

CADASTER

Case studies on the Development and Application of in-Silico Techniques for Environmental hazard and Risk assessment

Grant agreement no.: 212668

Collaborative Project

Sub-Priority ENV2007 3.3.1.1: In-silico techniques for hazard-, safety-, and environmental risk assessment

| |
|-------------------------------------------------------------------------------------------------------------------|
| Overview of (Q)SAR models and their specific features for assessing fate and effects (Deliverable 2.2) |
|-------------------------------------------------------------------------------------------------------------------|

Due date of deliverable: December 31, 2009

Actual submission date: December 31, 2009

Start date of project: 1 January 2009

Duration: 4 years

Lead Contractor: National Institute of Public Health and the Environment (RIVM),
Laboratory for Ecological Risk Assessment

Corresponding author of document: Mojca Kos Durjava¹, Willie Peijnenburg², Paola Gramatica³

1. Public Health Institute Maribor, Center for Risk Assessment of Chemicals with laboratory, Maribor, Prvomajska 1, Slovenija (mojca.durjava@zzv-mb.si)

2. RIVM Laboratory for Ecological Risk Assessment - P.O. Box 1 3720 BA Bilthoven, The Netherlands (willie.peijnenburg@rivm.nl)

3. QSAR Research Unit in Environmental Chemistry and Ecotoxicology, University of Insubria, Via J.H. Dunant 3 - 21100 Varese, Italy (paola.gramatica@uninsubria.it)

Deliverable no: 2.2 (Overview of (Q)SAR models and their specific features for assessing fate and effects)

Nature: Report restricted to consortium partners and EU services

| Project co-funded by the EU Commission within the Seventh Framework Programme | | |
|-------------------------------------------------------------------------------|---------------------------------------------------------------------------------------|---|
| Dissemination Level | | |
| PU | Public | |
| RE | Restricted to a group specified by the consortium (including the Commission Services) | |
| CO | Confidential, only for members of the consortium (including the Commission Services) | X |

WP 2: Database on experimental parameters and (Q)SARs for chemical and biological endpoints

Work Package Leader: Mojca Kos Durjava (Partner 2: Public Health Institute Maribor)

Task 2.2- Collection of (Q)SAR models and non-testing approaches. Overview of (Q)SAR models and their specific features for assessing fate and effects (Deliverable 2.2)

Overview

A survey of the existing QSAR/QSPR models for the four classes of chemicals selected in this project (Brominated flame retardants, Fragrances, Perfluorinated chemicals, Triazoles and Benzotriazoles), stated in previous reports, has been completed. At the moment, just a few QSAR models specifically developed on the four chemical classes of compounds studied in CADASTER, have already been published. The analysis of these models according to the requirements of the 'OECD principles for QSAR validation' for regulatory applicability is the topic of Deliverable 3.2. (due per December 2009) Publicly available EPI Suite models were also taken into consideration. In this case, it has been assessed, which of these models are reliably applicable to the four classes of chemicals.

At this stage of knowledge, QSPR models are available only for some SIDS physico-chemical properties of BFRs (Henry's low constant, vapor pressure, water solubility, LogK_{OW}, photodegradation rate), while for the other three classes of chemicals EPI Suite models are the only tools available to predict SIDS physico-chemical properties.

The same applies for existing QSAR models, which are predominantly developed for non-SIDS endpoints, such as endocrine disruption (for BFRs and PFCs) or skin sensitization (for fragrances). There is only one QSAR model based on acute toxicity to fish which is developed for a large data set containing a few substituted triazoles. In the absence of *ad hoc* QSAR models for the four classes of chemicals selected in this project, the ECOSAR estimation program included in EPI Suite could be used to predict acute and chronic toxicity to fish, aquatic invertebrates and algae (SIDS endpoint for ecotoxicity). The problems linked to the applicability of ECOSAR models to BFRs, fragrances, PFCs, TAZs and BTAZs are better explained in the following paragraph.

Activities performed

Below are summarized the existing QSARs collected for individual classes. Tables containing all the useful information and details on the models are attached in Appendix 2.2. The same models have been verified for their application of the OECD Principles of QSAR in regulation (Deliverable 3.2).

BFRs

QSPRs models have been published for the following physico-chemical and degradation properties:

- Henry's Law Constant, H (SIDS) (Xu et al., 2007);
- Subcooled Liquid Vapor Pressure, PL (SIDS) (Xu et al., 2007; Wania and Dugani, 2003; Wang et al., 2008; Öberg, 2002);
- Water Solubility, WS (SIDS) (Wania and Dugani, 2003);
- Octanol-Air Partition Coefficient, LogK_{OA} (Xu et al., 2007; Wania and Dugani, 2003 ; Wang et al., 2008; Chen et al., 2003 ; Zhao et al., 2005) ;
- Octanol-Water Partition Coefficient, LogK_{OW} (SIDS) (Wania and Dugani, 2003; Braekevelt et al., 2003; Li et al., 2008);
- Photodegradation rate, K_p, measured in different solvents (methanol/water, methanol, hexane) (SIDS) (Niu et al., 2006; Chen et al., 2007 ; Fang et al., 2009);
- Quantum yield, Φ_p, measured in different solvents (methanol/water, methanol, hexane) (Niu et al., 2006; Chen et al., 2007 ; Fang et al., 2009).

To our knowledge, no *ad hoc* QSAR models have already been published for ecotoxicity endpoints (acute toxicity to Algae, Daphnia and Fish), probably due to the lack of sufficient amounts of experimental data. However different QSARs have been found in literature for endpoints related to endocrine disrupting activity of BFR, that is:

- Aryl hydrocarbon Receptor Relative Binding Affinity, RBA (Wang et al, 2005; Wang et al., 2006; Zheng et al., 2007);
- *Luciferase* activity (Harju et al., 2002);
- Androgen Receptor Antagonism, AR_{ANT} (Harju et al., 2007; Yang et al., 2009).

In addition to the above mentioned *ad hoc* QSARs/QSPRs, different EPI Suite estimation programs can be used to predict SIDS physico-chemical properties, such as Henry's Law Constant, Melting Point (MP), Boiling Point (BP), Vapor Pressure (VP), Water Solubility (WS), LogK_{OA} and LogK_{OW}.

For most of EPI Suite models no experimental data of brominated flame retardants, and in particular of PBDEs, are included in the training set used to build the models. Comparing available experimental physico-chemical data of BFRs with those estimated by the EPI Suite programs, the latter ones tend to either overestimate or underestimate the experimental values, especially for highly brominated compounds. This evidence suggests that BFRs are not included in the structural applicability domain of EPI Suite models. Reported below is a table containing comparisons between EPI Suite models and models recently published by the University of Insubria (UI) within the CADASTER Project (Papa *et al.*, QSAR & Comb.Sci, 2009), realized through the calculation of RMSE values (*Root Mean Square of Errors*).

Table 1. Comparison between EPI Suite models and *ad hoc* models by UI (Papa *et al.*, 2009).

| Phys-Chem Property | Model | N _{obj} ^a | RMSE |
|-------------------------------|-----------|-------------------------------|-------|
| LogH [Pa m ³ /mol] | EPI Suite | 7 | 0.79 |
| | BEHe7 | | 0.11 |
| MP [°C] | EPI Suite | 25 | 57.47 |
| | X2A | | 18.92 |
| LogVp [Pa] | EPI Suite | 34 | 0.93 |
| | T(O..Br) | | 0.16 |
| LogWS [mol/L] | EPI Suite | 12 | 1.88 |
| | Mor23m | | 0.22 |
| LogKoa | EPI Suite | 30 | 0.81 |
| | T(O..Br) | | 0.23 |
| LogKow | EPI Suite | 20 | 0.91 |
| | T(O..Br) | | 0.19 |

^a experimental data collected by UI

As it can be seen in Table 1, RMSE values of EPI Suite are always higher than those of the *ad hoc* models developed by UI.

Another important tool included in EPI Suite is the ECOSAR estimation program, which predicts toxicity to fish, aquatic invertebrates and algae using an extensive set of structure-activity relationship models (SARs). The majority of BFRs (PBDEs, CH₃O-BDE metabolites, hexabromocyclododecane, hexabromobenzene, 2,4,6-tribromophenol) are classified by ECOSAR as “Neutral Organics”. For these chemicals baseline toxicity is estimated by logKow.

Since acute toxicity models (e.g. Fish LC50 96-h, Daphnid LC50 48-h, Green Algae EC50 96-h ...) can be used only for compounds with LogKow values less than 5.0 (6.4 for Green Algae), these models are not applicable to BFRs, which are characterized by LogKow values higher than 5.

For chronic toxicity endpoints (i.e. Fish ChV 30-d, Daphnid ChV 16-d, Green Algae ChV) ECOSAR models can be used for compounds with LogKow values less than 8.0. Thus, these models could be applicable for most of BFR, with the exception of higher brominated PBDE.

In any case, ECOSAR highlights that, according to water solubilities estimated by the program WSKOWWIN, BFRs “may not be soluble enough to measure the predicted effects”.

Fragrances

Till now, to our knowledge, no *ad hoc* QSAR/QSPR models have been developed for the prediction of physico-chemical properties and environmental toxicity of fragrances. Nevertheless different QSARs exist for skin sensitization, an endpoint related to human toxicity but not included in SIDS.

Hence, the only existing models for the prediction of SIDS endpoints for fragrances are EPI Suite models. EPI Suite programs can be used to estimate physico-chemical properties such as Melting Point, Boiling Point, Vapor Pressure, Water Solubility, LogK_{OA} and LogK_{OW}.

In general, EPI Suite estimations are quite in agreement with experimental data, although they are less accurate than those obtained using *ad hoc* models recently developed by UI within the CADASTER

Project. In Table 2 a comparison between EPI Suite models and *ad hoc* models developed by UI is reported (Papa et al., SAR & QSAR Environ. Res., in press, 2009).

Table 2. Comparison between EPI Suite models and *ad hoc* models by UI.

| Phys-Chem Property | Model | N _{obj} ^a | RMSE |
|--------------------|-------------------|-------------------------------|------|
| LogVp [Pa] | EPI Suite | 37 | 1.91 |
| | piPC01 nHDon | | 0.2 |
| LogWS [mg/L] | EPI Suite | 37 | 0.3 |
| | BEHm3 JGI3 nCconj | | 0.45 |
| LogKow | EPI Suite | 52 | 0.64 |
| | X2v RDCHI | | 0.47 |

^a experimental data collected by UI

In relation to the ECOSAR models, they classify the fragrances based on the different fragments of the chemicals and, in many cases, give different toxicity results for a single compound, depending on the class to which the compound is assigned. As an example, for the fragrance Cinnamyl acetate (CAS: 103-54-8) ECOSAR identifies two classes, “Esters” and “Vinyl/Allyl Esters”; in addition, the program gives also baseline toxicity predictions based on the Neutral Organic SAR. The predicted toxicity values are highly different depending on the assigned class (Table 3).

Table 3. ECOSAR predictions (mg/L) for the fragrance 103-54-8.

| Endpoint | Ester SAR | Vinyl/Allyl Esters SAR | Neutral Organic SAR |
|---------------------------|------------|------------------------|---------------------|
| Fish LC50 - 96 hr | 7.845 | 0.932 | 26.007 |
| Fish (SW) LC50 – 96 hr | 10.791 | | |
| Fish LC50 - 14 day | 316.753 * | | |
| Daphnid LC50 – 48 hr | 14.454 | 5.896 | 16.627 |
| Green Algae EC50 – 96 hr | 5.623 | 1.475 | 10.152 |
| Mysid Shrimp LC50 – 96 hr | 7.241 | | |
| Earthworm LC50 – 14 day | 1224.109 * | | |
| Fish ChV | 0.575 | 0.016 ! | 2.423 |
| Fish (SW) ChV | 1.846 ! | | |
| Daphnid ChV | 7.331 | 0.175 ! | 1.784 |
| Green Algae ChV | 1.868 | 0.315 | 4.289 |
| Mysid Shrimp (SW) ChV | 92.035 | | |

* Chemical may not be soluble enough to measure this predicted effect.

! The toxicity value was determined from a predicted SAR using established acute-to-chronic ratios and ECOSAR regression techniques which are documented in the supporting Technical Reference Manual. When possible, this toxicity value should be considered in a weight of evidence approach.

In addition, most of the models used by the ECOSAR program to estimate ecotoxicity for fragrances are not validated QSARs, and based on very few experimental data.

Perfluorinated chemicals (PFCs):

For QSPRs on SIDS physico-chemical properties, data on boiling point, Fluorophilicity - Fluorous partition coefficient are modeled (Rucker et al., 2005; Kiss et al., 2001). In addition, commercial softwares were used to derive “polyparameter linear free energy relationships” for various end points.

EPI Suite models have been also considered. Their performances have been compared with those of some preliminary models developed by UI on MP, BP and VP (presented in Conferentia Chemometrica 2009, Siofok, Hungary).

Table 4. Comparison between EPI Suite models and *ad hoc* models by UI.

| Phys-Chem Property | Model | N _{obj} ^a | RMSE |
|--------------------|-----------------------|-------------------------------|-------|
| MP [°C] | EPI Suite | 111 | 46.68 |
| | AAC F02[C-F] C-013 | | 40.36 |
| BP [°C] | EPI Suite | 130 | 43.05 |
| | Ms ATS1m nROH | | 27.57 |
| LogVp [Pa] | EPI Suite | 35 | 1.12 |
| | CIC0 MATS1v TPSA(Tot) | | 0.83 |

^a experimental data collected by UI

QSAR models on T4-TTR binding are published using the PLS approach (Weiss et al., 2009). In addition, ECOSAR predictions can be considered as toxicity models for PFCs but as in above cases (eg. Fragrances) they are derived from a fragment based approach, with either few or with predicted data, and the baseline toxicity value which is used belongs to different chemical classes and not specifically for the PFCs. In some cases, the differences between the predictions from different chemical classes (eg. CAS 360-58-7) are 10 fold (Table 5).

Table 5. Examples of ECOSAR predictions (mg/L) for some PFCs compounds.

| Compound | Fragments | Fish LC50 96-hr | Daphnid LC50 48-hr | Green Algae EC50 96-hr |
|----------|-------------------------|-----------------|--------------------|------------------------|
| 376-89-6 | Halo Nitriles | No current SARS | | |
| | Nitriles, Polyaliphatic | 628.510 | 332.588 | 106.871 |
| 360-58-7 | Vinyl/Allyl Ethers | 0.270 | 9.711 | 12.056 |
| | Halo Ethers | 20.597 | 5.583 | No data |

Triazoles and Benzotriazoles (TAZs and BTAZs)

QSPR models specifically on TAZs and BTAZs have been not found in the literature. Only logP data are modeled where few TAZs are part of the larger dataset. Regarding EPI Suite models, their predictions for MP, VP, LogK_{ow} and WS do not show large deviations from available experimental data. However, preliminary *ad hoc* QSPRs developed by UI for triazoles and benzotriazoles have RMSE values always

lower than those calculated for EPI Suite models, the main exception being the LogKow model (Table 6).

Table 6. Comparison between EPI Suite models and *ad hoc* preliminary models developed by UI .

| Phys-Chem Property | Model | N _{obj} ^a | RMSE |
|--------------------|-----------------------|-------------------------------|--------|
| LogVp [mmHg] | EPI Suite | 31 | 1.725 |
| | RBN MATS4e EEig15x | 31 | 0.785 |
| LogWS [mg/L] | EPI Suite | 49 | 1.247 |
| | AMW CICO MATS7e | 49 | 0.537 |
| MP [°C] | EPI Suite | 56 | 71.728 |
| | X1A GGI4 R2e F03[N-N] | 56 | 26.634 |
| LogKow | EPI Suite | 62 | 0.604 |
| | T(0..0) GATS1m W3D | 62 | 0.685 |

^a experimental data collected by UI

QSAR models on the following end-points are found for TAZs/BTAZs (Trohalaki et al., 2002; Klink, 2003; Wei et al., 2006):

1. EC₂₅MTT (mM)
2. LSCROS (mM)
3. EC₂₅LDH (mM)
4. EC₅₀GSH (mM)
5. Sensitization and/or irritancy potential
6. Fungicide (wheat head blight) (FA) (50 ug/ml)

Although these models are not specifically for SIDS endpoints, they are related to human and to plant toxicity. In addition, models based on “LC₅₀-96h acute toxicity of rainbow trout” are developed for large datasets which contain only few TAZs (Benfenati, 2006; Benfenati, 2008).

ECOSAR models on TAZs and BTAZs have the same problems as described above. In Table 7 ECOSAR predictions for the compounds 8-Azaxanthine (CAS 1468-26-4) and Acetamide, N-(3-(5-nitro-2-furyl)-s-triazol-5-yl) (CAS 1704-66-1) are reported as an example.

Table7. ECOSAR predictions (mg/L) for the compounds 1468-26-4 and 1704-66-1.

| Compound | Classes/Fragments | Fish LC50 96-hr | Daphnid LC50 48-hr | Green Algae EC50 96-hr |
|-----------|-----------------------|-----------------|--------------------|------------------------|
| 1468-26-4 | Imides | 14909.459 | 12330.157 | 35.158 |
| | Amides | 6905.285 | 1280.314 | 2.649 |
| | Triazoles (Non-Fused) | 3062.8 | 14933.703 | 125.656 |
| 1704-66-1 | Amides | 463.87 | 155.959 | 1.359 |
| | Triazoles (Non-Fused) | 407.283 | 1329.625 | 34.103 |

In general, most of ECOSAR models available are based on very few data and thus they are of limited utility for the specific classes of compounds studied under the project CADASTER. From this analysis

the need of the development of specific QSAR models in CADASTER Project (WP3) is highly evident. The main problem in this respect is the lack of SIDS data in sufficient amounts to develop new QSARs. For this reason, all the available data that have been collected will be modeled.

Activities foreseen

In line with the Description of Work of CADASTER, the collection and evaluation of available QSAR models will continue during the whole Project.

Appendix to Deliverable 2.2

The Appendix to Deliverable 2.2 contains tables containing all the useful information and details on the QSAR models highlighted above.

Appendix to Deliverable 2.2

Existing QSPR models on Brominated Flame Retardants (BFRs)

| Reference | Method / Tool | Endpoint | Train obj. | Test obj. | n°Var | n° PLS comp | R ² % | Q ² _{cum} /R ² _{CV} % | Experimental Data set | |
|---------------------------------|---------------|---------------------------------------------------------------|-----------------------------------------------------------|-----------|-------|-------------|------------------|---------------------------------------------------------------|---------------------------------------------------------------|---------------|
| Xu <i>et al.</i> (2007) | MLR | Log H | 7 | - | 2 | | 99.74 | 99.79 | Cetin and Odabasi (2005) | |
| | | Log P _L | 22 | - | 2 | | 98.13 | 97.59 | Wong <i>et al.</i> (2001) | |
| | | Log K _{OA} | 22 | - | 2 | | 97.61 | 97.25 | Wania <i>et al.</i> (2002) | |
| | | -Log RBA | 18 | - | 4 | | 64.73 | 28.94 | Chen <i>et al.</i> (2001) | |
| Wania and Dungani (2003) | LR | Log P _L | 6 | - | 1 | | 99.62 | - | Wania and Dugani (2003) | |
| | | Log S | 6 | - | 1 | | 99.19 | - | | |
| | | Log K _{OW} | 6 | - | 1 | | 97.54 | - | | |
| | | Log K _{OA} | 6 | - | 1 | | 99.41 | - | | |
| Braekevelt <i>et al.</i> (2003) | LR | Log K _{OW} | 9 | - | 1 | | 97 | - | Braekevelt <i>et al.</i> (2003) | |
| Wang <i>et al.</i> (2008) | MLR | Log P _L | 23 | 7 | 3 | | 99.71 | 99.58 | Wong <i>et al.</i> (2001), Tittlemier <i>et al.</i> (2002) | |
| | | Log K _{OA} | 22 | 6 | 3 | | 99.73 | 99.64 | Wania <i>et al.</i> (2002), Harner and Shoeib (2002) | |
| Öberg (2002) | PLS | Log P _L | 23 | 9 | ? | 1 | 99.2 | | Wong <i>et al.</i> (2001), Tittlemier <i>et al.</i> (2002) | |
| Chen <i>et al.</i> (2003) | PLS | Log K _{OA} | 9 (?) | - | 10 | | 97.9 | 97.5 | Harner and Shoeib (2002) | |
| Zhao <i>et al.</i> (2005) | MLR | Log K _{OA} | 13 | - | 2 | | 92.74 | - | Harner and Shoeib (2002) | |
| Li <i>et al.</i> (2008) | PLS | Log K _{OW} | 9 | - | 3 | | 98.93 | 96.1 | Braekevelt <i>et al.</i> (2003) | |
| US-EPA | EPI SUITE | H, MP, VP, S, Log K _{OW} , Log K _{OA} | RMSE of UI models are lower than that of EPI suite models | | | | | | | SRC- PhysProp |
| Niu <i>et al.</i> (2006) | PLS | Log K _p MET/H ₂ O | 15 | - | 6 | 1 | 95.8 | 95.7 | Eriksson <i>et al.</i> (2004) | |
| | | Log K _p MET | 9 | - | 6 | 3 | 97.81 | 98.4 | Eriksson <i>et al.</i> (2004) | |
| | | Log Φ _p MET/H ₂ O | 11 | - | 8 | 3 | 98.2 | 91.4 | Eriksson <i>et al.</i> (2004) | |

| Reference | Method / Tool | Endpoint | Train obj. | Test obj. | n°Var | n° PLS comp | R ² % | Q ² _{cum} /R ² _{CV} % | Experimental Data set |
|---------------------------|---------------|-----------------------------------------|------------|-----------|-------|-------------|------------------|---------------------------------------------------------------|-------------------------------|
| Chen <i>et al.</i> (2007) | PLS | Log K _p MET/H ₂ O | 15 | - | 8 | 3 | 98.2 | 97.3 | Eriksson <i>et al.</i> (2004) |
| | | Log K _p MET | 9 | - | 9 | 2 | 95.8 | 95.8 | Eriksson <i>et al.</i> (2004) |
| | | Log Φ _p MET/H ₂ O | 11 | - | 4 | 2 | 86.1 | 70.2 | Eriksson <i>et al.</i> (2004) |
| Fang <i>et al.</i> (2009) | PLS | Log K _p HEX | 18 | - | 7 | 1 | 91.97 | 90.7 | Fang <i>et al.</i> (2009) |
| | | Log K _p MET | 18 | - | 9 | 1 | 89.87 | 88.3 | Fang <i>et al.</i> (2009) |
| | | Log Φ _p HEX | 18 | - | 5 | 1 | 78.85 | 75.4 | Fang <i>et al.</i> (2009) |
| | | Log Φ _p MET | 18 | - | 5 | 2 | 86.49 | 93.0 | Fang <i>et al.</i> (2009) |

Existing QSAR models on Brominated Flame Retardants (BFRs)

| Reference | Method / Tool | Endpoint | Train obj. | Test obj. | n°Var | n° PLS comp | R ² % | Q ² _{cum} /R ² _{CV} % | Experimental Data set |
|----------------------------|-------------------|-------------------------------------------|--------------------------------------------------------------------------------------------------------------------|-----------|-------|-------------|------------------|---------------------------------------------------------------|-----------------------------|
| Wang <i>et al.</i> (2005) | CoMFA | -Log RBA | 18 | - | | 6 | 99.5 | 58 | Chen <i>et al.</i> (2001) |
| | CoMSIA | -Log RBA | 18 | - | | 6 | 98.2 | 68 | Chen <i>et al.</i> (2001) |
| Wang <i>et al.</i> (2006) | Heuristic Met. | -Log RBA | 18 | - | 4 | | 90.3 | 84.4 | Chen <i>et al.</i> (2001) |
| Zheng <i>et al.</i> (2007) | SVM _{3f} | -Log RBA | 15 | 3 | 7 | | - | 88.9 | Chen <i>et al.</i> (2001) |
| Harju <i>et al.</i> (2002) | PLS | Luciferase activity | 17 | - | | 1 | 61.8 | 48.6 | Meerts <i>et al.</i> (1998) |
| Harju <i>et al.</i> (2007) | PLS | -Log IC ₅₀ ARant | 20 | - | ? | | 90 | 77 | Hamers <i>et al.</i> (2006) |
| Yang <i>et al.</i> (2009) | CoMSIA | -Log IC ₅₀ ARant | 15 | 4 | | 3 | 97.6 | 54.6 | Hamers <i>et al.</i> (2006) |
| US-EPA | ECOSAR | ecotoxicity endpoints (acute and chronic) | ECOSAR models are not applicable to BFR both for their high LogKow values and for their very low water solubility. | | | | | | |

Existing QSPR models on Perfluorinated chemicals (PFCs)

| Reference | Method / Tool | Endpoint | Train obj. | Test obj. | n°Var | R ² % | Q ² _{cum} / R ² _{cv} % | Experimental Data set |
|-----------------------------|-------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------|------------------------------|-------|------------------|-------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------|
| Rucker <i>et al.</i> (2005) | MLR / MOLGEN -QSPR | BP | 82 | - | 7 | 0.99 | 0.98 | |
| Kiss <i>et al.</i> (2001) | NN / 3D-NET | Fluorophilicity - Fluorous partition coef. | 59 | - | 8 | | 0.97 | Kiss (1998) Szlavik (1999) |
| Arp <i>et al.</i> (2006) | ClogP Sparc EPI Suite COSMOtherm | Log P _L ^a Log K _{AW} ^a Log K _{OW} Log K _i ^a (hexanadecane/air) Log K _{OA} ^a | 11 4 4 8 7 | 5 12 12 12 9 | | | | Prediction from commercial softwares ^a ClogP calculation was not available for these endpoints |
| Goss (2008) | Sparc COSMO-RS | pKa | 33 | - | | | | Prediction from commercial softwares Henne and Fox (1951) |
| US-EPA | EPI SUITE | MP, BP, VP | RMSE of UI models are lower than that of EPI suite models | | | | | SRC- PhysProp |

Reported as Polyparameter linear free energy relationships (pp-LFER) in article

Existing QSAR models on Perfluorinated chemicals (PFCs) - continued

| Reference | Method / Tool | Endpoint | Train obj. | Test obj. | n°Var | n° PLS comp | R ² % | Q ² _{cum} /R ² _{cv} % | Experimental Data set |
|----------------------------|---------------|-------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------|-----------|-------|-------------|------------------|---------------------------------------------------------------|----------------------------|
| Weiss <i>et al.</i> (2009) | PLS | T4-TTR binding | 23 | - | 56 | 2 | - | 0.41 | Weiss <i>et al.</i> (2009) |
| US-EPA | ECOSAR | ecotoxicity endpoints (acute and chronic) | Prediction of the toxicity based on the baseline toxicity value of different fragments of chemicals but not specifically for PFCs | | | | | | |

Existing QSPR models on Triazoles and Benzotriazoles (TAZs and BTAZs)

| References | Method /Tool | Endpoint | n | Model parameters | Experimental Data set |
|------------------|----------------|------------------|-----|------------------------------------------------------------------------------------------|-----------------------|
| Devillers (1999) | Neural Network | Log P | 593 | Few triazoles commonly used as Pesticides are studied in a bigger training set | |
| US- EPA | EPI Suite | MP, VP, WS, LogP | | RMSE of UI models (according to OECD principles) are lower than that of EPI suite models | SRC- PhysProp |

Existing QSAR models on Triazoles and Benzotriazoles (TAZs and BTAZs)

| References | Method / Tool | Endpoint | n | n of TAZ | Test obj. | n°Var | A | R ² % | R ² _{CV} % /Q ² _{LOO} |
|-----------------------|--------------------------------------------------------------|--------------------------------------------------------|---------------|----------|-------------------------------------------------------------------------|-------|----------------|------------------|---------------------------------------------------------------|
| Benfenati (2006) | fragment-based QSAR | LC ₅₀ -96 h acute toxicity of rainbow trout | 282 | | | | | | |
| Benfenati (2008) | | LC ₅₀ -96 h acute toxicity of rainbow trout | 125 | | Few triazoles commonly used as Pesticides are studied in the bigger set | | | | |
| Benfenati (2006) | | Daphnia toxicity | 220 (42 test) | | | | | | |
| Trohalaki (2002) | Heuristic and Best MLR / CODESSA Quantum-Chemical descriptor | EC ₂₅ MTT (mM) | | | | | | | |
| | | LSCROS (mM) | 13 | 2 | - | | Not applicable | | |
| | | EC ₂₅ LDH (mM) | | | | | | | |
| | | EC ₅₀ GSH (mM) | | | | | | | |
| Klink (2003) | TOPKAT 6.0 | Sensitization and/or irritancy potential | 2 | 2 | - | | | | use of commercial software to predict the response |
| Wei Q-L et al. (2006) | MLR/SPSS | Fungicide (wheat head blight) (FA) (50 ug/ml) | 18 | 18 | - | 2 | - | 96.9 | - |

References

- Arp,H.P.H. et al. (2006) *Environmental Science and Technology* 40, 7298.
- Benfenati E. (2006) *Chemical Research in Toxicology* 19, 1533-1539.
- Benfenati E. (2006) *Bioorganic and Medicinal Chemistry* 14, 2779-2788.
- Benfenati E. (2008) *Journal of Environmental Science and Health, Part B* 43, 633-637.
- Breakevelt,E. et al. (2003) *Chemosphere* 51, 563-567.
- Cetin,B. and Odabasi,M. (2005) *Atmospheric Environment* 39, 5273-5280.
- Chen,G. (2001) *Environmental Science and Technology* 35, 3749-3756.
- Chen,J. et al. (2007) *Ecotoxicology and Environmental Safety* 66, 348-352.
- Chen,J.W. et al. (2003) *Chemosphere* 51, 577-584.
- Devillers J. (1999) *SAR & QSAR in Environmental Research* 10, 249-262.
- Eriksson,J. (2004) *Environmental Science and Technology* 38, 3119-3125.
- Fang,L. et al. (2009) *Ecotoxicology and Environmental Safety* 72, 1587-1593.
- Goss, K.U. (2006) *Environmental Science and Technology* 40, 3572-3577.
- Goss, K.U. (2008) *Environmental Science and Technology* 42, 456.
- Hamers,T. (2006) *Toxicological Science* 92, 157-173.
- Harju,M. et al. (2002) *Chemosphere* 47, 375-384.
- Harju,M. et al. (2007) *Environmental Toxicology and Chemistry* 26, 816-826.
- Harner,T. and Shoeib,M. (2002) *Journal of Chemical and Engineering Data* 47, 228-232.
- Henne, A.L. and Fox, C.J. (1951) *Journal of the American Chemical Society* 73, 2323–2325.
- Kaiser, M.A. et al. (2005) *Journal of Chemical and Engineering Data* 50, 1841-1843
- Kiss, L.E. et al. (2001) *Journal of Fluorine Chemistry* 108, 95.
- Kiss, L.E. et al. (1998) *Synlett* 1243
- Klink, K.J. (2003) *Toxicological Science* 75, 89-98.
- Krusic, P.J. et al. (2005) *Journal of Physical Chemistry A* 109, 6232-6241.
- Li,L. et al. (2008) *Chemosphere* 72, 1602-1606.

Meerts, I.A.T.M. (1998) *Organohalogen Compounds* 37, 147-150.

Niu, J. et al. (2006) *Chemosphere* 64, 658-665.

Öberg, T.G. (2002) *Environmental Science and Pollution Research* 6, 405-411.

Rucker, C. et al. (2005) *Journal of Chemical Information and Modeling* 45, 74.

Shoeib, M. et al. (2004) *Environmental Science and Technology* 38, 1313-1320.

Szlavik, Z. et al. (1999) *Journal of Fluorine Chemistry* 98, 83.

Tittkemier, S.A. et al. (2002) *Environmental Toxicology and Chemistry* 21, 1804-1810.

Trohalaki, S. et al. (2002) *Toxicological Science* 68, 498-507.

US-EPA EPI Suite, available on-line at: <http://www.epa.gov/oppt/exposure/pubs/episuite.htm>

Wang, Y. et al. (2005) *Environmental Science and Technology* 39, 4961-4966.

Wang, Y. et al. (2006) *Chemosphere* 64, 515-524.

Wang, Z.Y. et al. (2008) *Science of the Total Environment* 389, 296-305.

Wania, F. and Dugani, C.B. (2003) *Environmental Toxicology and Chemistry* 22, 1252-1261.

Wania, F. et al. (2002) *Analytical Chemistry* 74, 3476-3483.

Wei, Q-L et al. (2006) *Bioorganic and Medicinal Chemistry* 14, 7146.

Weiss et al. (2009) *Toxicological Science* 109, 206-216.

Wong, A. et al. (2001) *Journal of Chemical and Engineering Data* 46, 239-242.

Xu, H.-Y. et al. (2007) *Chemosphere* 66, 1998-2010.

Yang, W. et al. (2009) *Chemosphere* 75, 1159-1164.

Zhao, H.X. et al. (2005) *Chemosphere* 59, 1421-1426.

Zheng, G. et al. (2007) *QSAR & Combinatorial Science* 26, 536-541.