Overview of activities within WP 3

Development and validation of QSAR models

Paola Gramatica

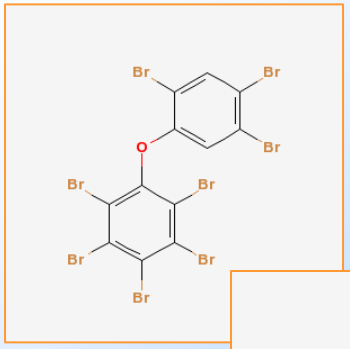
WP3 leader

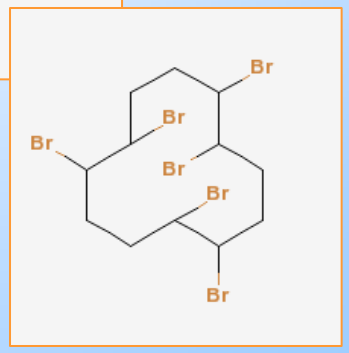
QSAR Research Unit in Environmental Chemistry and Ecotoxicology
DBSF -University of Insubria, Varese - Italy

E-mail: paola.gramatica@uninsubria.it
http://www.qsar.it

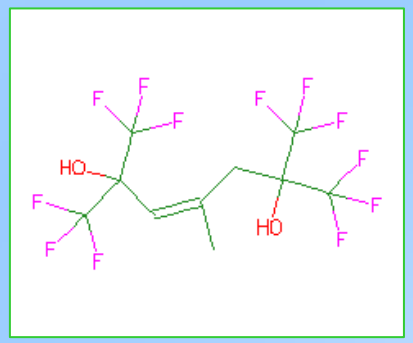
Emerging Pollutants – CADASTER classes

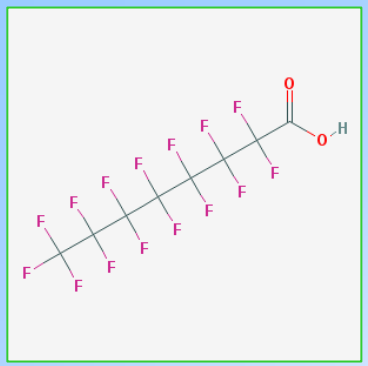
PolyBrominated DiPhenyl Ethers (PBDE) and other Brominated Flame Retardants (BFRs)



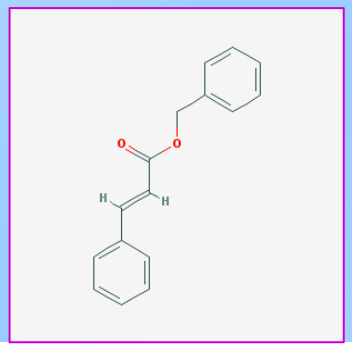


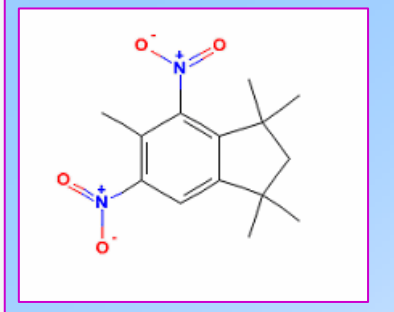
Poly- and Perfluorinated compounds (PFCs)



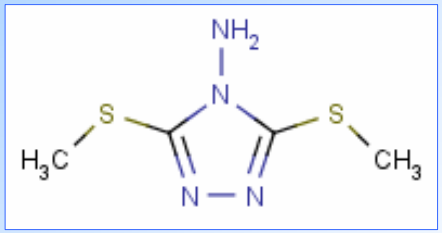


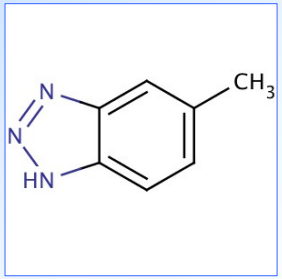
Substituted musks/Fragrances





Triazoles and Benzotriazoles (B)TAZs





CADASTER

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

WP3 Partners

- University of Insubria (Italy)- P. Gramatica
- Linneus University (Sweden)- T. Oberg
- IVL Swedish Environ. Res. Inst. (Sweden)- M. Rahmberg
- IDEAConsult Ltd. (Bulgaria) - N. Jeliazkova
- Helmholtz Zentrum München (Germany) - I.Tetko

WP3 Tasks

- Task 3.1 - Preparation of the chemical structures and molecular descriptors database for the chemicals of the four selected classes.
- Task 3.2 - Evaluation of existing QSARs.
- Task 3.3 - Gap analysis.
- Task 3.4 - Prioritization for experimental tests.
- Task 3.5 - Development of new QSARs.
- Task 3.6 - Development of multi-model approaches (consensus modeling).
- Task 3.7 - External Validation of QSAR models with experimental data from the Project.

CADASTER Partners at 21° SETAC meeting May 2011, Milan (Italy)

QSAR Research Unit

in Environmental Chemistry and Ecotoxicology

http://www.qsar.it

DBSF - University of Insubria
Varese - Italy

Staff

Prof. Paola Gramatica

Dr. Ester Papa, Ph.D

Dr. Simona Kovarich → Brominated Flame Retardants

Dr. Mara Luini → Fragrances

Dr. Barun Bhhatarai, Ph.D → Perfluoroalkylate Substances, Triazoles & Benzotriazoles

Dr. Jiazhong Li, Ph.D

CADASTER

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

Task 3.1 - Preparation of the chemical structures and molecular descriptors database for the chemicals of the 4 selected classes.

a) Minimization of Chemical Structures by Hyperchem

b) Calculation of DRAGON molecular descriptors

Database

c) Calculation of CADASTER online molecular descriptors

OECD Principles for QSAR models in REACH

To facilitate the consideration of a QSAR model for regulatory purposes, it should be associated with the following information:

- a defined endpoint
- an unambiguous algorithm
- a defined domain of applicability
- appropriate measures of goodness of fit, robustness and predictivity
- a mechanistic interpretation, if possible

- In literature, very few local models for the specific CADASTER classes, mainly
  - not validated and without information on their Applicability Domain (AD)
- Only general models (as EPISUITE), mainly without AD
- Very few experimental data for SIDS end-points, useful for Risk Assessment

- Need for specific **local models for CADASTER chemicals** verifying external validation and AD

Task 3.4 - Prioritization for experimental tests.

Prioritization through structural similarity analysis and ranking methods, based on available toxicity data

BFRs
prioritization based on toxicity profile (exp+pred ED potency) (UI), structural diversity (LNU/UI)

PFCs
prioritization based on toxicity profile (exp+pred mammalian toxicity) (UI), structural representativity (LNU/UI)

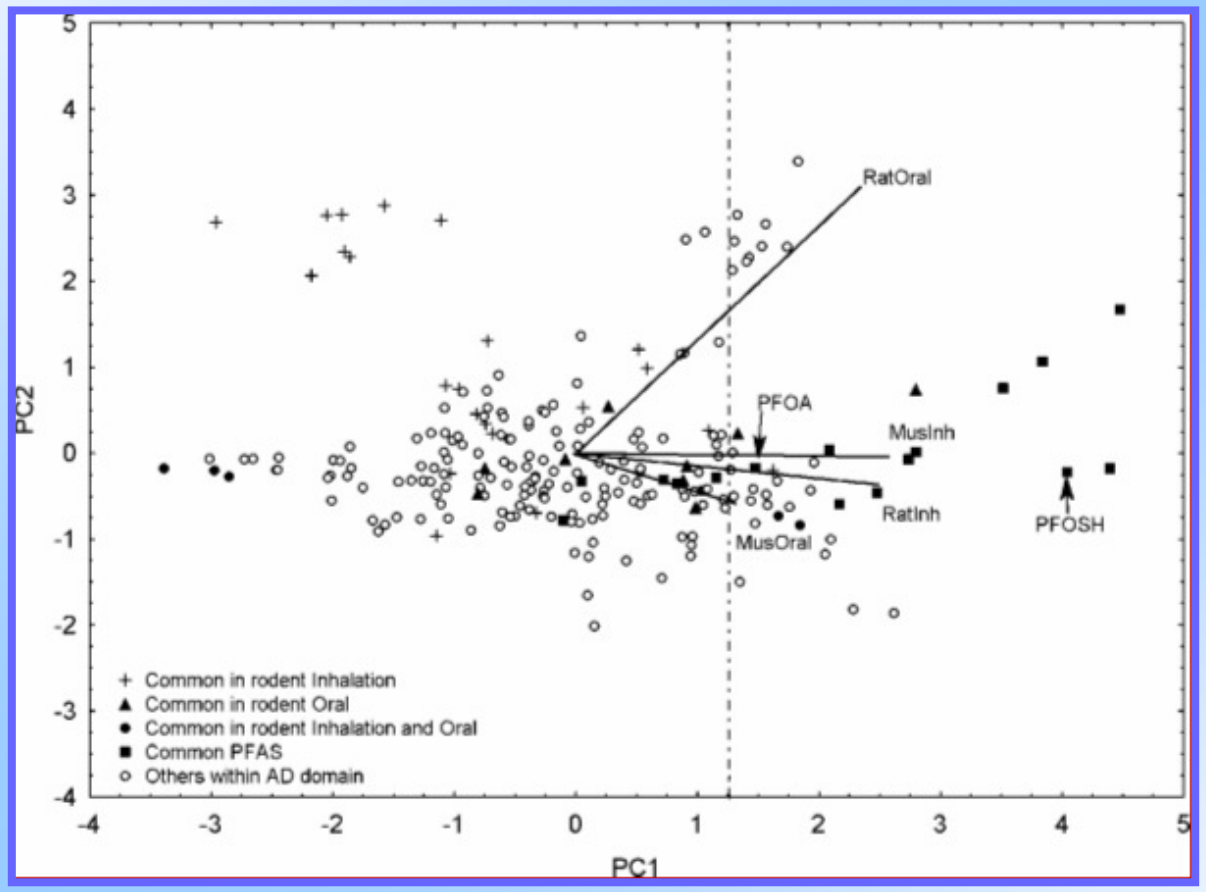
FRAGRANCES
prioritization based on toxicity profile (exp+pred mammalian and cyto-toxicity), structural representativity (UI).

(B)TAZs
prioritization based on toxicity profile (exp aquatic toxicity) (UI) and structural diversity (UI-LNU).

and compound's commercial availability.

Task 3.4 :
Prioritization for experimental tests in the Project

Prioritization based on toxicity profile (exp+pred mammalian toxicity: rat and mouse oral and Inhalation toxicity) (UI)



PFOA and PFOSH among the more toxic

Task 3.5 - Development of new QSARs.

QSAR prediction of physico-chemical properties and biological activities

- Development of QSAR models for **available** end-points (sufficient experimental data for modeling), paying attention to external validation and structural applicability domain analysis.

- Identification of more toxic and environmental hazardous chemicals based on the studied end-points

Summary on BFRs

Endpoint (N_{obj})	SIDS		Model D 3.5	Desc	Data set	Pred & AD	LNU	Ref
	Phys-chem	Fate	(*on-line) UI			(243 BFRs)		
H * (7)	X		MLR *	Dragon 5.5	Published + CADASTER web	Published + uploaded in CAD. Database		*a*
MP * (25)	X		MLR *	" "				*a-b*
VP * (34)	X		MLR *	" "				*a*
WS * (12)	X							*a*
$logK_{OA}$*(30)					STER			*a-b*
$logK_{OW}$*(20)	X				base			*a-b*
Kp (15)		X	MLR	" "	CADASTER Database			
HLp (15)		X	MLR	" "				
BCF							X	*c*
ED Activity (~20)					ublished + AD. Dat.	" "		*d-e*

Prediction of SIDS endpoints and characterization of environmental behaviour

Exp and Pred data used for Prioritization (Del 3.4) → experimental testing

Predictions compared with EPI Suite software

[a] Papa, E.; Kovarich, S.; Gramatica P., *QSAR Comb. Sci.* 2009, 28, 790-796.
[b] Papa, E.; Kovarich, S.; Gramatica P., *Molecular Informatics* 2011 , 30, 232-240 (Proceedings issue of Euro-QSAR2010).
[c] Oberg T., *Environ Sci & Pollut Res* 2002, 9, 405-411.
[d] Papa, E.; Kovarich, S.; Gramatica, P., *Chem. Res. Toxicol.* 2010, 23, 946-954.
[e] Kovarich, S.; Papa, E.; Gramatica, P., *J. Hazardous Mat.* 2011,190, 106-112.

Summary on Fragrances

Endpoint (N_{obj})	SIDS		Model Del 3.5 (* on-line) UI	Descriptors	Data set	Pred & AD (79 frag)	Ref
	Phys-chem	Tox					
VP* (37)	X		MLR	Dragon 5.5	Exp data uploaded in CADASTER web / base		a
WS* (37)	X		" "	" "			
logKow* (37)	X						
Rodent acute toxicity (23)					a set shed + STER eb / Database	Published	b
Cytotoxicity (~20)							

Prediction of SIDS endpoints and characterization of environmental behaviour

Exp and Pred data used for Prioritization (Del 3.4) → experimental testing

* Predictions compa[illegible] E[illegible]

[a] Papa, E.; Luini, M.; Gramatica, POSTER presented at SETAC-Europe 2009, Göteborg, Sweden, 31 May – 4 June 2009.

[b] Papa, E.; Luini, M.; Gramatica, P. *SAR QSAR Environ. Res.*, 2009, 20, 767–779.

Summary on PFCs

Endpoint (N_{obj})	SIDS		Model & Desc (* on-line) UI	Data set	Pred & AD (221 PFCs)	Ref	Models by WP3 Partners	Cons ensus
	Ph-ch	Tox						
MP* (94)	X		MLR-Drag5.5			*a*	LNU, IDEA, HMGU	X
BP* (93)	X							X
VP* (35)	X					*b*	LNU (PLS) ref c	
WS* (20)	X							
CMC (10)			MLR-Drag5.5	CAD.				
Rodent acute toxicity (~55)		X				*d-e*		
ED activity (24)						*f*		

Prediction of SIDS endpoints and characterization of environmental behaviour

Exp and Pred data used for Prioritization (Del 3.4) → experimental testing

*** Predictions compared with EPI Suite software**

[a] Bhhatarai, B. et al. (**WP3 partners**), *Molecular Informatics* , 2011 30, 189-204 (Proceedings issue of Euro-QSAR2010).

[b] Bhhatarai, B.; Gramatica, P., *Environ. Sci. Technol.,* 2010, Online first (2011 45: Special issue on PFCs), Oral at CMTPI11

[c] Oberg T.; Liu T., *Chemo Lab,* 2011, 107, 59-64.

[d] Bhhatarai, B.; Gramatica, P. , *Molecular Diversity*, 2010, 15, 467-476.

[e] Bhhatarai, B.; Gramatica P., *Chem. Res. Toxicol.,* 2010, 23, 528-539.

[f] Kovarich, S.; Papa, E., Gramatica, P Submitted to *SAR & QSAR Environ. Res.*, 2011.

Summary on (B)TAZs

Endpoint (N_{obj})	SIDS		Model D3.5 (*on-line) UI	Desc	Data set	Pred&AD (351 TAZ)	Ref UI	WP3 Partners
	Ph-ch	Ecotox						
WS* (49)	X		MLR *	Drag 5.5	Published	Published	*a*	
LogKow* (64)	X							
VP* (33)	X				eb / ase			
MP* (56)	X							
EC_{50} Algae		X					*b*	
EC_{50} Daphnia		X	in progress					
LC_{50} Fish		X			xp data loaded AD. web atabase			in progress
BCF								
LC_{50} Earthworm		X						
LD_{50} Honeybees		X	(MLR)					
LD_{50} Bird		X						

Prediction of SIDS endpoints and characterization of environmental behaviour

Experimental data used for Prioritization (Del 3.4) → experimental testing

* Predictions compared with EPI Suite software

[a] Bhhatarai, B.; Gramatica, P., *Water Research*, 2011, 45 (3) 1463-1471. Oral at CMTPI11

[b] Kovarich S.et UI group, Oral at CMTPI1

WP3 in WP5 : Models in CADASTER web

A limited amount of developed and published local models for CADASTER chemicals are now uploaded in the CADASTER database:

- **BFRs:** VP, MP, logKow, logKoa
- **PFCs:** BP, WS, VP
- **(B)TAZs**: WS, logKow

Better results than EPI Suite and check of structural AD: **definition of interpolation or extrapolation**

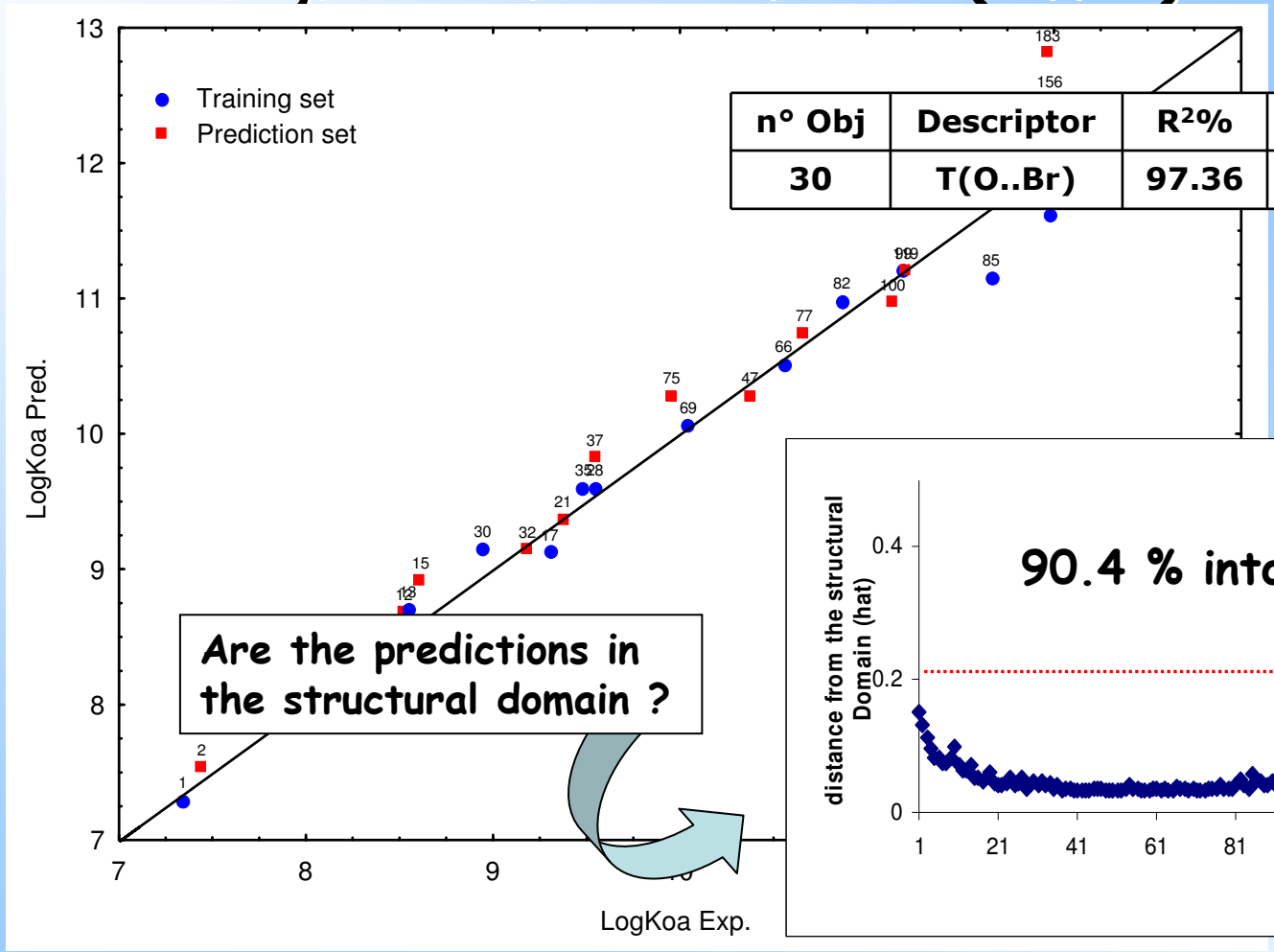
Ongoing modelling for (B)TAZ are based also on descriptors calculated in the CADASTER database → uploadable models

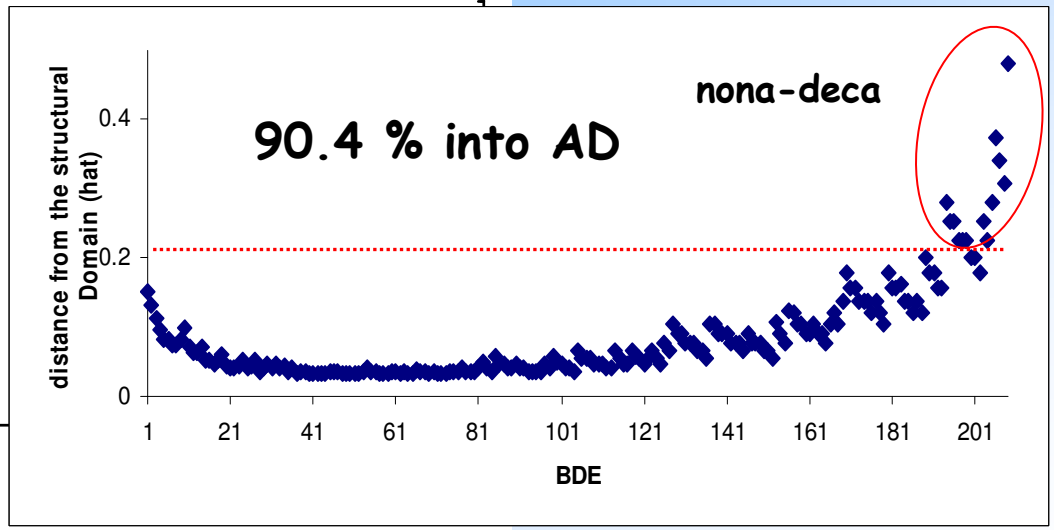
Example: Local Model for Log Koa of PBDE

Experimental range of LogKoa: 7.34 (mono-BDE) – 11.96 (hepta-BDE)

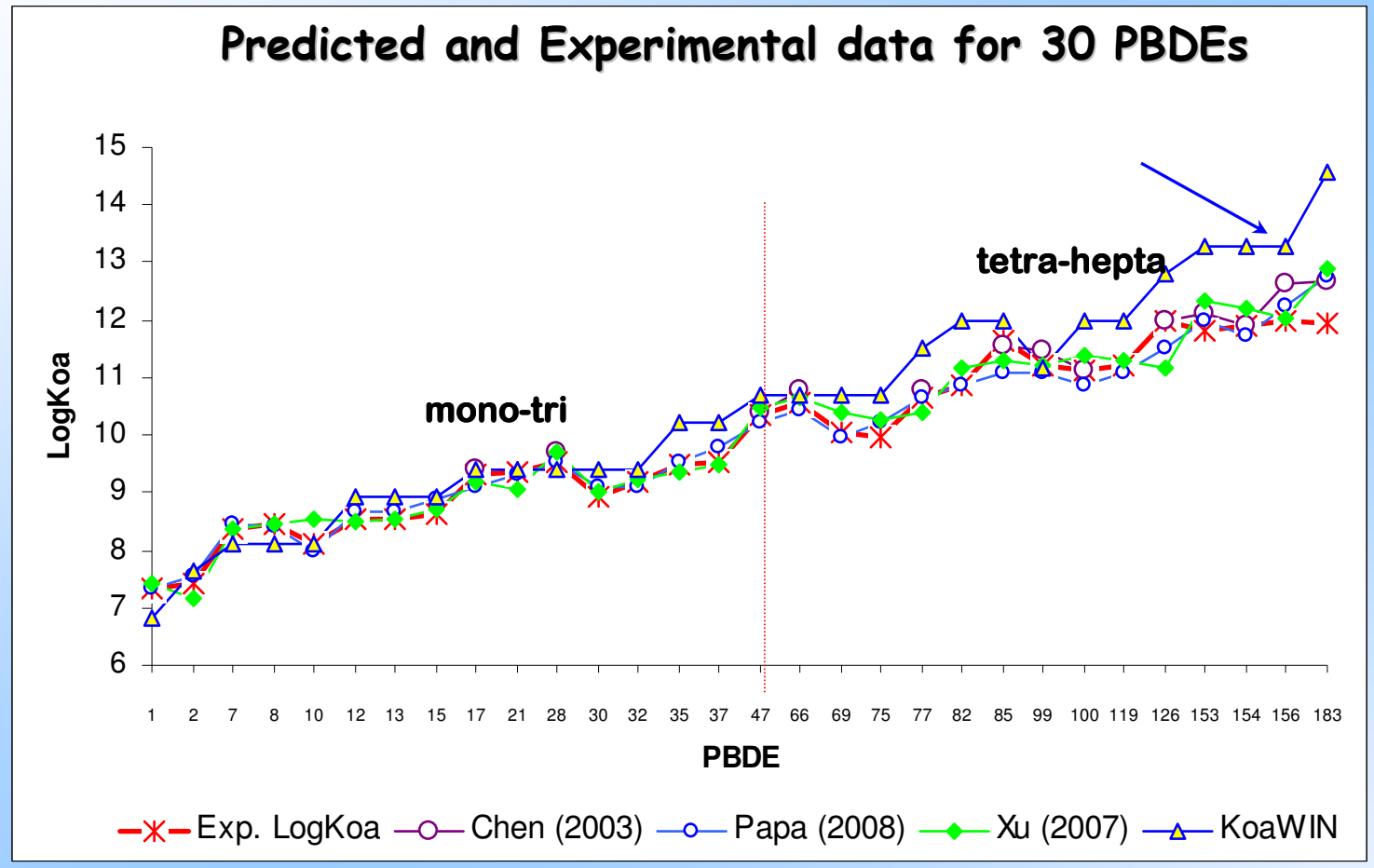
LogKoa= 6.654 +0.222 T(O..Br)

n° Obj	Descriptor	R^2%	Q^2_{boot}%	$Q^2_{EXT(rand20\%)}$ %
30	T(O..Br)	97.36	96.77	99.56





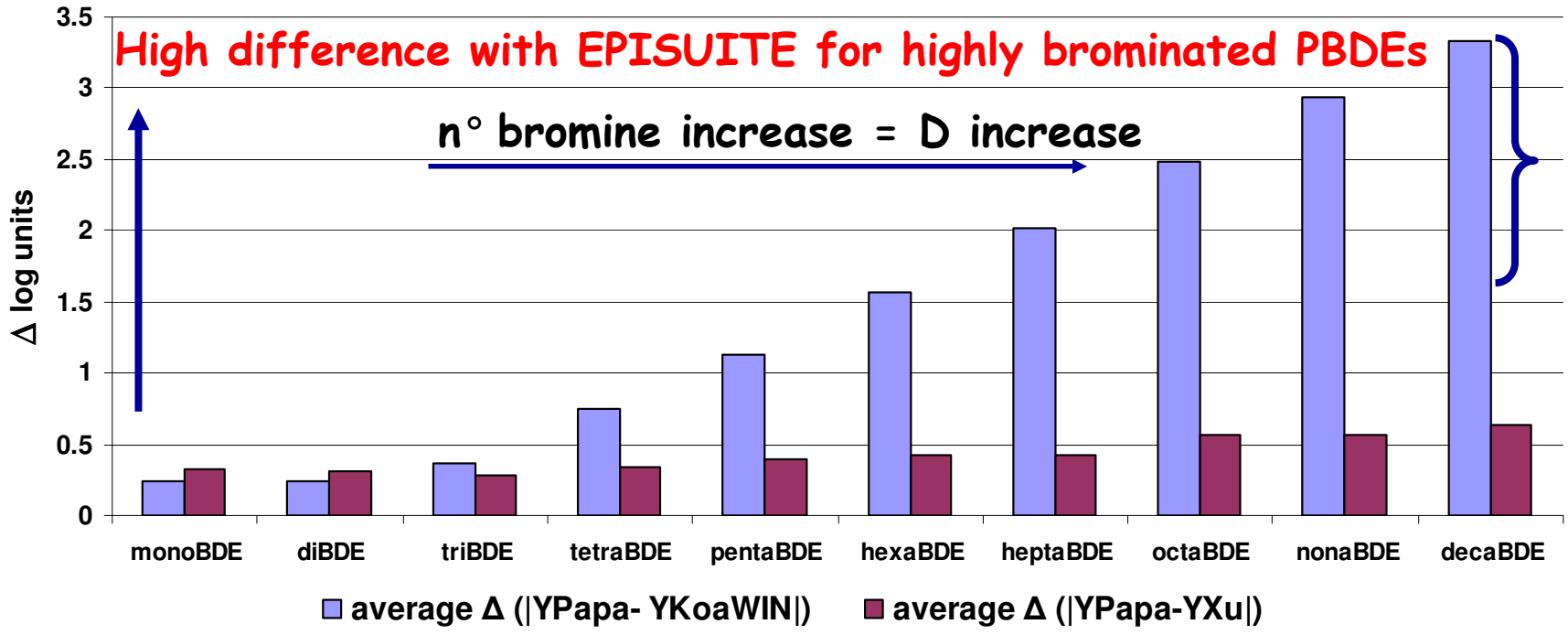
Comparison with some existing models, in particular EPISuite



Author	Method	N° obj.	N° vars	R^2%	Q^2_{LOO}%	Q^2_{EXT} %	RMSE (30 obj)
Papa et al.	MLR	30	1	97.4	96.8	99.6	0.23
Xu et al.	MLR	22	2	97.6	97.2	-	0.31
Chen et al.	PLS	13	10	97.9	97.5	-	-
KoaWIN (Episuite)	K_{OW}/K_{AW}						0.81

Comparison with some existing models, in particular EPISuite

Predictions for all the 209 PBDEs congeners



Y_{Papa} = Predictions by our model (range Log Koa: 7.32 – 15.09)
$Y_{Episuite}$ = Predictions by KoaWIN (Dmax = 3.33 log units; range Log Koa: 6.81-18.23)
Y_{Xu} = Predictions by Xu et al. (2007) (Dmax =1.06 log units; range Log Koa: 7.4-15.73)

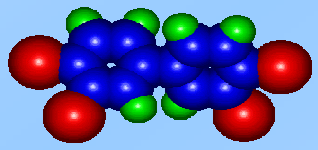
Papa, E.; Kovarich, S.; Gramatica P., *Molecular Informatics* 2011 , 30, 232-240 (Proceedings issue of Euro-QSAR2010).

Aim of WP3 in CADASTER Project

to develop and propose in:

http://www.cadaster.eu

- local QSAR models
- specific for the 4 classes of emerging pollutants
- externally validated
- verified for their structural Applicability Domain



Thanks to all for your attention !!

http://www.qsar.it

paola.gramatica@uninsubria.it

Prof. Paola Gramatica - QSAR Research Unit - DBSF - University of Insubria - Varese (Italy)