco, J.A., Jansen, H.J., Pieterse, B., Brouwer, A., Schoonen, W.G. and van der Burg, B. (2005) Comparison of in vitro and in vivo screening models for androgenic and estrogenic activities. Toxicological Sciences 89(1), 173-187.

Thomas, K., Langford, K., Petersen, K., Smith, A. and Tollefsen, K. (2009) Effect-directed identification of naphthenic acids as important in vitro xeno-estrogens and anti-androgens in North Sea offshore produced water discharges. Environmental science & technology 43(21), 8066-8071.

Thomas, K.V., Balaam, J., Hurst, M.R. and Thain, J.E. (2004) Identification of in vitro estrogen and androgen receptor agonists in North Sea offshore produced water discharges. Environmental Toxicology and Chemistry: An International Journal 23(5), 1156-1163.

Tollefsen, K.-E., Harman, C., Smith, A. and Thomas, K.V. (2007) Estrogen receptor (ER) agonists and androgen receptor (AR) antagonists in effluents from Norwegian North Sea oil production platforms. Marine pollution bulletin 54(3), 277-283.

Total (2012a) Material Safety Data sheet FINASOL OSR 51 SDS#: 30033 Version 1.01

Total (2012b) Safety Data sheet FINASOL OSR 52 SDS#: 30034 Version 7.02

van der Linden, S.C., von Bergh, A.R., van Vught-Lussenburg, B.M., Jonker, L.R., Teunis, M., Krul, C.A. and van der Burg, B. (2014) Development of a panel of high-throughput reporter-gene assays to detect genotoxicity and oxidative stress. Mutation Research/Genetic Toxicology and Environmental Mutagenesis 760, 23-32.

Van der Oost, R., Beyer, J. and Vermeulen, N.P. (2003) Fish bioaccumulation and biomarkers in environmental risk assessment: a review. Environmental toxicology and pharmacology 13(2), 57-149.

Villanger, G., Lydersen, C., Kovacs, K., Lie, E., Skaare, J. and Jenssen, B. (2011) Disruptive effects of persistent organohalogen contaminants on thyroid function in white whales (Delphinapterus leucas) from Svalbard. Science of The Total Environment 409(13), 2511-2524.

Whitehouse, B.G. (1984) The effects of temperature and salinity on the aqueous solubility of polynuclear aromatic hydrocarbons. Marine Chemistry 14(4), 319-332.

Wickham Hadley (2017) Tidyverse: EasilyInstall and Load the “Tidyverse” R package version 1.2.1 http://CRAN.R-project.org/package=tidyverse

Xie, W.-H., Shiu, W.-Y. and Mackay, D. (1997) A review of the effect of salts on the solubility of organic compounds in seawater. Marine environmental research 44(4), 429-444.

Weblinks:

Total (2019) http://totalspecialfluids.com / 11.07.19 7:03

XI Appendix

Table 4 Overview of different WAF approaches used in WP3 of GRACE. All data shown were included in the analysis of the present report.

Endpoints (laboratory)	WAF approach	Dispersant	Medium	Water:oil ratio (stock) (v:w)	DOR (w/w)	Incubation temperature [°C]	Incubation time [h]	Mixing energy
Acute toxicity in *D. rerio* (RWTH)	LEWAF		Artificial fish medium (freshwater)	1:50		10	40	low (avoid vortex)
	CEWAF	Finasol OSR 51 (NNA crude oil) Finasol OSR 52 (MGO, IFO180)	Artificial fish medium (freshwater)	1:200	1:10	10	40	high (25% vortex)
Acute toxicity in *D. magna* (RWTH)	LEWAF		Artificial fish medium (freshwater)	1:50		10	40	low (avoid vortex)
	CEWAF	Finasol OSR 51 (NNA crude oil)	Artificial fish medium (freshwater)	1:200	1:10	10	40	high (25% vortex)
In vitro based endpoints on cytotoxicity (MTT), endocrine disruption (ERα-CALUX®, A-YES®),oxidative stress (Nrf2-CALUX®), genotoxicity (MNC, Ames Flu) (RWTH)	LEWAF		Deionized water / Artificial brackish water (6 ‰)	1:50		10	40	low (avoid vortex)
	CEWAF	Finasol OSR 51 (NNA crude oil) Finasol OSR 52 (MGO, IFO180)	Deionized water / Artificial brackish water (6 ‰)	1:200	1:10	10	40	high (25% vortex)

Acute toxicity in *C. finmarchicus* (NTNU)	LEWAF			1:40	1:10	10	72	low (avoid vortex)
	CEWAF	Finasol OSR 52 (all oil types)		1:40	1:10	10	72	high (25% vortex)
Acute toxicity in *P. Lividus* (UPV/EHU)	LEWAF		Seawater (29 ‰)	1:200	1:10	10	40	low (avoid vortex)
	LEWAF+D	Finasol OSR 52 (all oil types)	Seawater (29 ‰)	1:200	1:10	10	40	low (avoid vortex)
In vitro based endpoints in *M. galloprovicialis* hemocytes (UPV/EHU)	LEWAF		Seawater (33 ‰)	1:200	1:10	10	40	low (avoid vortex)
	LEWAF+D	Finasol OSR 52 (NNA crude oil)	Seawater (33 ‰)	1:200	1:10	10	40	low (avoid vortex)
Biomarker in *M. trossulus* (SYKE)	LEWAF		Brackish water (5.6 and 15 ‰)	1:200	1:10	10	40	low (avoid vortex)
	WAF-D	Finasol OSR 51 (NNA crude oil)	Brackish water (5.6 and 15 ‰)	1:200	1:10	10	40	low (avoid vortex)

Table 5 Background information on the classification of effect data.

Endpoint	Classification basis	Background
Endocrine disruption (ERα-CALUX®, A-YES®)	Effect-based trigger values (EBT) biological reference activities in water bodies effect extrapolation to population	EBT ERα-CALUX®: 0.1 ng E2/L EBT A-YES®: 0.56 ngE2/L [1] class limits based on:<EBT; 3* EBT; 10*EBT; 100*EBT EEQ range wastewater effluents: <LOD – 100 ngE2/L [2, 3] Fish population decline: 5-6 ng EE2/L [4] EQS surface waters inland: E2: 0.4 ng/L EE2: 0.035 ng/L References on petroleum products: [5, 6]
Oxidative stress (Nrf2-CALUX®)	Effect-based trigger values (EBT)	EBT Nrf2-CALUX®: 21 µg Curcumin/L (26 µg Dichlorvos/L) [1] class limits based on:<EBT; 3* EBT; 10*EBT; 100*EBT
Micronucleus induction (MNC) in ZF-L cells	Induction factor (IF): response corrected for untreated control (NC) Baseline on historical MNC data in ZF-L, (V79, RTLW-1) cell line Biological effects in reference studies (IFs)	Baseline genotoxic substances: ZF-L: 4-NQO: IF = 3.54 (± 0,85) V79: CPP: IF = 3-15 (own data + [7]) V79: EMS: IF = 2-12 (own data + [7]) RTLW-1: 4-NQO: IF = 3.5 [8] References on petroleum products: [8-10]
Cytotoxicity in different cell lines (		

Mutagenicity in Ames Flu assay	# of revertants
	Baseline on historical Ames Flu data (tester stain and approach specific (+/-S9))
	Biological effects in reference studies
		Threshold # revertants: <10 (NC); > 25 (PC) [11]
		Threshold= 0.4 ([11]
		Own historical data (# revertants):
		TA 98 +S9:
		NC: 1.56 ± 1.83
		PC: 41.79 ± 9.7
		TA 98 -S9:
		NC: 2.47 ± 2.47
		PC: 41.8 ± 5.57
		TA 100 +S9:

		NC: 3.18 ± 3.28 PC: 43.13 ± 5.33 TA 100 +S9: NC: 4.02 ± 2.64 PC: 45.93 ± 1.71 References water effluents: [12] References on petroleum products: [9, 13-16]
Acute toxicity in model and regional relevant species	Baseline on historical acute toxicity data and validity criteria	Validity criteria *D. magna* and *D. rerio*: [17, 18] NC < 10% effects PC > 30 % effects (*D. rerio*)
Biomarker in *D. rerio*	Induction factor (IF): response corrected for untreated control (NC) Biological effects in reference studies	Own data on PAHs, PCBs: IF FE-ERODmax = 3 – 5 Reference on petroleum products: [19]
Biomarker in *M. trossulus*	Individual control fluctuations (CT)	Mean control of the individual biomarker and treatment with strandard deviation criteria above and below CT_{mean} class limits based on: CT_{mean} +/- 0,5; +/- 1; +/-1.5; +/- 2

1. Escher, B.I., et al., Effect-based trigger values for in vitro and in vivo bioassays performed on surface water extracts supporting the environmental quality standards (EQS) of the European Water Framework Directive. Science of The Total Environment, 2018. 628: p. 748-765.
2. Leusch, F.D., et al., Comparison of five in vitro bioassays to measure estrogenic activity in environmental waters. Environmental science & technology, 2010. 44(10): p. 3853-3860.
3. Jarošová, B., et al., What level of estrogenic activity determined by in vitro assays in municipal waste waters can be considered as safe? Environment international, 2014. 64: p. 98-109.
4. Kidd, K.A., et al., Collapse of a fish population after exposure to a synthetic estrogen. Proceedings of the National Academy of Sciences, 2007. 104(21): p. 8897-8901.
5. Vrabie, C.M., et al., Specific in vitro toxicity of crude and refined petroleum products: II. Estrogen (α and β) and androgen receptor-mediated responses in yeast assays. Environmental toxicology and chemistry, 2010. 29(7): p. 1529-1536.
6. Villeneuve, D., et al., Relative potencies of individual polycyclic aromatic hydrocarbons to induce dioxinlike and estrogenic responses in three cell lines. Environmental Toxicology: An International Journal, 2002. 17(2): p. 128-137.

7. Reifferscheid, G., et al., Untersuchung von Abwasserproben auf Gentoxizität. Umweltwissenschaften und Schadstoff-Forschung, 2007. 19(1): p. 7-16.
8. Brinkmann, M., et al., Genotoxicity of heterocyclic PAHs in the micronucleus assay with the fish liver cell line RTL-W1. PloS one, 2014. 9(1): p. e85692.
9. Bluhm, K., et al., Genotoxicity of three biofuel candidates compared to reference fuels. Environmental toxicology and pharmacology, 2018. 64: p. 131-138.
10. Lachner, D., L.F. Oliveira, and C.B. Martinez, Effects of the water soluble fraction of gasoline on ZFL cell line: Cytotoxicity, genotoxicity and oxidative stress. Toxicology in Vitro, 2015. 30(1): p. 225-230.
11. Reifferscheid, G., et al., International round-robin study on the Ames fluctuation test. Environmental and molecular mutagenesis, 2012. 53(3): p. 185-197.
12. Leusch, F.D., et al., Assessment of the application of bioanalytical tools as surrogate measure of chemical contaminants in recycled water. Water research, 2014. 49: p. 300-315.
13. Pérez, A.S., et al., Assessment of the mutagenic potency of sewage sludges contaminated with polycyclic aromatic hydrocarbons by an Ames fluctuation assay. Environmental Toxicology and Chemistry: An International Journal, 2003. 22(11): p. 2576-2584.
14. Petrilli, F., G.D. Renzi, and S.D. Flora, Interaction between polycyclic aromatic hydrocarbons, crude oil and oil dispersants in the Salmonella mutagenesis assay. Carcinogenesis, 1980. 1(1): p. 51-56.
15. Lockard, J., et al., Comparative study of the genotoxic properties of Eastern and Western US shale oils, crude petroleum, and coal-derived oil. Mutation Research/Genetic Toxicology, 1982. 102(3): p. 221-235.
16. Smith, K.E., et al., The dosing determines mutagenicity of hydrophobic compounds in the Ames II assay with metabolic transformation: Passive dosing versus solvent spiking. Mutation Research/Genetic Toxicology and Environmental Mutagenesis, 2013. 750(1-2): p. 12-18.
17. OECD, Fish embryo acute toxicity (FET) test OECD guideline for testing of chemicals No 236 2013.
18. OECD, OECD 202 Daphnia sp., Acute Immobilisation Test. OECD guideline for testing of chemicals No 202, 2004.
19. Bräunig, J., et al., Time-dependent expression and activity of cytochrome P450 1s in early life-stages of the zebrafish (Danio rerio). Environmental Science and Pollution Research, 2015. 22(21): p. 16319-16328.