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

Evaluation of existing QSARs according to OECD principles
(Deliverable 3.2)

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: Paola Gramatica[1], Willie Peijnenburg[2]

1. QSAR Research Unit in Environmental Chemistry and Ecotoxicology, University of Insubria, Via J.H. Dunant 3 - 21100 Varese, Italy (paola.gramatica@uninsubria.i)
2. RIVM Laboratory for Ecological Risk Assessment - P.O. Box 1 3720 BA Bilthoven, The Netherlands (willie.peijnenburg@rivm.nl)

Deliverable no: 3.2 (Evaluation of existing QSARs according to OECD principles)
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 3: Development and validation of QSARs

Work Package Leader: Paola Gramatica (Partner 3: University of Insubria)

Task 3.2 Evaluation of existing QSARs.
Evaluation of existing QSARs according to OECD principles (report)
(Deliverable 3.2)

Overview

The existing QSARs (collected in WP2, Deliverable 2.2) have been evaluated according to the OECD principles for QSAR models validation for regulatory application:

A QSAR model for regulatory purposes should be associated with the following information:

1) a defined endpoint;

2) an unambiguous algorithm;

3) a defined domain of applicability;

4) appropriate measures of goodness-of–fit, robustness and predictivity;

5) a mechanistic interpretation, if possible.

Some models for BFRs and PFCs are listed here as examples, while a complete analysis of the existing QSARs, also for TAZ/BTAZs, is reported in the Tables of Appendix 3.2.

Analysis of some QSAR Models for BFRs according to the OECD Principles

Reference	Endpoint	Method	Chemicals within Applicability Domain	Validation status	Mechanistic interpretation
Xu et al. (2007)	LgH, LgP_L, LgK_{OA}, -LgRBA	MLR equations	no	No external validation	yes
Wania and Dungani (2003)	LgP_L, LgS, LgK_{OW}, $LogK_{OA}$	LR equations	no	No internal and external validation	yes
Braekevelt et al. (2003)	LgK_{OW}	LR equations	no	No internal and external validation	yes
Wang et al. (2008)	LgP_L, LgK_{OA}	MLR equations	?	yes	yes
Öberg (2002)	LgP_L	PLS	yes	yes	yes
Chen et al. (2003)	LgK_{OA}	PLS	no	No external validation	yes
Zhao et al. (2005)	LgK_{OA}	MLR equation	no	No internal and external validation	yes
Li et al. (2008)	LgK_{OW}	PLS equation	no	No external validation	yes
Niu et al. (2006)	$LogK_p$, $Log\Phi_p$	PLS equations	no	No external validation	yes
Chen et al. (2007)	$LogK_p$, $Log\Phi_p$	PLS equations	no	No external validation	yes
Wang et al. (2005)	-LgRBA	PLS	no	No external validation	?
Wang et al. (2006)	-LgRBA	Heuristic	no	No external validation	yes

		Method			
Zheng et al. (2007)	-LgRBA	Support Vector Machine	no	yes	yes
Harju et al. (2007)	-$LgIC_{50}$ARant	PLS	no	No external validation	yes

Analysis of some QSAR Models for PFCs according to the OECD Principles

Reference	Endpoint	Method	Chemicals within Applicability Domain	Validation status	Mechanistic interpretation
Rucker, C. 2005	BP	MLR	no	r^2 and r^2_{cv} (leave-one-out cross-validation) and checked on the external test set	no
Kiss, L.E. 2001	logP fluorous partition coefficient	NN	no	no	yes
Arp, H.P.H. 2006	different as reported above	different algorithm based on software used	no	no external validation on studied chemicals	no
Goss K-U. 2008	pKa		no		no

The SAR models on T4-TTR binding are published using the PLS approach. However, the applicability domain was not detailed and Mechanistic interpretations of the selected descriptors are absent.

The collection and evaluation of available QSAR models will continue for the whole duration of the CADASTER Project.

Main findings

In general, the majority of the existing QSARs were not externally validated, and a definition of the applicability domain is in general lacking. Thus, they do not fulfil the 'OECD principles for QSAR validation' for regulatory applicability. Hence, these models are of limited utility for the specific classes of compounds studied under the project CADASTER.

From this analysis, the need for the development of specific QSAR models in the CADASTER Project (WP3), developed according to the OECD principles, is highly evident. Currently, the major problem is the lack of SIDS data in sufficient amounts to develop new QSAR models. For this reason, all the available experimental data pertaining to toxicity and physic-chemical properties have been collected (Deliv. 2.1) and will be modeled in WP3.

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

Appendix to Deliverable 3.2

Existing QSPR models on Brominated Flame Retardants (BFRs)

Reference	Method	Endpoint	Train obj.	Test obj.	n°Var	n° PLS comp	R^2%	Q^2_{cum}/ R^2_{CV} %	$R(Q)^2_{ext}$ %	AD	(Mechanistic) Interpretation of desc.	Experimental Data set
Xu *et al.* (2007)	MLR	LgH	7	-	2		99.74	99.79	-	no	yes	Cetin and Odabasi (2005)
		LgP_L	22	-	2		98.13	97.59	-	no	yes	Wong et al. (2001)
		LgK_{OA}	22	-	2		97.61	97.25	-	no	yes	Wania et al. (2002)
		-LgRBA	18		4		64.73	28.94	-	no	yes	Chen et al. (2001)
Wania and Dungani (2003)	LR	LgP_L	6	-	1		99.62	-	-	no		Wania and Dugani (2003)
		LgS	6	-	1		99.19	-	-	no		
		LgK_{OW}	6	-	1		97.54	-	-	no		
		$LogK_{OA}$	6	-	1		99.41	-	-	no		
Braekevelt et al. (2003)	LR	LgK_{OW}	9	-	1		97	-	-	no	partial	Braekevelt et al. (2003)
Wang *et al.* (2008)	MLR	LgP_L	23	7	3		99.71	99.58	98.6	?	yes	Wong et al (2001), Tittlemier et al. (2002)
		LgK_{OA}	22	6	3		99.73	99.64	98.1	?	yes	Wania et al (2002), Harner and Shoeib (2002)
Öberg (2002)	PLS	LgP_L	23	9	?	1	99.2		99.4	yes	partial	Wong et al (2001), Tittlemier et al. (2002)
Chen *et al.* (2003)	PLS	LgK_{OA}	9 (?)	-	10		97.9	97.5	-	part	yes	Harner and Shoeib (2002)
Zhao *et al.* (2005)	MLR	LgK_{OA}	13	-	2		92.74	-	-	no	yes	Harner and Shoeib (2002)
Li et al. (2008)	PLS	LgK_{OW}	9	-	3		98.93	96.1	-	no	yes	Braekevelt et al. (2003)
US-EPA	EPI SUITE	H, MP, VP, S, LgK_{OW}, LgK_{OA}	RMSE of UI models are lower than that of EPI suite models									SRC- PhysProp
Niu *et al.* (2006)	PLS	$LgK_{p\ MET/H2O}$	15	-	6	1	95.8	95.7	-	no	yes	Eriksson et al. (2004)
		$LgK_{p\ MET}$	9	-	6	3	97,81	98,4	-	no	yes	Eriksson et al. (2004)
		$Lg\Phi_{p\ MET/H2O}$	11	-	8	3	98.2	91.4	-	no	yes	Eriksson et al. (2004)

Reference	Method	Endpoint	Train obj.	Test obj.	n°Var	n° PLS comp	R^2%	Q^2_{cum}/ R^2_{CV} %	$R(Q)^2_{ext}$ %	AD	(Mechanistic) Interpretation of desc.	Experimental Data set
Chen *et al.* (2007)	PLS	$LgK_{p\ MET/H2O}$	15	-	8	3	98.2	97.3	-	no	yes	Eriksson et al. (2004)
		$LgK_{p\ MET}$	9	-	9	2	95.8	95.8	-	no	yes	Eriksson et al. (2004)
		$Lg\Phi_{p\ MET/H2O}$	11	-	4	2	86.1	70.2	-	no	yes	Eriksson et al. (2004)
Fang *et al.* (2009)	PLS	$LgK_{p\ HEX}$	18	-	7	1	91.97	90.7	-	no	yes	Fang et al. (2009)
		$LgK_{p\ MET}$	18	-	9	1	89.87	88.3	-	no	yes	Fang et al. (2009)
		$Lg\Phi_{p\ HEX}$	18	-	5	1	78.85	75.4	-	no	yes	Fang et al. (2009)
		$Lg\Phi_{p\ MET}$	18	-	5	2	86.49	93.0	-	no	yes	Fang et al. (2009)

Existing QSAR models on Brominated Flame Retardants (BFRs)

Reference	Method	Endpoint	Train obj.	Test obj.	n°Var	n° PLS comp	R^2%	Q^2_{cum}/ R^2_{CV} %	$R(Q)^2_{ext}$ %	AD	(Mechanistic) Interpretation of desc.	Experimental Data set
Wang *et al.* (2005)	CoMFA	-LgRBA	18	-		6	99.5	58	-	no	?	Chen et al. (2001)
	CoMSIA	-LgRBA	18	-		6	98.2	68	-	no	?	Chen et al. (2001)
Wang *et al.* (2006)	Heuristic Met.	-LgRBA	18	-	4		90.3	84.4	-	no	yes	Chen et al. (2001)
Zheng *et al.* (2007)	SVM_{3f}	-LgRBA	15	3	7		-	88.9	98.5	no	yes	Chen et al. (2001)
Harju *et al.* (2002)	PLS	*Luciferase* activity	17	-		1	61.8	48.6	-	yes	partial	Meerts et al. (1998)
Harju *et al.* (2007)	PLS	-LgIC_{50} ARant	20	-	?		90	77	-	no	yes	Hamers et al. (2006)
Yang et al. (2009)	CoMSIA	-LgIC_{50} ARant	15	4		3	97.6	54.6	55.5 (r^2_{pred})	no	yes	Hamers et al. (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	Endpoint	Train obj.	Test obj.	n°Var	R^2%	Q^2_{cum}/ R^2_{CV} %	$R(Q)^2_{ext}$ %	AD	(Mechanistic) Interpretation of desc.	Experimental Data set
Rucker *et al.* (2005)	MLR / MOLGEN - QSPR	Boiling point	82	-	7	0.99	0.98		No	Yes	
Kiss et al. (2001)	NN / 3D-NET	Fluorophilicity - Fluorous partition coefficient	59	-	8		0.97		No	Yes	Kiss (1998) Szlavik (1999)
Arp et al. (2006)	ClogP Sparc EPI Suite COSMOtherm	Log P_L[a]	11	5	Prediction from commercial softwares				No	No	Shoeib et al. (2004), Krusic et al. (2005), Kaiser et al. (2005)
		Log K_{AW}[a]	4	12					No	No	Goss (2006)
		Log K_{OW}	4	12					No	No	Goss (2006)
		Log Ki[a] (hexanedecane/ air)	8	12	[a] ClogP calculation was not available for these endpoints				No	No	Goss (2006)
		Log K_{OA}[a]	7	9					No	No	Shoeib et al. (2004)
Goss (2008)	Sparc COSMO-RS	pKa	33	-					No	No	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 ebergy relationships (pp-LFER) in article

Existing QSAR models on Perfluorinated chemicals (PFCs)

Reference	Method	Endpoint	Train obj.	Test obj.	n°Var	n° PLS comp	R^2%	Q^2_{cum}/ R^2_{CV} %	$R(Q)^2_{ext}$ %	AD	(Mechanistic) Interpretation of desc.	Experimental Data set
Weiss *et al.* (2009)	PLS	T4-TTR binding	23	-	56	2		0.41		No	No	Weiss *et al.* (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	AD	(Mechanistic) Interpretation of desc.	Experimental Data set
Devillers (1999)	Neural Network	Log P	593	Few triazoles commonly used as Pesticides are studied in a bigger training set	No	No	
US- EPA	EPI Suite	MP, VP, WS, LogP		RMSE of UI models (according to OECD principles) are lower than that of EPI suite models	No	No	SRC- PhysProp

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

References	Method / Tool	Endpoint	n	n of TAZ	Test obj.	n°Var	R^2%	R^2_{CV}% /Q^2_{LOO}	AD	(Mechanistic) Interpretation of desc.
Wei Q-L et al. (2006)	MLR/SPSS	Fungicide (wheat head blight) (FA) (50 ug/ml)	18	18	-	2	96.9	-	No	No
Benfenati (2006)	fragment-based QSAR	LC_{50}-96 h acute toxicity of rainbow trout	282	Few triazoles commonly used as Pesticides are studied in the bigger set					No	No
Benfenati (2008)		LC_{50}-96 h acute toxicity of rainbow trout	125						No	No
Benfenati (2006)		Daphnia toxicity	220 (42 test)						No	No
Trohalaki (2002)	Heuristic and Best MLR / CODESSA Quantum-Chemical descriptor	EC_{25}MTT (mM)	13	2	-	Not applicable			No	No
		LSCROS (mM)							No	No
		EC_{25}LDH (mM)							No	No
		EC_{50}GSH (mM)							No	No
Klink (2003)	TOPKAT 6.0	Sensitization and/or irritancy potential	2	2	-	Use of commercial software to predict the response			No	No

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.