QSAR models for aquatic toxicity of triazoles and benzo-triazoles: WP3 results within the FP7 European Project CADASTER

Stefano Cassani[1], Simona Kovarich[1], Ester Papa[1], Partha Pratim Roy[1], Magnus Rahmberg[2], Sara Nilsson[2], Tomas Öberg[3], Nina Jeliazkova[4], Nikolay Kochev[5], Ognyan Pukalov[5], Stefan Brandmaier[6], Igor Tetko[6], Paola Gramatica[1]

[1] QSAR Research Unit in Environmental Chemistry and Ecotoxicology, University of Insubria, Varese, Italy.; [2] IVL Swedish Environmental Research Institute Ltd, Stockholm, Sweden; [3] School of Natural Sciences, Linnaeus University, Kalmar, Sweden; [4] IdeaConsult Ltd., Bulgaria; [5] University of Plovdiv, Department of Analytical Chemistry and Computer Chemistry Plovdiv, Bulgaria; [6] Helmholtz Zentrum München, German Research Center for Environmental Health.

e-mail: stefano.cassani@uninsubria.it

INTRODUCTION & OBJECTIVES

Triazoles and benzo-triazoles (BTAZs) are potentially hazardous chemicals that adversely affect humans and other non-target species, and are on the list of substances of very high concern (SVHC) in the European regulation of chemicals REACH.

TAZ/BTAZs are synthetic molecules used in various industrial processes (to obtain pharmaceuticals and agricultural products), and have a wide application as anti-corrosives, cleaning agents for textiles, flame retardants, photographic emulsions, etc... Furthermore they are abundantly used as components of liquid deicing agents for aircraft and airport runways. Because of their wide use they have been found distributed throughout the environment, mainly in water compartments. The amount of experimental data available for these molecules is insufficient for a comprehensive characterization of their environmental and toxicological profile and they have been included among the four classes of chemicals studied in the European FP7 Project CADASTER (CAse studies on the Development and Application of in Silico Techniques for Environmental hazard and Risk assessment) [1].

OBJECTIVES:

-Development of QSAR models, by different modeling approaches, for the three key organisms for the aquatic ecosystem (Algae, *Daphnia* and Fish), in order to define the potential aquatic toxicological profile of BTAZs. (Deliverable 3.5)

-Definition of *Daphnia*-Fish Interspecies Quantitative Correlation with a QAAR approach.

-Future development of consensus models, based on WP3 results, for the aquatic toxicity of BTAZs. (Deliverable 3.6)

GENERAL MATERIALS & METHODS

DATA SETS Experimental data of aquatic toxicity of BTAZs (also included in the ECHA list) were collected from the Footprint PPDB database [2]. Additional data for azo-aromatic compounds (diazines, triazines and similar heterocycles) were collected to improve the robustness and the predictivity of models.

ENDPOINTS Algae (*Pseudokirchneriella subcapitata*): EC50 96h; Daphnia (*Daphnia magna*): EC50 48h; Fish (*Oncorhynchus mykiss*): LC50 96h. Experimental toxicities (mol/L) were transformed into the logarithms of the inverse effect/lethal concentrations (pEC50 and pLC50).

FINAL GOAL OF CADASTER

To exemplify the integration of information, models and strategies for carrying out hazard and risk assessments for four classes of emerging pollutants:

- Brominated Flame Retardants
- Fragrances
- Perfluorinated Compounds
- Triazoles/ benzotriazoles

WP2 Collection of data and models

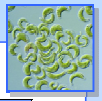
WP3 Development and validation of QSARs

WP4 Integration of QSARs within hazard and risk assessment

WP5 Development of website and stand-alone tools for dissemination and project results

	University of Insubria	IVL Swedish Environmental Research Institute	Linneaus University	HelmoltzZentrum	Idea Consult Ltd.
	UI	IVL	LnU	HMGU	IDEA
Endpoint	EC_{50} algae EC_{50} *Daphnia* LC_{50} fish	EC_{50} algae EC_{50} *Daphnia* LC_{50} fish	LC_{50} fish	EC_{50} algae EC_{50} *Daphnia* LC_{50} fish	LC_{50} fish EC_{50} *Daphnia*
Algorithm	MLR-OLS	PLSR	PLSR	kNN, ASNN, FSMLR, PLS, MLRA, SVM	MLR
Molecular Descriptors	DRAGON 5.5, PaDEL [3], CADASTER (1D-2D) [4]	DRAGON 6 (1D-2D-3D)	DRAGON 6 (1D-2D-3D)	CADASTER (1D-2D-3D)	DRAGON 5.5
Applicability Domain	Leverage	DModX	Leverage/Euclidean distance to the model	STD of ASNN	Leverage
Validation	Internal (R^2, Q^2_{LOO}, Y-sc etc) External (Q^2_{ext} F1[5]-F2-[6]-F3[7], CCC [8])	Internal (R^2,Q^2, RMSEE) External (RMSEP)	Internal (R^2, RMSE) External (Q^2_{ext}, RMSE)	Internal and External (R^2, Q^2, RMSE and MAE)	Internal (R^2, Q^2_{LOO}, Y-sc etc) External (Q^2_{ext} F1-F2-F3, CCC)
Input formats of chemical structures	Hyperchem files (.hin) for DRAGON, MOL for PaDEL and SMILES for CADASTER	Hyperchem files (.hin)	SMILES	3D sdf files prepared by Corina	SMILES

ALGAE TOXICITY



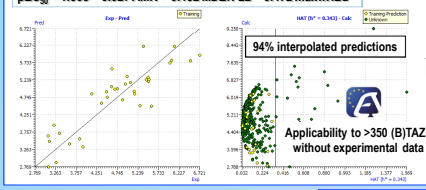
- Endpoint: pEC_{50} 72hr.
- external validation on different prediction sets; external parameters are related to split models

Partner	N Obj	Descriptors	N Descriptors	R^2	Q^2_{LOO}	RMSEtr.	RMSEp	Q^2ext	CCC
UI	35(17 BTAZ)	DRAGON 5.5, PaDEL 2.12, CADASTER online	3	0.8-0.82	0.73-0.77	0.41-0.44	0.35-0.47	0.67-0.89 *	0.82-0.91
IVL	15	DRAGON 6.0	2440	0.97	0.56	0.15	-	-	-
HMGU	35(17 BTAZ)	DRAGON 6.0	1144	0.77	0.72	0.48	0.61	<0	-

* Range of different parameters for Q^2_{EXT}

UI PaDEL model

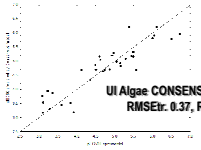
pEC_{50} = 1.505 + 0.027 AMR + 0.432 MDEN-22 + 0.472 maxwHBa



94% interpolated predictions

Applicability to >350 (B)TAZs without experimental data

Paper in preparation by UI, with a final consensus model averaging the predictions from three different developed models (DRAGON, PaDEL and CADASTER).



UI Algae CONSENSUS model RMSEtr. 0.37, R^2 0.82

DAPHNIA TOXICITY

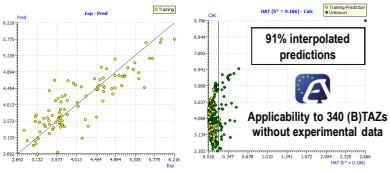
- Endpoint: pEC_{50} 48hr.
- external validation on different prediction sets; external parameters are related to split models

Partner	N Obj	Descriptors	N Descriptors	R^2	Q^2_{LOO}	RMSEtr.	RMSEp	Q^2ext	CCC
UI	97(46 BTAZ)	DRAGON 5.5, CADASTER online	5-3	0.75-0.78	0.73-0.75	0.39-0.40	0.37-0.43	0.69-0.83 *	0.85-0.89
IVL	29	DRAGON 6.0	1715	0.97	0.88	0.18	0.37	-	-
HMGU	97(46 BTAZ)	Adriana	136	0.70	0.70	0.45	0.36	0.76	-
IDEA	97(46 BTAZ)	DRAGON 5.5	5	0.85	0.83	0.34	0.31	0.4	-

* Range of different parameters for Q^2_{EXT}

UI DRAGON logP free model

pEC_{50} = 3.725 – 0.019 TPSA(NO) + 0.009 Aeigm + 0.048 nCar + 0.192 nHDon + 0.027 H-052

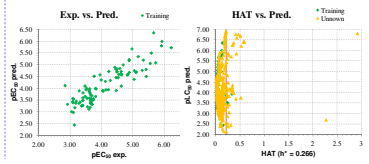


91% interpolated predictions

Applicability to 340 (B)TAZs without experimental data

IDEA model

pEC_{50} = 4.18512 - 0.53597*ATS4m + 0.06588*nCar + 0.20931*N-072 + 0.48392*ALOGP - 0.51458*nR09^2



FISH TOXICITY

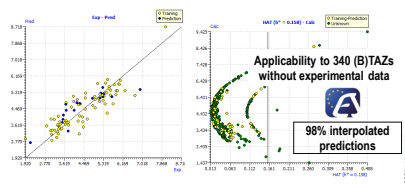
- Endpoint: pLC_{50} 96hr.
- external validation on a common evaluation set (EV, n=18 BTAZ)

Partner	N Obj	Descriptors	N Descriptors	R^2	Q^2_{LOO}	RMSEtr.	RMSEp	Q^2ext	CCC
UI	76 (28 BTAZ)	DRAGON 5.5, PaDEL 2.12	4-3	0.79-0.82	0.77-0.79	0.48-0.51	0.41-0.58	0.71-0.87 *	0.81-0.93
IVL	19	DRAGON 6.0	2275	0.98	0.79	0.18	0.54	-	-
LnU	76 (28 BTAZ)	DRAGON 6.0	677	0.94	0.81	0.29	0.48	0.81	-
HMGU	76 (28 BTAZ)	ChemAxon	90	0.63	0.63	0.68	0.60	0.67	-
IDEA	79 (30 BTAZ)	DRAGON 5.5	5	0.83-0.85	0.79-0.81	0.51-0.55	0.45-0.49	0.77-0.79 *	0.81 -0.93

* Range of different parameters for Q^2_{EXT}

UI PaDEL model

pLC_{50} = 2.486 + 0.367 VP-1 – 0.047 SHBint2 - 0.312 maxHaaCH

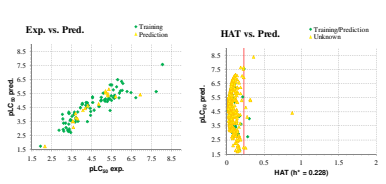


Applicability to 340 (B)TAZs without experimental data

98% interpolated predictions

IDEA model

pLC_{50} = 2.94479 + 10.80548 Mp+(-0.13221) nN+(-8.29644) SIC1+ + 0.5821 EEig07d+(-0.25934) O-058



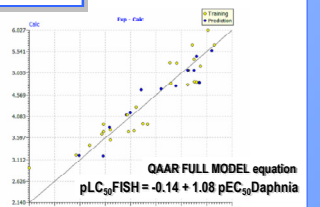
UI Quantitative Activity-Activity Relationship (QAAR) based on Interspecies Correlation : DAPHNIA-FISH

QAAR (Quantitative Activity-Activity Relationships) model based on interspecies correlations were developed by UI to provide direct estimation of the acute aquatic toxicity of untested BTAZs from *Daphnia* to Fish.

The here proposed QAAR model was obtained by using EC_{50}48h measured in *Daphnia* as independent variable, and LC_{50}96h measured in Fish as response endpoint.

MODELS	N_{TR}	N_P	R^2	Q^2_{LOO}	Q^2_{EXT}*	$RMSE_{tr}$	$RMSE_p$
Split	27	13	0.85	0.82	0.91-0.93	0.34	0.24
FULL	40	--	0.87	0.85	--	0.31	--

* Range of different parameters for Q^2_{EXT}



QAAR FULL MODEL equation

$pLC_{50}FISH = -0.14 + 1.08\ pEC_{50}Daphnia$

UI ECOSAR comparisons

Predicted data, derived from the UI local models here proposed for aquatic toxicity of BTAZs, were compared with those obtained by the freely available tool ECOSAR [9] ("triazoles"equation). The accuracy in prediction of our models and ECOSAR models was compared by calculating the parameter RMSE (* range of different UI models).

MODEL	*Algae* (n=12 BTAZs) RMSE	*Daphnia* (n=32 BTAZs) RMSE	Fish (n=33 BTAZs) RMSE
UI	0.29-0.41*	0.44-0.45*	0.47-0.53*
ECOSAR	0.51	0.63	0.84

CONCLUSIONS and future work

- Different robust and externally predictive QSAR models have been developed by WP3 partners, with different modeling approaches, to predict the aquatic toxicity of BTAZs in Algae, *Daphnia* and Fish.
- Quantitative Activity-Activity Relationship (QAAR) based on Interspecies Correlation, externally validated, has been developed by UI to predict Fish acute toxicity from *Daphnia* toxicity data.
- Consensus modeling approach will be applied for the aquatic toxicity of BTAZ, considering all the results obtained by WP3 partners. This work will be presented in the Deliverable 3.6.

REFERENCES

[1] CADASTER FP7 PROJECT - www.cadaster.eu.

[2] Footprint PPDB (Pesticide Properties DataBase), available on-line at: http://sitem.herts.ac.uk/aeru/footprint/en/index.htm.

[3] Yap C.W. *J. Comput. Chem.* 2011, 32 (7): 1466-1474. available on-line at: http://padel.nus.edu.sg/software/padeldescriptor/index.html

[4] www.cadaster.eu/database/ (http://qspr-thesaurus.eu/static/home.do)

[5] Shi L. M.; Fang H.; Tong W.; Wu J.; Perkins R.; Blair R. M.; Branham W. S.; Dial S. L.; Moland C. L.; Sheehan D. M. J Chem Inf Comput Sci 2001, 41, 186-195.

[6] Schüürmann, G.; Ebert R.U.; Chen J.; Wang B.; Kühne R. J Chem Inf Model 2008, 48, 2140-2145.

[7] Consonni, V.; Ballabio, D.; Todeschini, R. J Chem Inf Model 2009,49, 1669-1678.

[8] Chirico N., Gramatica P. J Chem. Inf. Model. 2011,51, 2320-2335.

[9] EPI Suite v.4.1. http://www.epa.gov/oppt/exposure/pubs/episuite.htm

Financial support, given by European Union through the project CADASTER (FP7-ENV-2007-1-212668), is gratefully acknowledged