Supporting Information to

# Safe and Sustainable chemicals by design: a computer-based approach to redesign for reduced environmental hazards

Joanke van Dijk[1,2*,#], Hannah Flerlage[2,3,#], Steven Beijer[3], J. Chris Slootweg[3], Annemarie P. van Wezel[2]

[1] Copernicus Institute of Sustainable Development, Utrecht University, 3584 CB Utrecht, the Netherlands

[2] Institute for Biodiversity and Ecosystem Dynamics, University of Amsterdam, PO Box 94240, 1090 GE Amsterdam, the Netherlands

[3] Van 't Hoff Institute for Molecular Sciences, University of Amsterdam, PO Box 94157, 1090 GD Amsterdam, the Netherlands

*Corresponding author: j.vandijk3@uu.nl

[#] Contributed equally

# Contents

# 1 Methods

## 1.1 Selection and applicability of QSAR models

In this study, the potential persistency (P), bioaccumulation (B), mobility (M) and toxicity (T) of structures were assessed. These parameters were selected as they are used to identify problematic substances under REACH (Regulation (EC) No. 1907/2006). For the selection of QSAR models, only models were considered that are freely available, able to be used for large sets of structures and that are currently included in Appendix 1 of the Practical Guide - How to use and report (Q)SARs (ECHA, 2016).

The applicability domain of a (Q)SAR model is the response and chemical structure space in which the model makes predictions with a given reliability, and therefore defined by the nature of the chemicals in the training set (Gramatica et al., 2012; Netzeva et al., 2005). In table S1 the structural and physicochemical parameters used to define the QSAR applicability domains according to the OECD Guidance Document (OECD, 2007) are shown. This guidance is also in accordance with the recommended way of setting the applicability domain as described in the EPISUITE manuals (US EPA, 2012). Models included in VEGA for the prediction of persistency, carcinogenicity and ecotoxicity were not used because the applicability was evaluated to be unsatisfactory. While LogKow and molecular weight (MW) ranges are met for most of these models, the training sets only contain few organophosphates or none in the case of the QSARs for persistency. For carcinogenicity and ecotoxicity models, only few organophosphates were found of which many contained halogens and aromatic side groups, which are mechanistically relevant for these rather complex toxicity endpoints and could lead to a misclassification of organophosphates. Ecotoxicity QSAR Ecosar which is included in EPISUITE does not facilitate the high-throughput prediction of *in silico* generated structures needed for our approach. Thus, persistency was evaluated based on LogKow and biodegradability and toxicity based on mutagenicity and endocrine disruption.

*Table S1. Log Kow and MW range and number of organophosphate structures, number of P containing structures and total number of structures in the training /validation datasets of the used QSAR models.*

| Attribute | Endpoint kind | Model | Log Kow range (KOWWIN) (exp) | MW range [g mol$^{-1}$] | n of Organophosphates | n of P containing structures | Total n of structures |
|---|---|---|---|---|---|---|---|
| **P** | Biodegradability<br>Mix of expert judgement and experiments | EpiSuite: BIOWIN, Aerobic: Biowin1-6, anaerobic: Biowin7 | -4.2 - 8.65 | 30 - 959 | 20 | 45 | 1263 |
| | Log Kow<br>Experiments | EpiSuite: KOWWIN | -11.96 - 10.2 | 4 - 1203 | 66 | 362 | 15809 |
| **B** | BCF<br>Factor/ experiment | EpiSuite BCFBAF | -4.5 - 8.6 | 68 - 992 | 21 | 60 | 685 |
| **M** | Log Koc<br>Experiments | VEGA: OPERA KOC | -5.4 - 8.68 | 32 - 665 | 9 | 68 | 729 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | EpiSuite: KOCWIN | -5.4 - 8.68 | 32 - 1053 | 13 | 74 | 788 |
| **T** | Mutagenicity<br><br>Experiment (Ames test) | VEGA: Mutagenicity consensus model - CAESAR | -4.22 - 8.65 | 30 - 3080 | 42 | 102 | 4204 |
| | | VEGA: Mutagenicity consensus model - ISS | -3.7 - 7.6 | 30 -1255 | 8 | 29 | 670 |
| | | VEGA: Mutagenicity consensus model - KNN | -5.4 - 8.65 | 28 -1550 | 42 | 117 | 5770 |
| | | VEGA: Mutagenicity consensus model - SARPY | -4.22 - 8.65 | 30 - 3080 | 42 | 102 | 4204 |
| | Endocrine disruption<br><br>Expert judgement / experiments | VEGA: Androgen Compara | -3.7 - 8.65 | 42 - 1701 | 22 | 73 | 1664 |
| | | VEGA: Estrogen Cerapp | -3.7 - 8.1 | 42 - 973 | 22 | 64 | 1529 |
| | | VEGA: TRALPHA NRMEA | -5.08 - 9.05 | 31 - 1638 | 59 | 188 | 5462 |
| | | VEGA: TRBETA NRMEA | -5.08 - 9.05 | 31 - 1638 | 59 | 188 | 5487 |
| | | | Log Kow range (KOWWIN) (exp) | MW range<br><br>[g $mol^{-1}$] | n of Organophosphates | n of P containing structures | Total n of structures |
| **Database of generated structures** | | | -6.26 – 6.27 (predicted values) | 180 - 336 | all | all | 32350 |

## 1.2 Testing of target compound

### 1.2.1 LC-MS

The LC-MS spectrum of the target compound was measured with an ultrahigh-performance LC system (Nexera Shimadzu, Den Bosch, The Netherlands) coupled to a maXis 4G high resolution quadrupole time-of-flight HRMS (q-ToF/HRMS) with an added HD collision cell ($N_2$) and an electrospray ionization (ESI) source (Bruker Daltonics, Wormer, The Netherlands). The used chromatography column was a polar reversed-phase core–shell Kinetex biphenyl LC column having 1.7 µm particle size, pore size of 100 Å and dimensions of 150 × 2.1 mm (Phenomenex, Utrecht, The Netherlands). The column was heated at 40 °C. For the mobile phase, pure $H_2O$ (A) and MeOH acidified with 0.05% acetic acid (B) were used at a flow rate of 0.3 mL/min. The LC gradient program was 0% B, 100% A from 0 to 2 min and reached 100% B and 0% A at 17 min. From 17 to 25 min the eluent was kept at 100% B. 20 µL of the sample solution with a concentration of target I of 500 µg $L^{-1}$ in 95% water and 5 % methanol (volume) were injected. The autosampler had a temperature of 15 °C. The MS detector with ESI source was internally calibrated prior to the start of the analysis by infusing a 50 µM sodium acetate solution in H2O:MeOH (1:1, volume) with a loop injection of 20 µL and a loop rinse of 20 µL. A spray voltage of +3.5 kV was used for the positive ESI mode with a

resolving power of 30,000–60,000 at full width at half maximum (FWHM). Nitrogen was used as curtain gas to lose neutral compounds. The capillary temperature was 300 °C. MS/MS spectra were recorded in data-dependent acquisition mode with a minimum resolving power of 20,000 at FWHM.

### 1.2.2 CALUX assays

CALUX® assays were performed by BioDetection Systems (BDS), Amsterdam, the Netherlands in order to identify hormone-like properties. The assays were performed according to a similar protocol earlier described by Besselink et al. (2004), to identify both agonistic and antagonistic effects on the estrogen receptor (ER alpha), androgen receptor (AR), progesterone receptor (PR) and thyroid receptor (TR beta). Firstly, cytotoxic effects of TiBP and the target compound on BDS cell line were tested in order to determine suitable concentration for the receptor assays.

For the CALUX assays, cells were cultured in 384 well plates. Subsequently, the cells were exposed to a dilution series of TiBP and the Target Compound. The test was performed in triplicate. Cells were also exposed to a concentration series of a reference compound. After exposure, light production in the wells were quantified. The activity evoked by the TiBP and the Target Compound were derived by interpolation in the response curve of the reference compound.

Additionally, Nrf2, P53, P53 + S9 CALUX assays were performed to elucidate potential genotoxic and/or cytotoxic effects.

### 1.2.3 Ready biodegradability tests

The ready biodegradability of the compounds was investigated in a manometric respirometry test (OECD301F) over a period of 28 days (OECD, 1992). The biodegradation was followed by the oxygen uptake of the microorganisms during exposure. As a reference item sodium benzoate was tested simultaneously under the same conditions as the test item, and functioned as a procedure control. Aerobic activated sludge (microorganisms from a domestic wastewater treatment plant) was supplied by the sewage treatment plant of Rossdorf, Germany. Degradation rate of compounds was calculated by the oxygen consumption of the aerobic activated sludge microorganisms after 28 days of incubation. The testing was conducted at 22°C ± 1°C (darkness) according to GLP standards by Ibacon GmHb (Rossdorf, Germany).

### 1.2.4 TGA

To get a first indication about the functionality of the designed and synthesized alternative compound as a flame retardant, thermogravimetric analysis (TGA) was performed (Markwart et al., 2019). The thermogravimetric analysis of TiBP and di-*n*-butyl (2-hydroxyethyl) phosphate (target I) were performed with a Mettler-Toledo TGA/DSC 3+ instrument with autosampler. Ca. 10 mg sample was transferred into a 100 µL aluminum crucible using a micropipette. The lid was pierced with a needle and then used to seal the crucible. The temperature was increased from 25 °C to 600 °C at a rate of 10 K $min^{-1}$. Both compounds were analyzed in air and in a nitrogen atmosphere at a gas flow of 40 mL $min^{-1}$. Measurements were performed in duplicate.

# 2 Results

## 2.1 Top 25 of the PBMTS MAUT analysis

Desirability functions were designed to calculate subscores for LogKoc and LogKow, indicating optimal ranges for the values based on reported criteria, as both very high and very low values are not desirable. The desirability functions for the PBMTS MAUT analysis are shown in Figure S1, together with distributions (kernel density) of each endpoint used in the MAUT analysis, both for the whole dataset and for the top 500 structures. The scoring of the selected target structure is visualized by the blue line. It must be noted that the distribution of the anaerobic biodegradability sub score moved to lower values in the set of structures ranked under the top 500 structures compared to the total dataset. For aerobic biodegradability and non-mutagenicity sub scores, the kernel density estimate curve of the top 500 structures shifted to higher scores. Also, the distribution of the LogKow and LogKoc shifted to a more desirable range, as defined by the desirability functions, indicating the effectiveness of the PBMTS scoring approach.

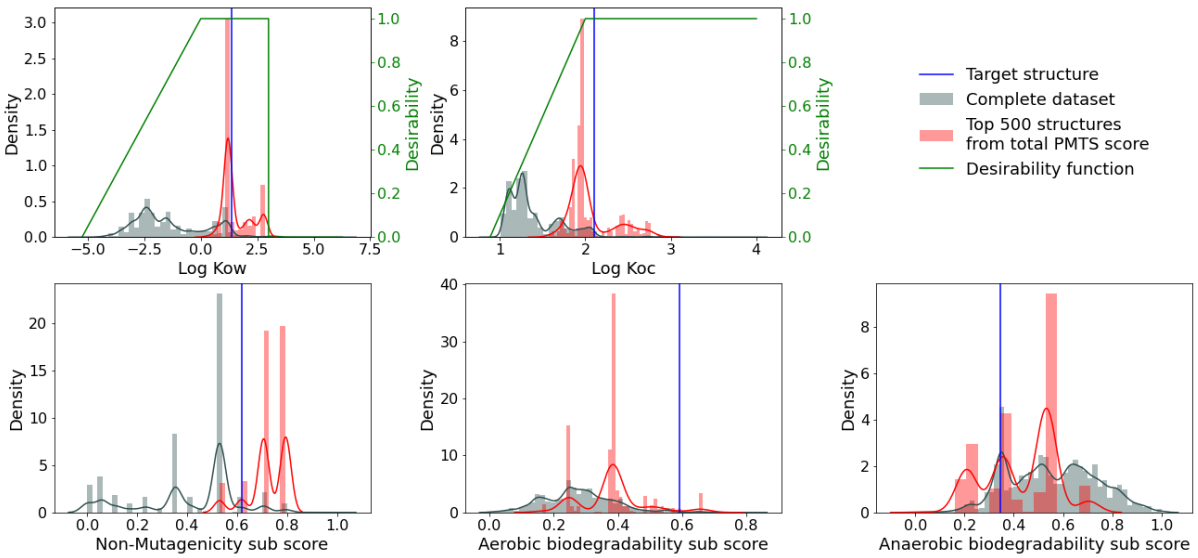



*Figure S1. Histograms and corresponding kernel density estimates showing the distribution of properties in the complete dataset and the top 500 structures ranked according to the PBMTS score. Desirability functions are plotted in green, values corresponding to the selected target structure are marked with blue vertical lines.*

Figure S2 shows the top 25 ranked structures from the PBMTS MAUT analysis and their scores are shown. The highest ranked structure has a score of 0.7103. The final selected target structure ranks 22$^{nd}$ and has a score of 0.6673.

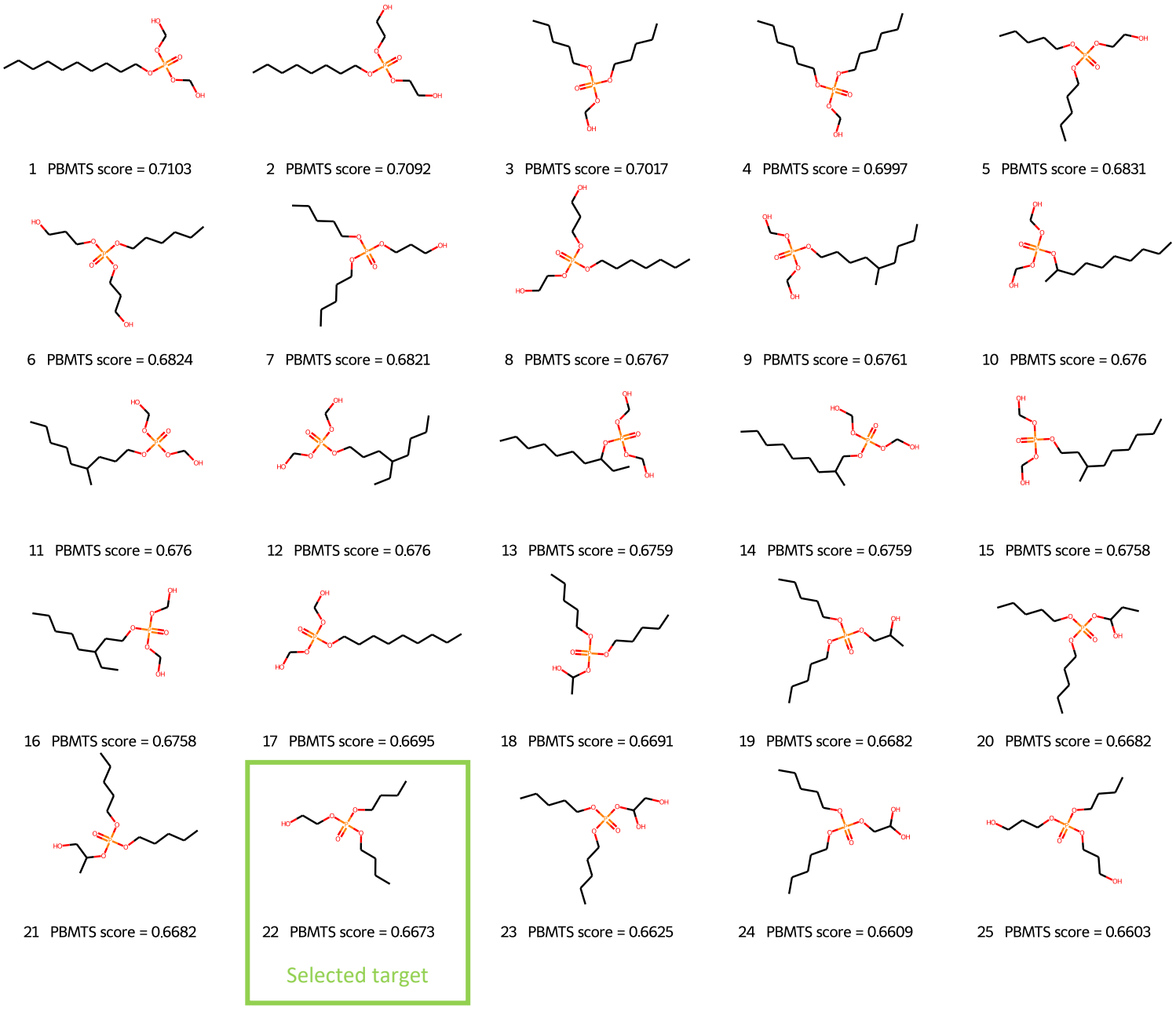



*Figure S2.Top 25 structures according to the PBMTS score.*

## 2.2 CALUX assays

As a first step, cytotoxic effects of TiBP and the target compound were tested to determine appropriate concentrations for the CALUX assays to ensure observed effects are not caused by the compound's cytotoxicity (Figure S3).

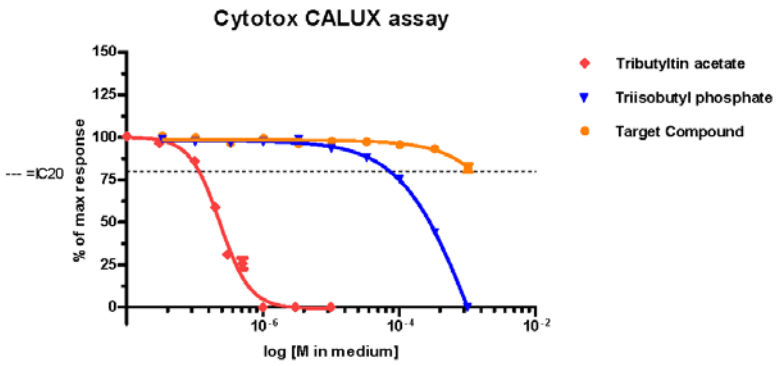
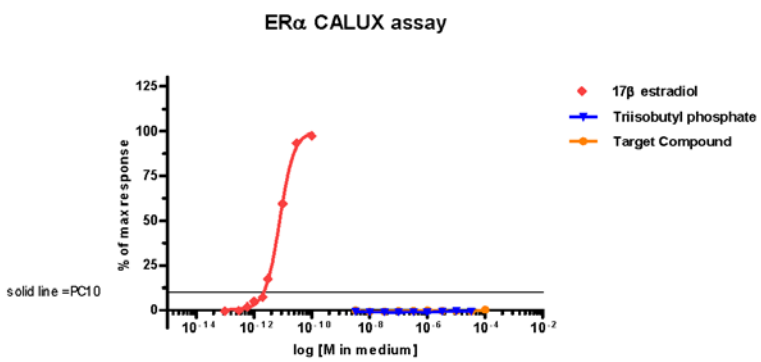
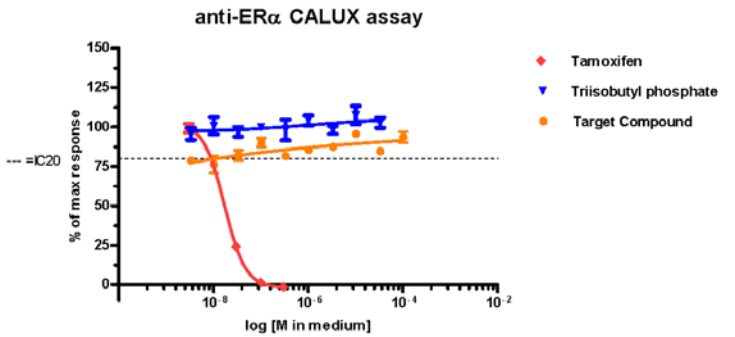
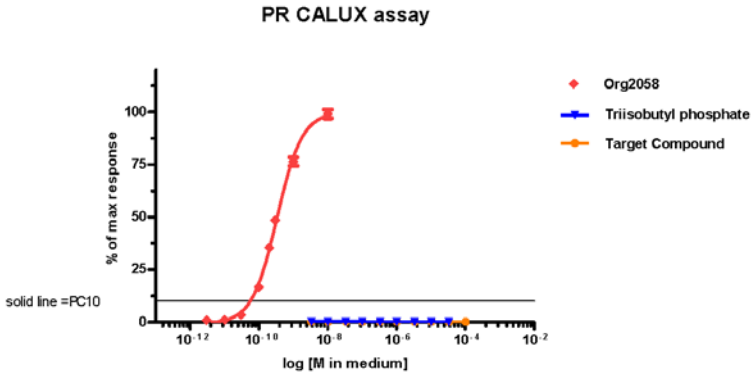
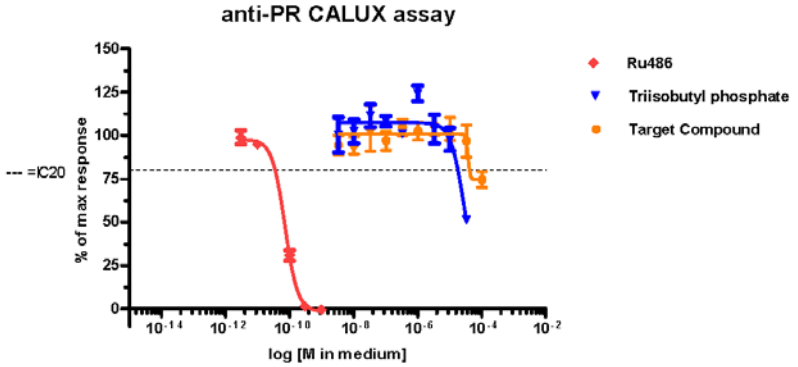



*Figure S3. Cytotoxic effects of TiBP, the Target Compound and the positive control (Tributyltine acetate).*

Table S2 shows the results of the CALUX assays used to identify hormone-like properties. For every test a positive control was used (shown in red) to verify the test. Both the TiBP and the target compound seem to have a slight antagonistic effect on the androgen and progesterone receptors in the highest concentrations tested (8.93 mg/L TiBP and 19.6 mg/L target compound). It will however be very unlikely that such high concentrations will be reached inside organisms in real-life, as the highest reported concentration of TiBP detected in organisms is $7.4*10^{-3}$ mg/kg ww (Brandsma et al., 2015). Moreover, both TiBP and the target compound were not found to be genotoxic in the CALUX assays (Table S3).

*Table S2. CALUX results for the identification of potential hormone-like properties*

| Receptor | Agonistic effects | Antagonistic effects |
|---|---|---|
| **Estrogen** |  |  |
| **Progesterone** |  |  |

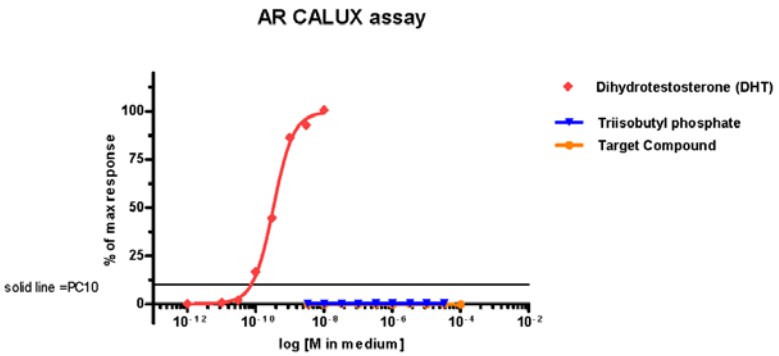
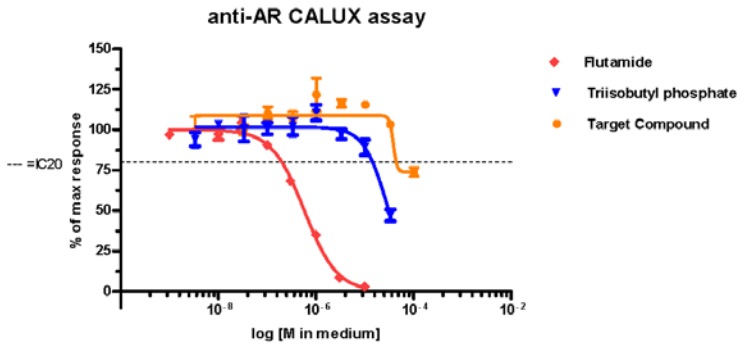
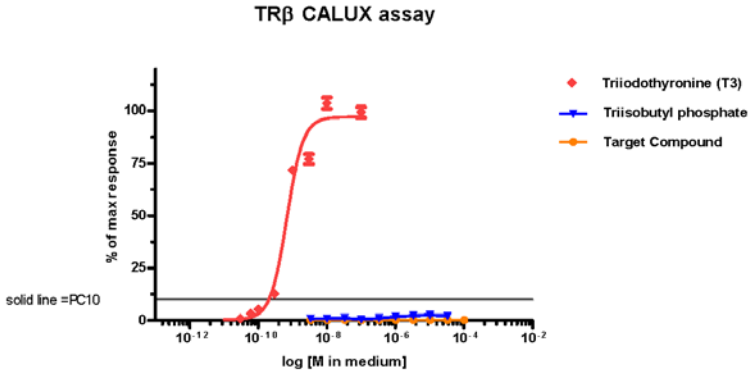
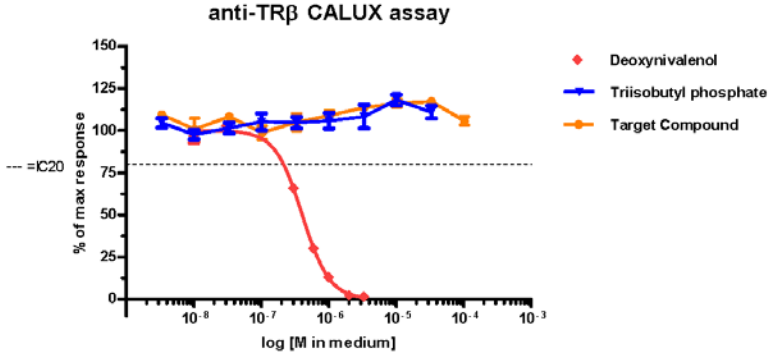
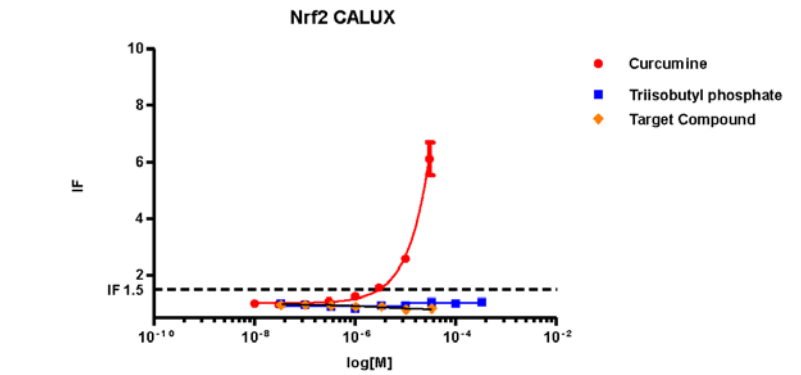
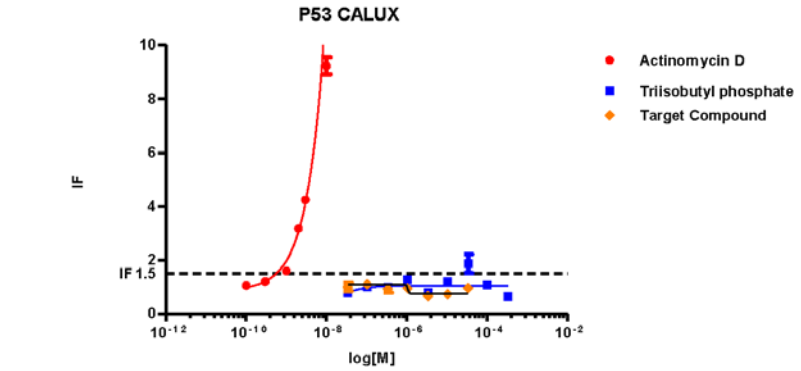
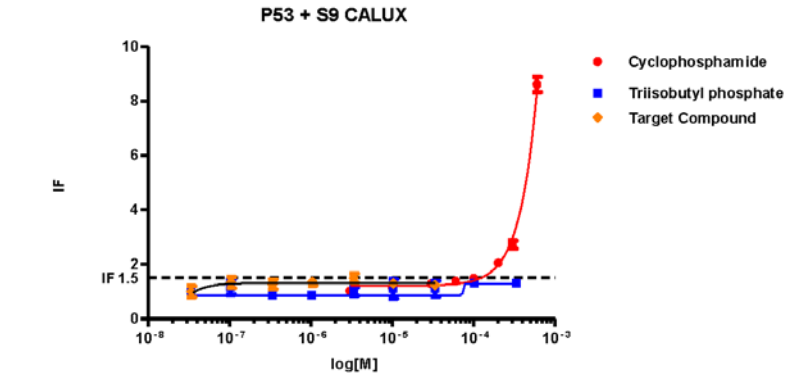
| | | |
|---|---|---|
| **Androgen** |  |  |
| **Thyroid** |  |  |

*Table S3. Results from the CALUX assays to identify potential genotoxic properties*

| Endpoint | Results |
|---|---|
| |  |
| |  |
| |  |

## 2.3 Ready biodegradability tests

Results from the OECD301F test are presented in Figure S4 (TiBP) and S5 (target compound). A mean biodegradation of 10% or more of TiBP was reached at day 15 (mean degradation of 14.5%) and for the target compound at day 18 (the mean degradation of 12%). At the end of the 10-day window at day 28, the mean degradation of TiBP was 41.5% and of the target compound was 27.5%. Therefore, for both substances the 10 day window criterion was not passed. The degradation rate of TiBP and the target compound also did not reach 60% within the 10-day window or after 28 days. Therefore, both TiBP and the target compound are considered to be not readily biodegradable according to the OECD301F test.

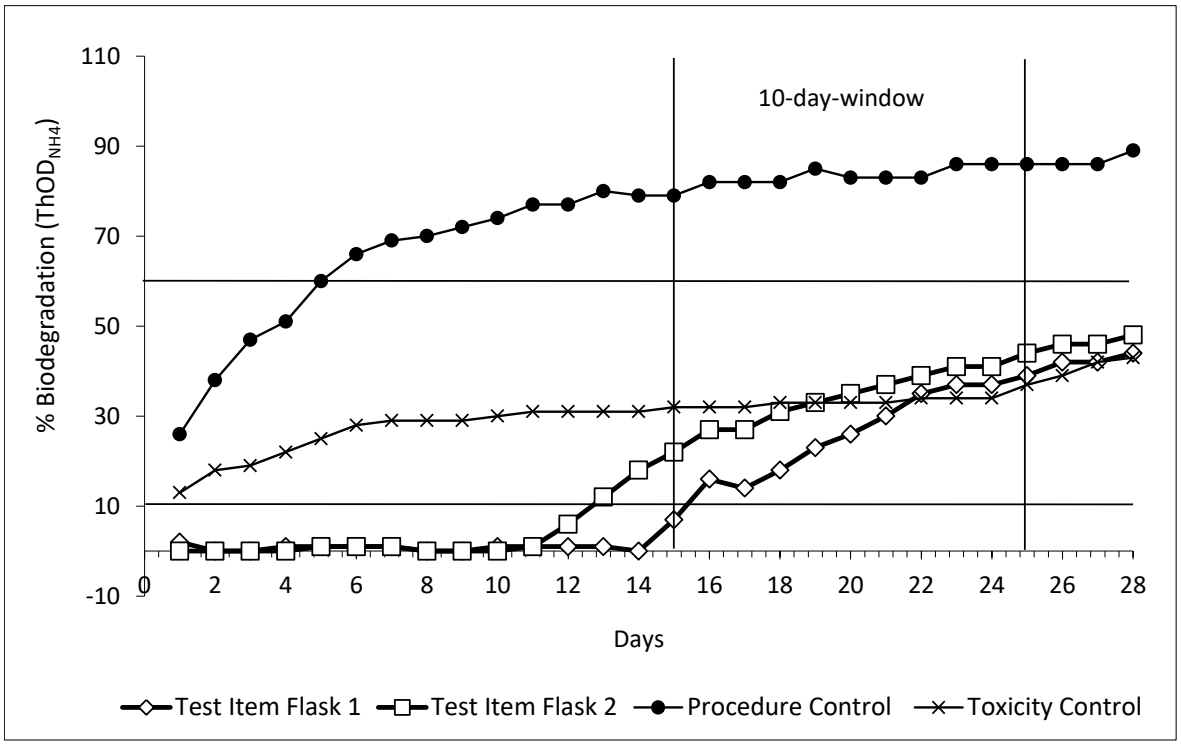



*Figure S4. Biodegradation during the Exposure Period of 28 days of TiBP (Flask 1 and 2), of the Toxicity Control and of the Reference Item Sodium Benzoate (Procedure Control) related to ThOD$_{NH4}$.*

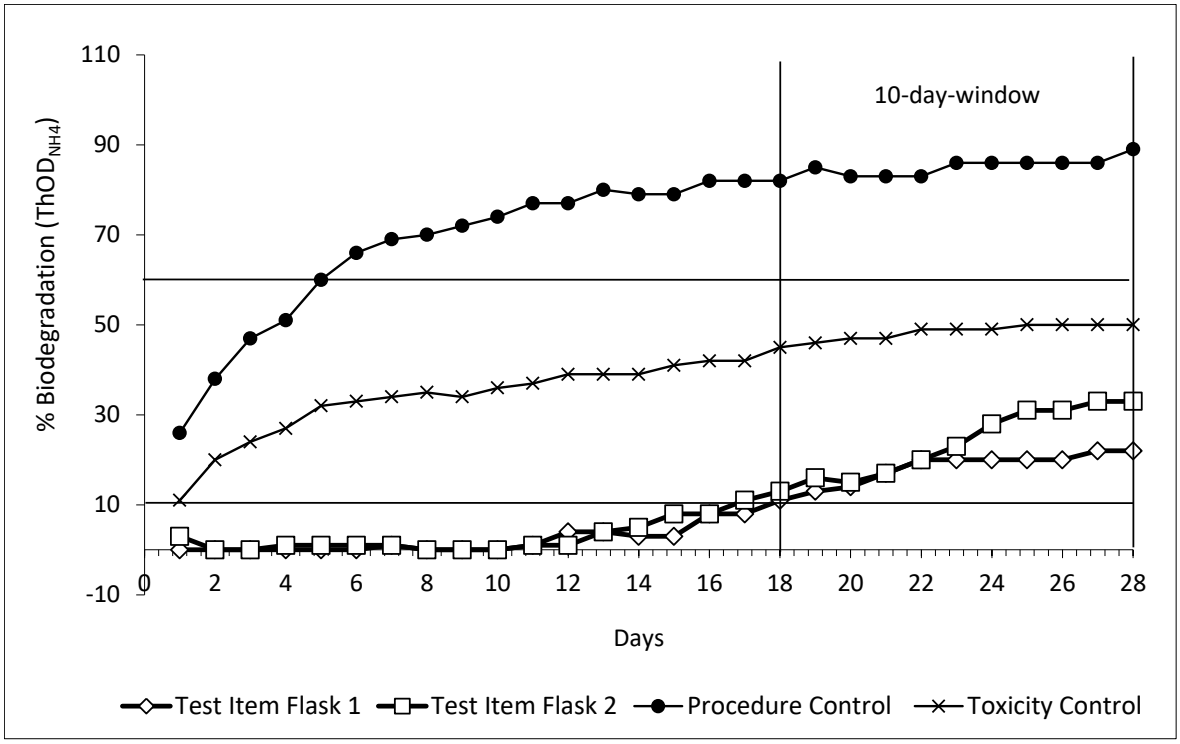



*Figure S5. Biodegradation during the Exposure Period of 28 days of the Target Compound (Flask 1 and 2), of the Toxicity Control and of the Reference Item Sodium Benzoate (Procedure Control) related to ThODNH4.*

## 2.4 TGA

The TGA monitors the weight loss as a function of temperature and the remaining char yield, which hints at condensed phase activity for flame retardancy. The char yield was 7% and 20% for TiBP and the target compound, respectively (Figure S2).

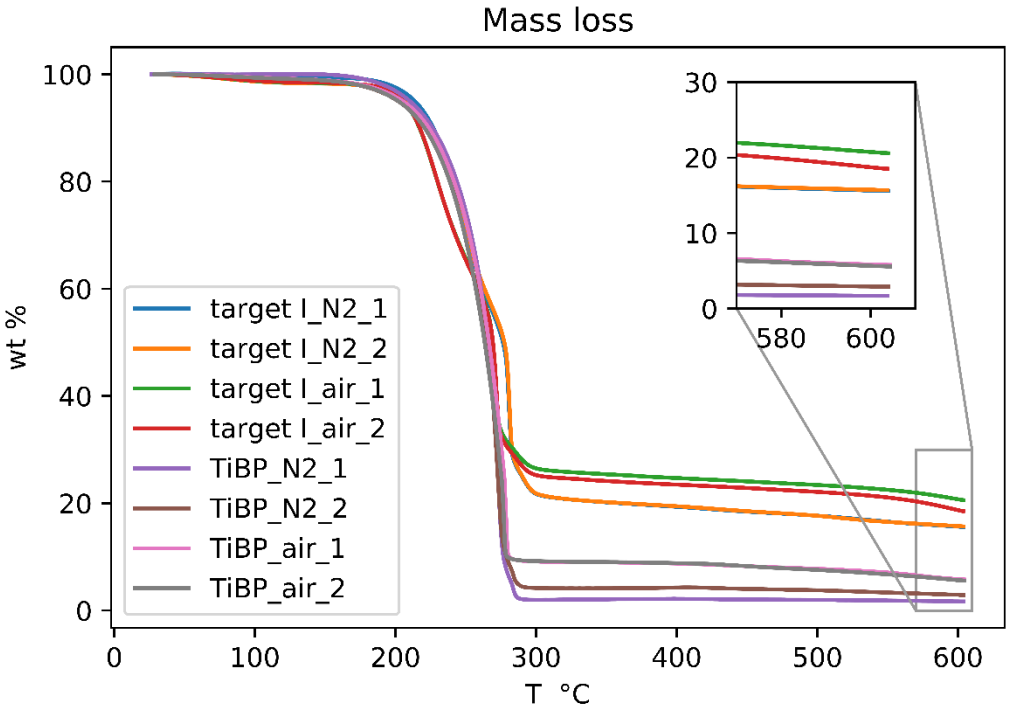



*Figure S6. Results of the thermogravimetric analysis of the selected target compound and the original compound TiBP show a higher char yield (residue as weight percent after heating, see inset) of the target compound, both in air and in nitrogen atmosphere.*

# 3 References


Besselink, H.T., Schipper, C., Klamer, H., Leonards, P., Verhaar, H., Felzel, E., Murk, A.J., Thain, J., Hosoe, K., Schoeters, G., Legler, J., Brouwer, B., 2004. Intra- and interlaboratory calibration of the DR CALUX® bioassay for the analysis of dioxins and dioxin-like chemicals in sediments. Environmental Toxicology and Chemistry 23, 2781–2789. https://doi.org/10.1897/03-542.1

Brandsma, S.H., Leonards, P.E.G., Leslie, H.A., de Boer, J., 2015. Tracing organophosphorus and brominated flame retardants and plasticizers in an estuarine food web. Science of The Total Environment 505, 22–31. https://doi.org/10.1016/j.scitotenv.2014.08.072

ECHA, 2016. Practical guide - How to use and report (Q)SARs. Version 3.1. (No. ECHA-16-B-09-EN).

Gramatica, P., Cassani, S., Roy, P.P., Kovarich, S., Yap, C.W., Papa, E., 2012. QSAR Modeling is not "Push a Button and Find a Correlation": A Case Study of Toxicity of (Benzo-)triazoles on Algae. Molecular Informatics 31, 817–835. https://doi.org/10.1002/minf.201200075

Markwart, J.C., Battig, A., Zimmermann, L., Wagner, M., Fischer, J., Schartel, B., Wurm, F.R., 2019. Systematically Controlled Decomposition Mechanism in Phosphorus Flame Retardants by Precise Molecular Architecture: P–O vs P–N. ACS Appl. Polym. Mater. 1, 1118–1128. https://doi.org/10.1021/acsapm.9b00129

Netzeva, T.I., Worth, A.P., Aldenberg, T., Benigni, R., Cronin, M.T.D., Gramatica, P., Jaworska, J.S., Kahn, S., Klopman, G., Marchant, C.A., Myatt, G., Nikolova-Jeliazkova, N., Patlewicz, G.Y., Perkins, R., Roberts, D.W., Schultz, T.W., Stanton, D.T., van de Sandt, J.J.M., Tong, W., Veith, G., Yang, C., 2005. Current Status of Methods for Defining the Applicability Domain of (Quantitative) Structure-Activity Relationships: The Report and Recommendations of ECVAM Workshop 521,2. Altern Lab Anim 33, 155–173. https://doi.org/10.1177/026119290503300209

OECD, 2007. OECD Environment Health and Safety Publications Series on Testing and Assessment No. 69. GUIDANCE DOCUMENT ON THE VALIDATION OF (QUANTITATIVE)STRUCTURE-ACTIVITY RELATIONSHIPS [(Q)SAR] MODELS (No. ENV/JM/MONO(2007)2). Paris, France.

OECD, 1992. Test No. 301: Ready Biodegradability (Text).

US EPA, 2012. Estimation Programs Interface Suite™ for Microsoft® Windows (EPISUITE), v 4.11. United States Environmental Protection Agency, Washington, DC, USA.