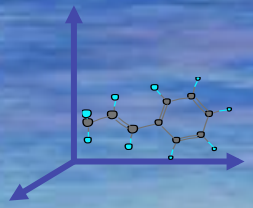
Integrated Testing Strategies (ITS) and QSAR Modeling

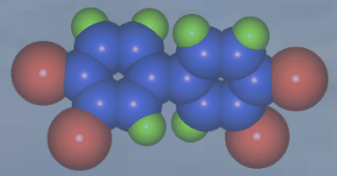


Prof. Paola Gramatica

Unità di Ricerca QSAR in Chimica Ambientale ed Ecotossicologia

DiSTA- Università dell' Insubria - Varese

http://www.qsar.it



L' UNIVERSO CHIMICO oggi

NUOVI
6.000.000 / anno

> 74.000.000 in C.A.S. (Nov/ 13)

> 308.000 regolate

NUOVI
10.000 / anno

EINECS TSCA
100.204

5-10%
dati
noti

Q
S
A
R

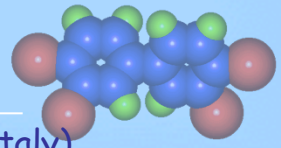
Metodi predittivi

esperimenti

Destino ambientale?

Effetti sull' uomo e sull' ecosistema?

EU-REACH



Nuova Legislazione Europea per le Sostanze Chimiche (1 giugno 2007) (White Paper 2/2001)

REACH: Registration Evaluation Authorisation and Restriction of Chemicals

Art. 1 Per assicurare:

- Alto livello di protezione per salute umana ed ambiente
- Promozione di metodi alternativi per la valutazione dei pericoli delle sostanze
- Libera circolazione delle sostanze
  (on their own, in preparations and in articles)
- Incremento della competitività ed innovazione

Situazione pre-REACH

Più di 40 legislazioni diverse

Diverse soprattutto per "existing" and "new" (cut-off 1981)

"Existing" chemicals: sul mercato EC tra 1/1/71 and 18 / 9/81 (EINECS: European Inventory of Existing Commercial Chemical Substances 100.204sostanze diverse)

"New" chemicals: introdotte nel mercato EC dopo il 1981, 4381 substances (1/23 of existing) ELINCS (European LIst of Notified Chemical Substances),

- "New" chemicals : con dati di registrazione (propr.Ch-Fis. e Tossicologiche) prima dell'immissione sul mercato
- "Existing" chemicals: non c'era tale obbligo

Mancanza di sufficiente informazione per la maggior parte delle sostanze

Situazione pre-REACH

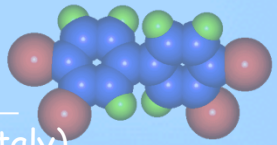
- **141 HPVCs valutati dall'autorità competente in 14 anni**

- **Elevato livello di non-conoscenza.**
  La maggior parte delle sostanze non è stata valutata e oltre l'80% non possiede dati sperimentale "di base"
  Public Availability of Data on HPVCs (Allanou, Hansen and van Der Bilt, 1999)

  - 14 %: base set data
  - 65%: less than base set
  - 21%: no data

  } 86%

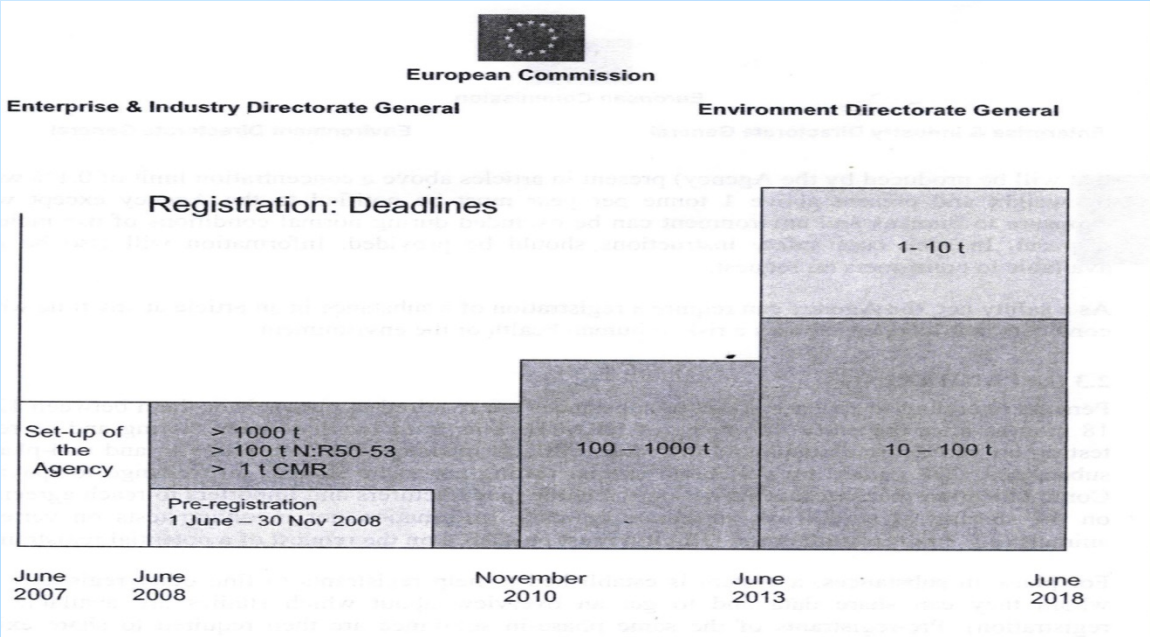
La valutazione completa di tutti i 2700 HPVC richiederebbe oltre 250 anni. (Schaafsma et al., 2009)



Registrazione a tappe : giugno 2007- giugno 2018

30.000 sostanze, ora in uso, da registrare in 11 anni

European Chemicals Agency (ECHA) in Helsinki



Pre-registrazione: per facilitare "data sharing" e ridurre
test su animali
costi per industrie

Sostanze ad alta pericolosità (CMR, ED e potenziali PBT/vPvBs) devono essere registrate prima

Attività richieste per la registrazione

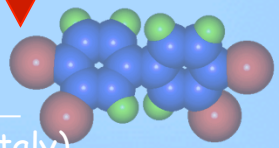
		Registrazione				Valutazione	Autorizzazione
		1-10 tonnellate/ anno	10-100 tonnellate/ anno	100-1.000 tonnellate/ anno	>1.000 tonnellate/ anno	- *Su richiesta della Europea Chemical Agency* - *Previsti per sostanze pericolose**	
Analisi richieste	**Proprietà chimico-fisiche**	←	- Punto di fusione/ congelamento - Punto di ebollizione - Densità relativa - Pressione del vapore - Tensione superficiale - Solubilità in acqua - Coefficiente di ripartizione		→	- Analisi ad hoc su richiesta della European Chemical Agency	- Utilità socio-economica della sostanza - Eventuale dimostrazione dell'impossibilità di sostituzione
				- Viscosità			
	Analisi tossicologiche	←	- Irritazioni cutenee/oculari - Mutagenicità su batteri		→		
			- Mutagenicità in vitro su cellule di mammiferi - Tossicità per la riproduzione, orale/epidermica, per dosi - Tossicocinetica		→		
					- Tossicità riproduttiva su seconda generazione		
	Eco-tossicità e studi ambientali	←	- Tossicità in acqua (Daphnia)		→		
			- Tossicità per fauna acquatica/alghe - Biodegrabilità				
				- Tossicità cronica su Daphnia	→		
Raccolta di informazioni			← - Notifica degli utilizzi		→	- Documentazione ad hoc	- Scenari di esposizione
Documentazione a supporto			- Chemical Safety Report - Safety Data Sheet, classificazione ed etichettatura		→	- Documentazione ad hoc	- Rapporto di adeguata copertura dei rischi

CMR (carcinogenicity, mutagenicity and reproductive toxicity), PBT (Persistent, bioaccumulative and Toxic) e vPvB (very persistent and very bioaccumulative)
onte: Commissione Europea

Dati richiesti dal REACH

Annex	Human Health	Environment
Annex VII (≥ 1 tpa)	•*In vitro* skin and eye irritation •Skin sensitisation •*In vitro* mutagenicity •Acute toxicity (one route)	•Short-term toxicity (daphnia, algae) •Degradation (biotic)
Annex VIII (≥ 10 tpa)	•*In vivo* skin and eye irritation •Further *in vitro* mutagenicity •Acute toxicity (2nd route) •Short-term RdT (28 days) •Reproductive toxicity screening •Assessment of toxicokinetics (not a testing requirement)	•Short term toxicity (fish) •Respiration inhibition test •Degradation (hydrolysis) •Fate (absorption/desorption)
Annex IX (≥ 100 tpa)	•Further mutagenicity studies (if + results) •Sub-chronic toxicity (90-days) •Reproductive toxicity tests	•Long-term toxicity (invertebrates, fish) •Biotic degradation (simulation studies) •Identification of degradation products •Fate: bioaccumulation in fish, further absorption/desorption •Short term toxicity- terrestrial organisms (invertebrates, MO, plants)
Annex X (≥ 1000 tpa)	•Further mutagenicity studies (if + results) •Further reproductive toxicity studies •*Chronic toxicity (may)* •*Carcinogenicity (may)*	•Further biotic degradation •Further fate •Long-term effects on terrestrial organisms •Long-term or reproductive toxicity to birds

Aumento costi



REACH: Stima dei costi (2009-2018)

Dati noti

Dati ignoti

30.000 Sostanze da registrare

2.1 - 3.9 Milioni di organismi vertebrati sacrificati

1.2 - 2.4 Miliardi di Euro per effettuare i tests

(EU commission - Van der Jagt 2004)

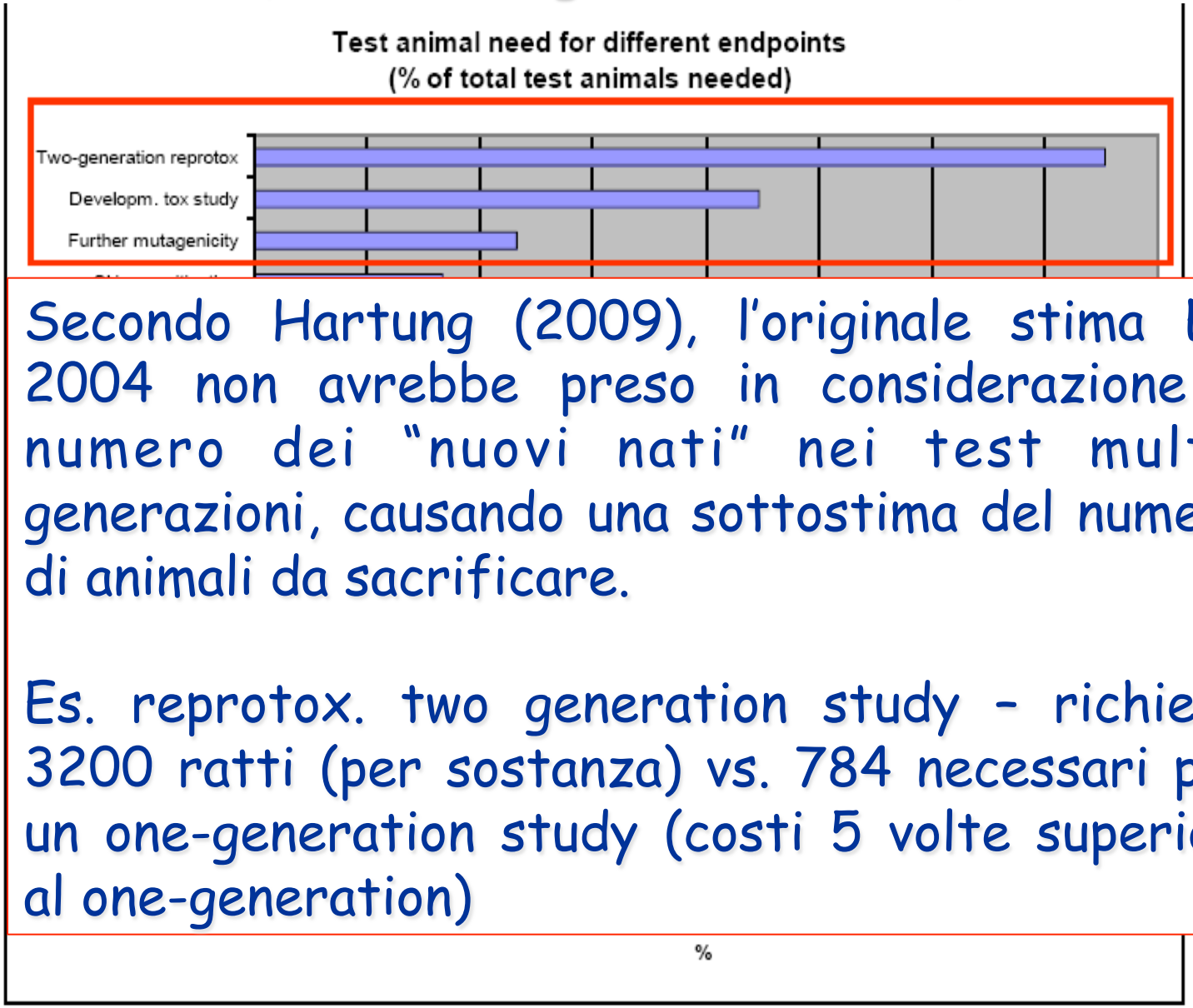
Nuove stime riportate da Hartung e Rovida (2009):

68.000 Sostanze da registrare

54 Milioni di vertebrati

9.5 Miliardi di Euro

Stima del fabbisogno di animali in diversi test (van der Jagt. et al. 2004)



Secondo Hartung (2009), l'originale stima EU 2004 non avrebbe preso in considerazione il numero dei "nuovi nati" nei test multi-generazioni, causando una sottostima del numero di animali da sacrificare.

Es. reprotox. two generation study - richiede 3200 ratti (per sostanza) vs. 784 necessari per un one-generation study (costi 5 volte superiori al one-generation)

Authorisation

Sostanze con caratteristiche di alta pericolosità devono essere soggette alla **authorisation**; l' Agenzia pubblicherà una lista di tali sostanze:

*Carcinogens Mutagens and Toxic to reproduction (CMR) , Endocrine Disruptors (ED)
Persistent Bioaccumulative Toxic (PBT, vPvB).

Finora 900, più > 600 possibili nuove individuate dal REACH: > 1500.

- Bando totale o parziale per rischi inaccettabili
- Il sistema di autorizzazione incoraggerà le industrie a cercare alternative più sicure.
- Infatti, ogni richiesta di autorizzazione deve includere una analisi delle alternative ed un piano di sostituzione.

"Data sharing" e riduzione dei tests su animali in REACH

Per ridurre al minimo l'uso di test su animali ed incoraggiare l'uso di metodi alternativi:

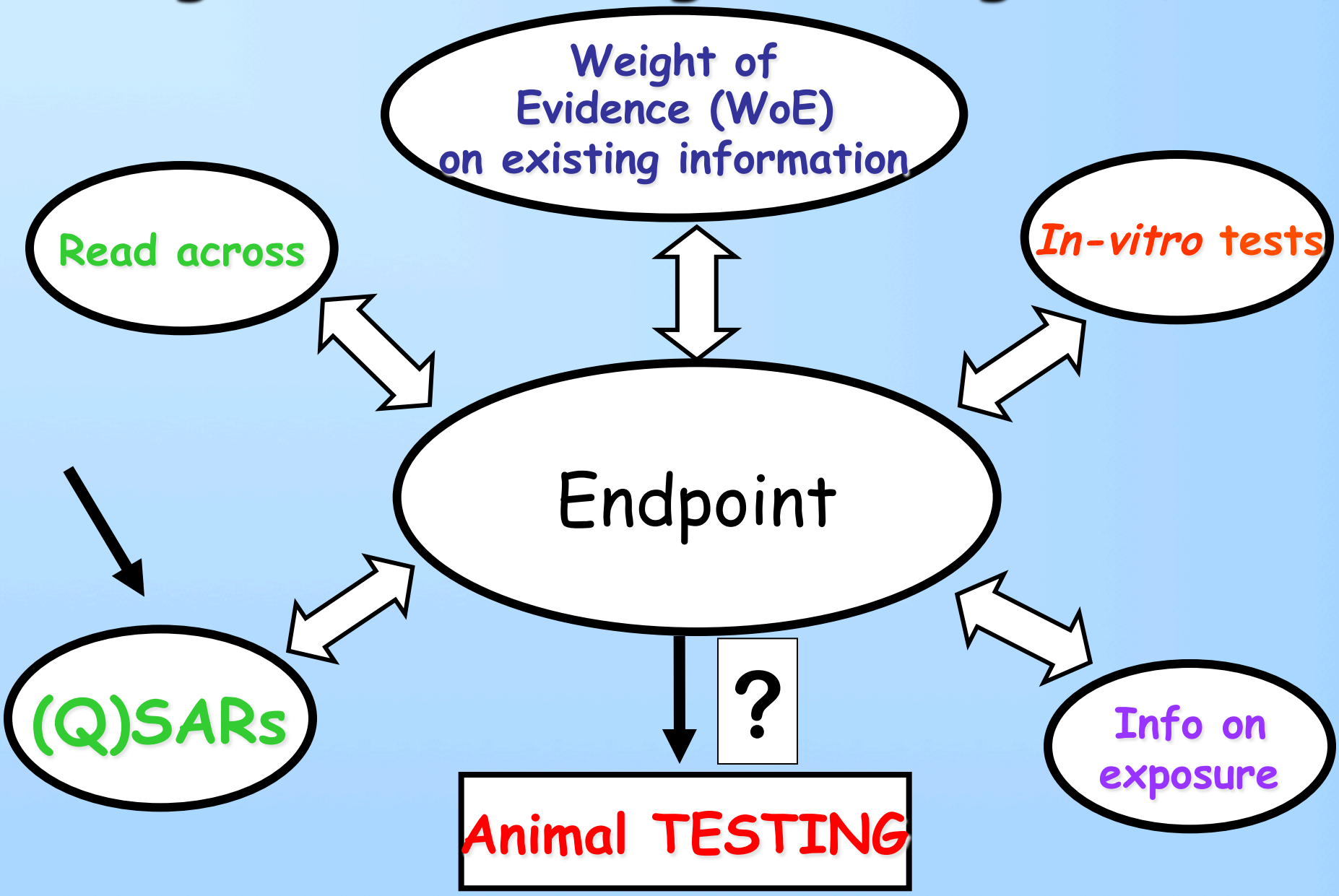
- Dati ottenuti con test su vertebrati devono essere condivisi, su pagamento
- Informazioni possono essere ottenute con metodi alternativi (*in vitro o in silico*, es. QSAR) (Art. 13) devono essere condivisi.

IUCLID5:database di dati esistenti
http://iuclid.echa.europa.eu/

Regole Generali per:
-uso di informazione esistente
-per sviluppo ed accettabilità di modelli *in silico* (es. QSAR)

Nuovi test sono richiesti solo (Art. 25: Ultima risorsa!) se non si ottengono informazioni in altri modi validati (*to replace, reduce or refine animal testing*, 3R approach)

Integrated Testing Strategies (ITS)



Potenzialità di "risparmio" con le ITS

Costo del REACH senza utilizzo di ITS:

- Numero di animali: 3.9 milioni di organismi vertebrati
- Costo dei tests: 2.4 miliardi di Euro

(Stime EU)

Costo del REACH con utilizzo di ITS:

- Numero di animali: 1.3-1.9 milioni
- Costo dei tests: 800-1130 milioni di Euro

(Van der Jagt et al., 2004; EUR report 21405)

QSAR use	Test Cost M Euro				
	1-10 t/a	10-100 t/a	100-1000 t/a	>1000 t/a	Tot
no QSAR	230	690	510	710	2130
Average QSAR	150	350	330	610	1430
Max QSAR	130	260	260	540	1190

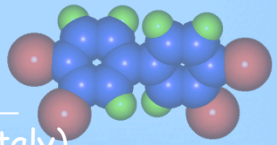
(EU-JRC)

(Q)SAR nel testo REACH

16 citazioni : Art.13, Art. 138, Annex III, Annexes VI-XI.

E' previsto l'utilizzo di dati ottenuti da modelli predittivi QSAR :

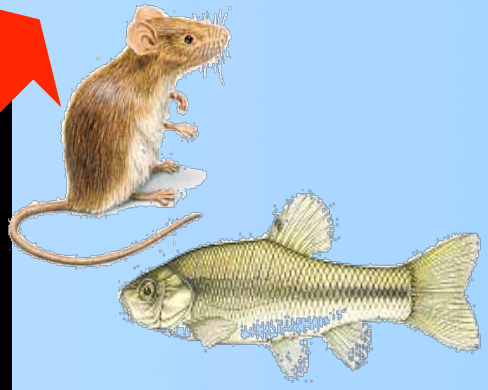
- Per dare indicazioni su sostanze più pericolose (liste di priorità)
- Per indirizzare la sperimentazione su sostanze prioritarie (focalizzazione esperimenti)
- Per colmare la mancanza di dati



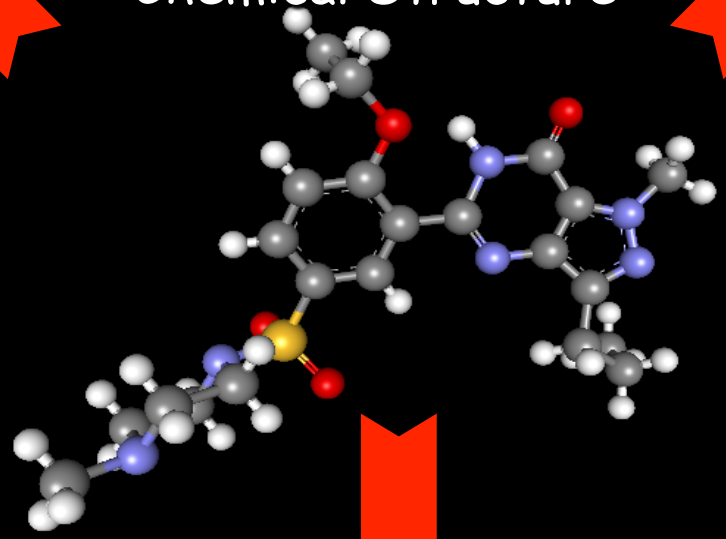
(Q)SAR
(Quantitative) Structure-Activity Relationships

PHYSICAL PROPERTIES

TOXICITY



Chemical Structure



ENVIRONMENTAL DISTRIBUTION

Mod da
A.Bassan, SIN

QSAR in U.S.

Dal 1979/80 ampio uso e "abuso"

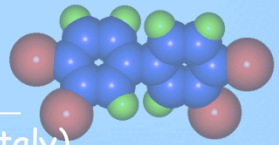
EPA / OPPT *Office of Pollution Prevention and Toxics*

TSCA: *Toxic Substances Control Act* inventory (~75.000 chem.)

NCP *New Chemicals Program (2000/yearPMN con dati QSAR)*

QSAR in E.U.

Uso molto limitato (quasi assente) prima del White Paper per il REACH (2001)



OECD (Organization for Economic Co-operation and Development)

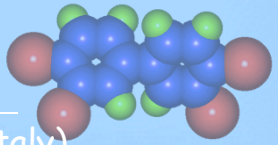
Programme for International Acceptability of QSARs in Regulation

Task Force of QSAR Experts

- Validation of existing QSAR models
- Guidance Documents – Application Tools

Workshop ICCA/CEFIC (2002): Setubal Principles

Workshops OECD (2003, 2004; 2006): OECD Principles



OECD Principles for the Validation, for Regulatory Purposes, of (Q)SAR Models

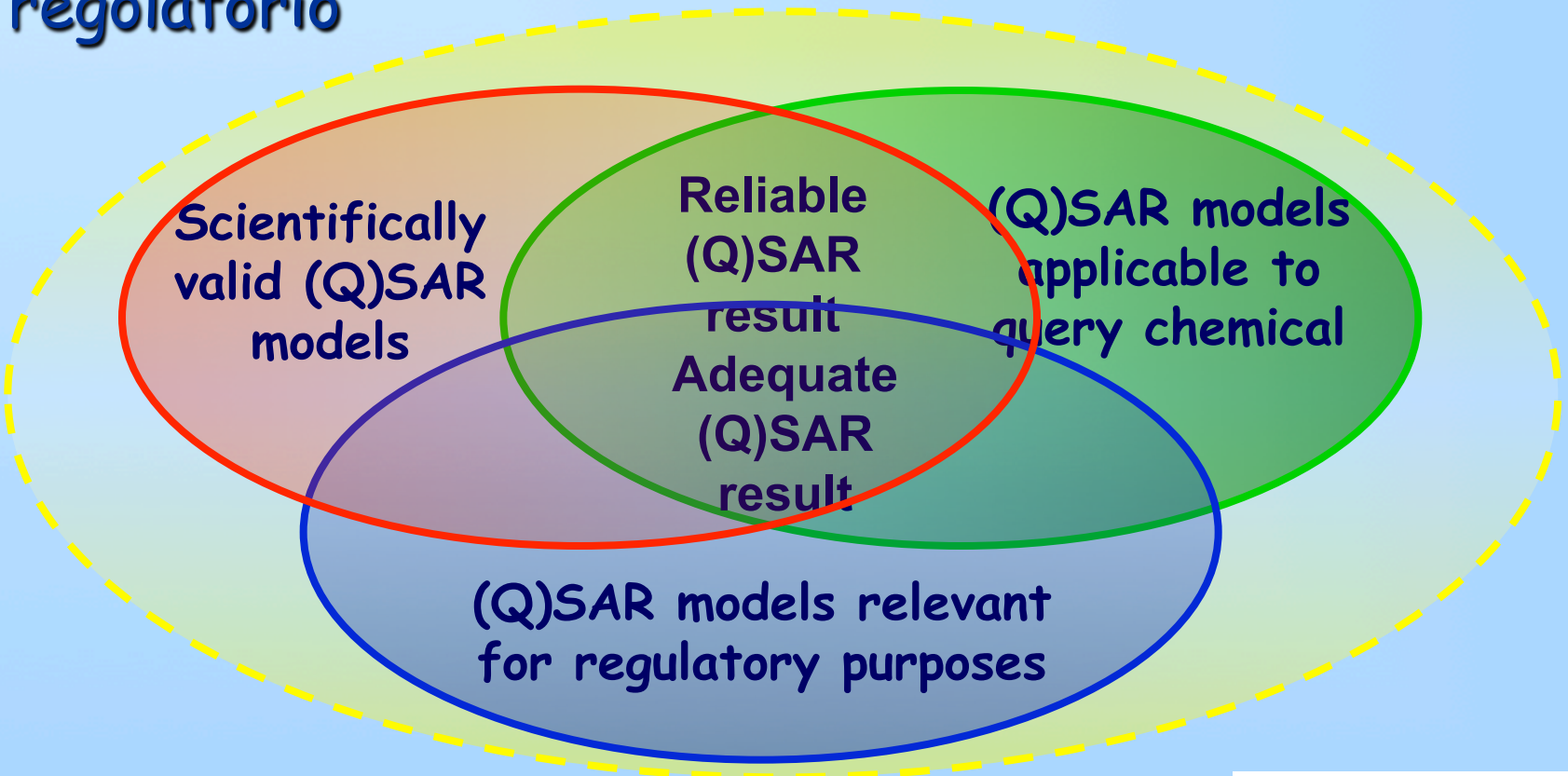
A QSAR model for regulatory purposes should be associated with the following information:

- 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;

OECD Guidance document: ENV/JM/MONO(2007)2

L' attendibilità e l' adeguatezza di una predizione QSAR dipendono da:

- Validità scientifica del modello
- Molecola nel dominio di applicabilità del modello
- Rilevanza della risposta modellata nell'ambito regolatorio



ECHA
European Chemicals Agency

ECHA
e i metodi
in silico

ECHA

Guidance on information requirements and chemical safety assessment

Chapter R.6: QSARs and grouping of chemicals

OECD: ENV/JM/MONO(2007)2

La guida riporta indicazioni su:

a) Come stabilire la validità di un modello (Q)SAR

b) Come definire l'adeguatezza di un risultato derivante da

(Q)SAR in ambito regolatorio

a) Come documentare e giustificare l'utilizzo di modelli

(Q)SAR in ambito regolatorio

a) Dove trovare informazioni sui modelli (Q)SAR

Quantitative Structure-Activity Relationship (QSAR)

" La struttura molecolare di un composto chimico influenza le sue proprietà ed attività biologiche "

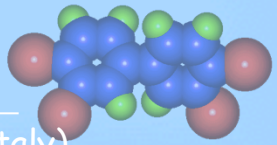
(Hansch 1964)

"Composti simili si comportano in modo simile"

Attività o Proprietà = f (Struttura)

E' possibile trovare una relazione matematica (f) tra Struttura Molecolare e comportamento (Attività o Proprietà) di un composto chimico

DATI PREDETTI



SAR

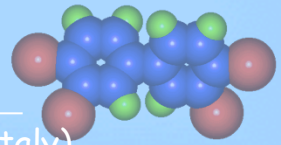
Identificazione di frammenti strutturali o proprietà (alerts) correlati ad una particolare attività (o proprietà)

- Analisi di Similarità
-
- Sistemi Esperti

Approccio Qualitativo
Attribuzione di Classe

Es. MultiCASE seleziona i descrittori (frammenti strutturali) automaticamente da un "learning set" di molecole attive/inattive.

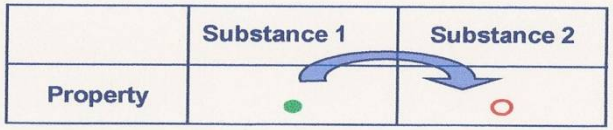
- Descrittori Attivanti (biofori)
- Descrittori Inattivanti (biofobi)



Strategia READ-ACROSS basata sulla similarità strutturale

The analogue approach refers to the grouping of chemicals and application of read-across for a given endpoint based on a relatively small number of analogues

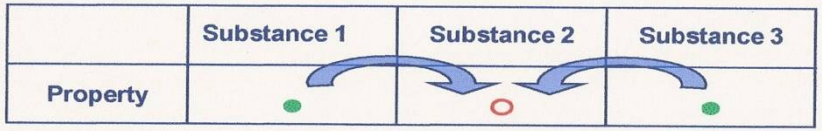
one-to-one



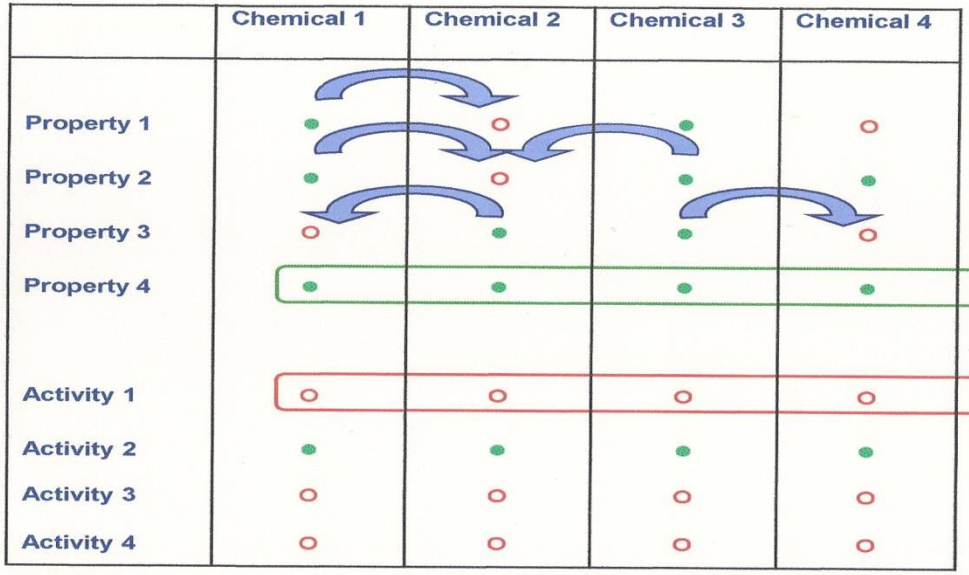
reliable data point

missing data point

many-to-one



JRC
EUROPEAN COMMISSION



read-across

QSAR
Trend analysis

reliable data point

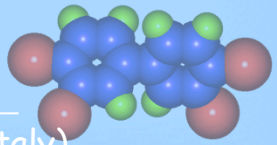
missing data point

The category approach refers to a wider approach, based on more analogues, multiple endpoints, and in which trends are also apparent

Gli studi QSAR, che utilizzano il maggior numero possibile di dati noti di input, sono basati sull'applicazione di diversi metodi matematici e statistici (metodi chemiometrici)

con lo scopo di trovare modelli quantitativi predittivi nelle relazioni tra struttura molecolare (descrittori molecolari: X_n) e attività biologica o proprietà (Y)

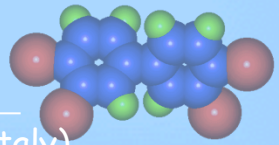
$$Y = b_0 + b_1X_1 + b_2X_2 + \ldots\ldots b_nX_n$$



Gli studi QSAR organizzano

la conoscenza esistente riguardo ad un "endpoint" (effetto o proprietà) con lo scopo di generalizzare questa conoscenza e permettere di fare

predizioni per altri composti chimici per i quali non si hanno dati sperimentali.



Storia del (Q)SAR

Alcani : m.p. and b.p._ ↑ n.C ↑ M.W.

SAR

Alcooli: solubilità in acqua ↓ n.C ↑ M.W.

PROPRIETA' CHIMICO-FISICHE ⟺ STRUTTURA

Alcooli tossicità ↑ SAR coeff.ripart. olio/acqua ↑ Log P

(Meyer-Overton 1899-1901)

ATTIVITA' BIOLOGICA ⟺ PROPRIETA' CHIMICO-FISICHE

ATTIVITA' BIOLOGICA ⟺ QSAR ⟺ STRUTTURA/ PROPRIETA'

(Hansch e Fujita 1964)

Esempio di modello QSAR

"Tossicità" = a + b logP + c E + d S

(Modello di Hansch)

logP

o log Kow, coefficiente di ripartizione tra ottanolo e acqua: termine di lipofilia

E

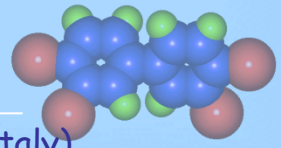
termine elettronico

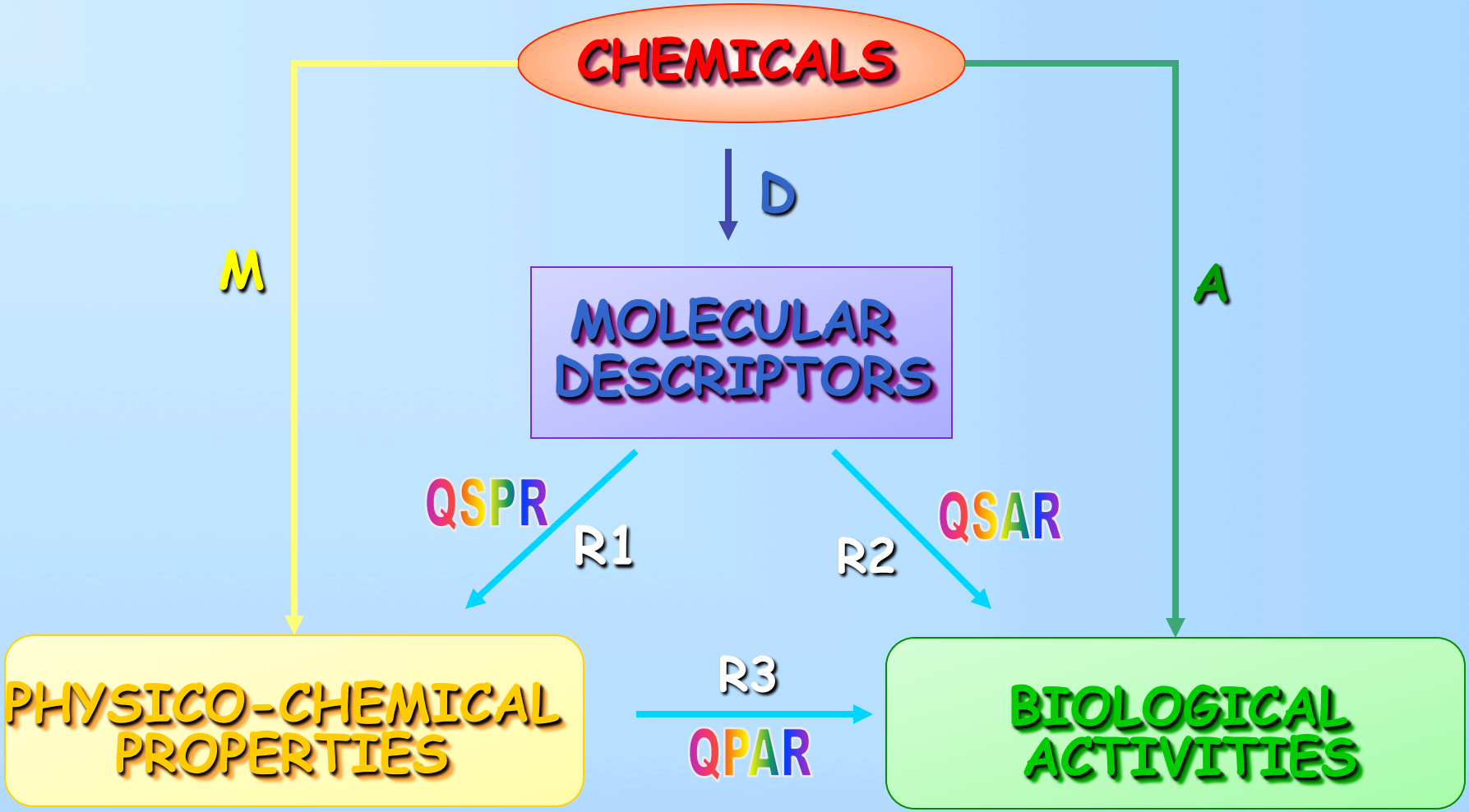
S

termine sterico

Probabilità o abilità del composto chimico di raggiungere il "target"

Possibilità del composto chimico di interagire con il " target" e di essere attivo





M: experimental measures of properties

A: experimental measures of activities

D: theoretical procedures for descriptors

R1, R2, R3: mathematical relationships

Le 3 necessità per il QSAR:

- **BUONI DATI di INPUT**

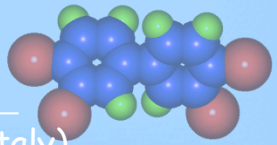
  Dati sperimentali di buona qualità sui quali trovare la Relazione Struttura-Attività

- **INFORMAZIONE STRUTTURALE**

  Buona rappresentazione della struttura chimica: descrittori molecolari

- **MODELLI PREDITTIVI**

  Modelli quantitativi con caratteristiche validate di predittività



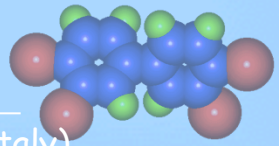
I Dati Sperimentali

Serve un numero "limitato" di **dati sperimentali *di ALTA QUALITA'*** sui quali sviluppare i modelli QSAR

- PIU' NUMEROSI POSSIBILE
- CORRETTI
- OMOGENEI (stesso lab, stesso metodo)

La qualità dei modelli teorici non può essere migliore dei dati di input

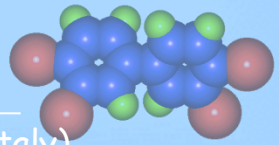
Garbage in, garbage out !!



ENVIRONMENTAL PARAMETERS

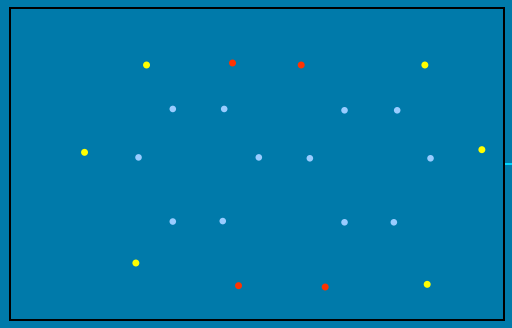
Priority setting / Risk Assessment

Parameters	Quality of QSAR models
Physico-chemical data m.p.; b.p.; vapour pressure; Henry law constant; water solubility; partition coefficients:Kow, Koc,..	OPTIMUM
Environmental fate and pathways chemical-, photo- and bio-degradation; bioaccumulation; half-lives	HIGH
Ecotoxicity algae; *Daphnia*; fish; ….	MEDIUM-HIGH
Mammalian toxicity skin-, eyes-, oral-, inhalation acute toxicity; mutagenicity; carcinogenicity; toxicity to reprod.	MEDIUM-LOW



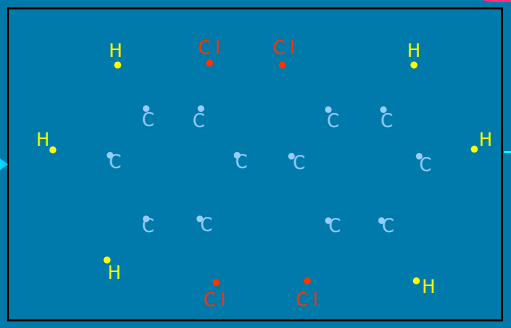
DESCRITTORI MOLECOLARI

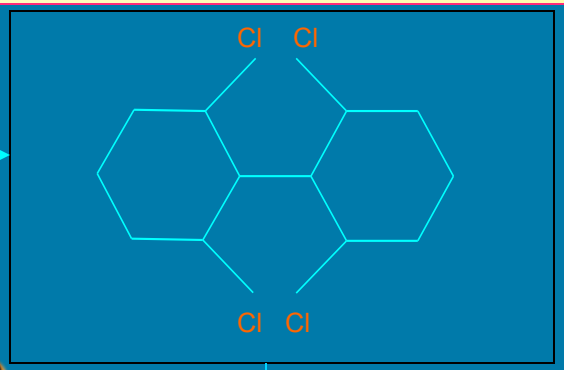
Numeri che "rappresentano" un composto chimico



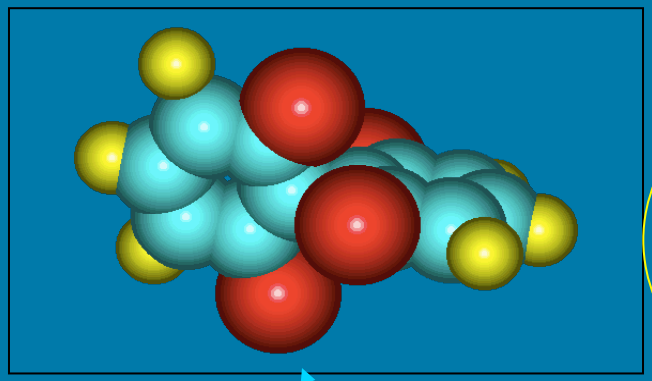
0D

1D

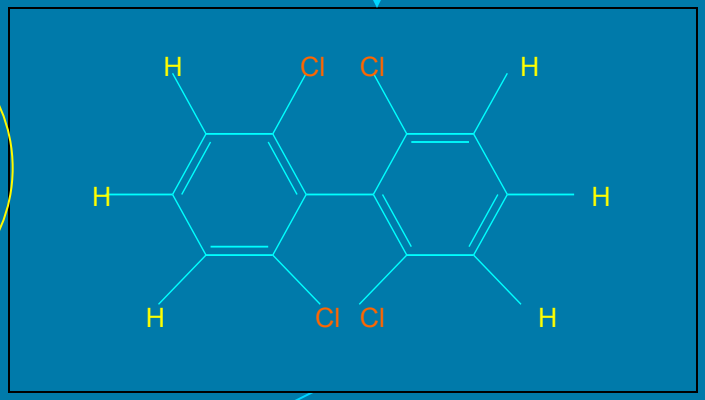


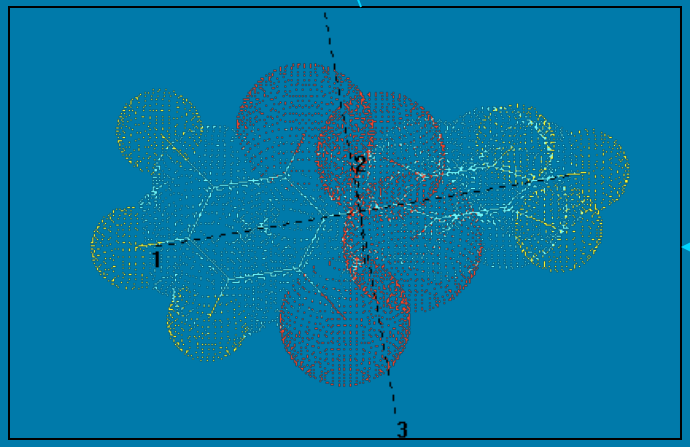


2D

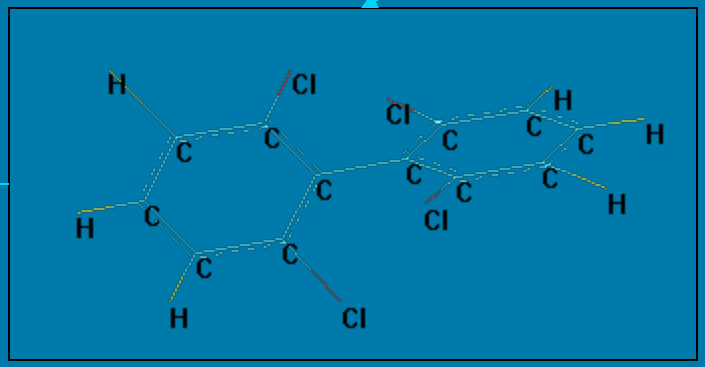


Global View



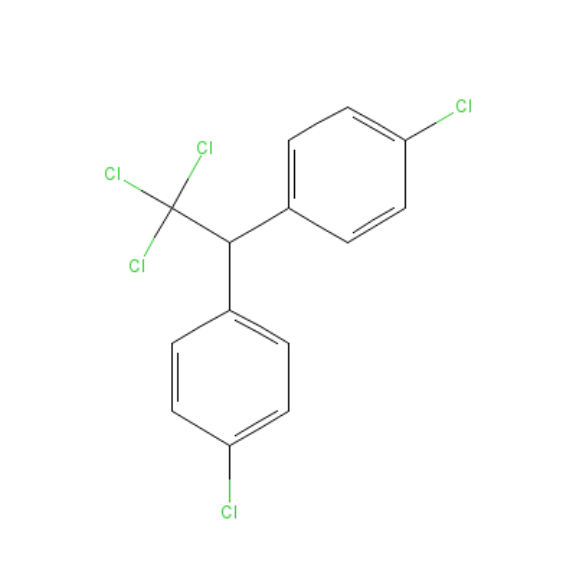


3D



INFORMAZIONE STRUTTURALE

I Descrittori Molecolari, calcolabili con vari software, codificano l'informazione strutturale e la convertono in numeri



CAS	AMW	Sv	Ss	Mv	Me	Ms
000050-29-3	12.66	21.69	45.81	0.77	1.03	2.41
000050-30-6	12.73	11.22	33.89	0.75	1.06	3.08
000050-31-7	15.03	11.92	37.67	0.79	1.09	3.14
000050-32-8	7.89	23.59	37.33	0.74	0.98	1.87
000051-28-5	10.83	11.14	49	0.66	1.1	3.77
000051-44-5	12.73	11.22	33.89	0.75	1.06	3.08
000055-38-9	8.98	19.38	35.81	0.63	1.01	2.24
000055-63-0	10.77	11.67	60.83	0.56	1.14	4.06
000056-23-5	30.76	5	17.69	1	1.21	3.54
000056-38-2	9.1	19.71	49.31	0.62	1.03	2.74
000057-15-8	11.83	9.6	24.83	0.64	1.05	3.1
000057-74-9	17.07	19.79	46.81	0.82	1.07	2.6
000058-89-9	16.16	13.79	32.67	0.77	1.07	2.72
000058-90-2	17.84	11.11	32.78	0.85	1.1	2.98
000059-50-7	8.91	10.6	23.11	0.66	1.01	2.57
000060-29-7	4.94	7.5	10.5	0.5	0.98	2.1
000060-51-5	9.55	14.17	31.97	0.59	1.02	2.66

Questi numeri sono le variabili utilizzate per costruire i modelli QSAR, trovando quali sono correlate con la risposta da modellare.

DESCRITTORI MOLECOLARI

DESCRIZIONE ESAUSTIVA

INFORMAZIONE

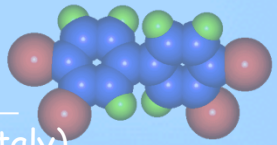
Modello con sola informazione rilevante

SELEZIONE

Risposta Y

$Y = f$ (descrittori selezionati)

MODELLO QSAR Predittivo



CHEMICALS

MOLECULAR DESCRIPTORS

EXPERIMENTAL DATA

x_1	x_2	...	x_n	Y
TRAINING SET				

EXPLORATIVE ANALYSIS:
- *Principal Component Analysis*
- *Cluster Analysis*

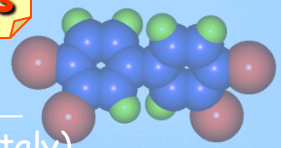
Chemometric Methods

CLASSIFICATION METHODS:
- *Classification Tree (CART)*
- *Discriminant Analysis*
- *Neural Networks*

REGRESSION METHODS:
- *Multivariate Linear Regression (MLR)*
- *Partial Least Squares Regression (PLS)*

Quantitative models for qualitative responses

Quantitative models for quantitative responses

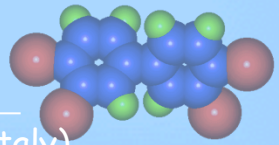


NON SOLO *FITTING* ……

Un modello che semplicemente spieghi ciò che in un sistema (molecola) è già conosciuto (dati noti) (approccio meccanicistico) non è sufficiente per garantire predizioni attendibili per nuove molecole!

E' essenziale costruire modelli validati per le loro capacità di predizione di eventi futuri (nuove molecole, anche non ancora sintetizzate: "chemical design")

VERIFICA della PREDITTIVITA' !!!
Approccio Statistico



INSUBRIA contributions

The Importance of Being Earnest: Validation is the Absolute Essential for Successful Application and Interpretation of QSPR Models

A. Tropsha, P. Gramatica, V.K. Gombar *QSAR & Comb.Sci.*, 2003: > 900 cit.

…"*in spite of their high fitted accuracy and apparent mechanistic appeal, some published QSPR models fail rigorous validation tests, and, thus, may lack practical utility as reliable screening tools.* "

Principles of QSAR models validation: internal and external

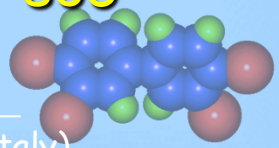
P. Gramatica *QSAR & Comb.Sci.*, 2007. *26*, 694-701. > 450 citazioni

Real External Predictivity of QSAR Models: Part 1 and 2

N. Chirico and P. Gramatica, *J. Chem. Inf. Model.*, 2011, 2012.

QSAR Modelling is not « Push a button and find a correlation » : …

P. Gramatica et al. *Molecular Informatics*, 2012, *31*, 817 – 835



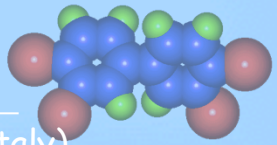
DOMINIO DI APPLICABILITA' (AD) ???

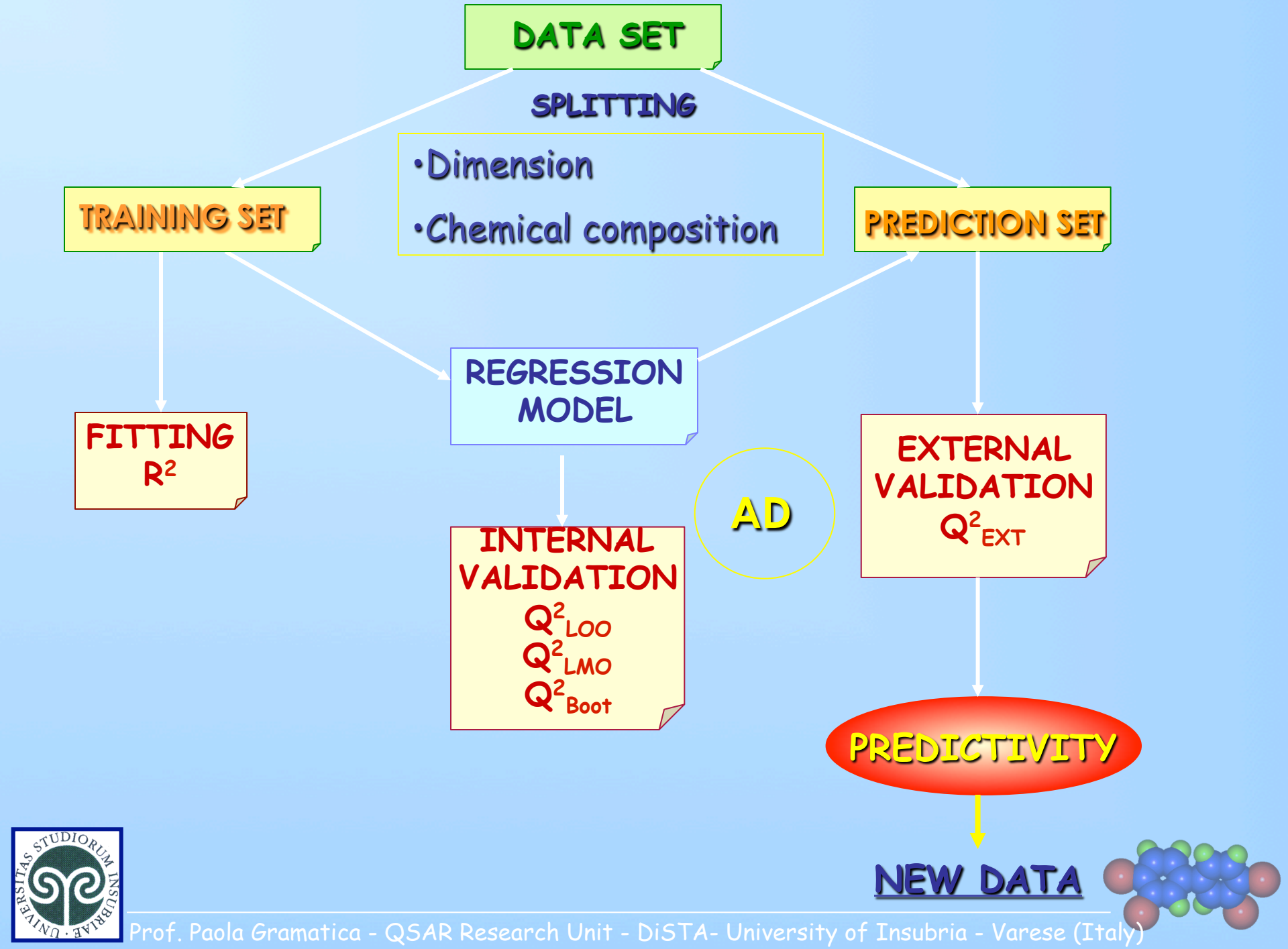
Non si può applicare qualsiasi modello a qualsiasi molecola!!!!

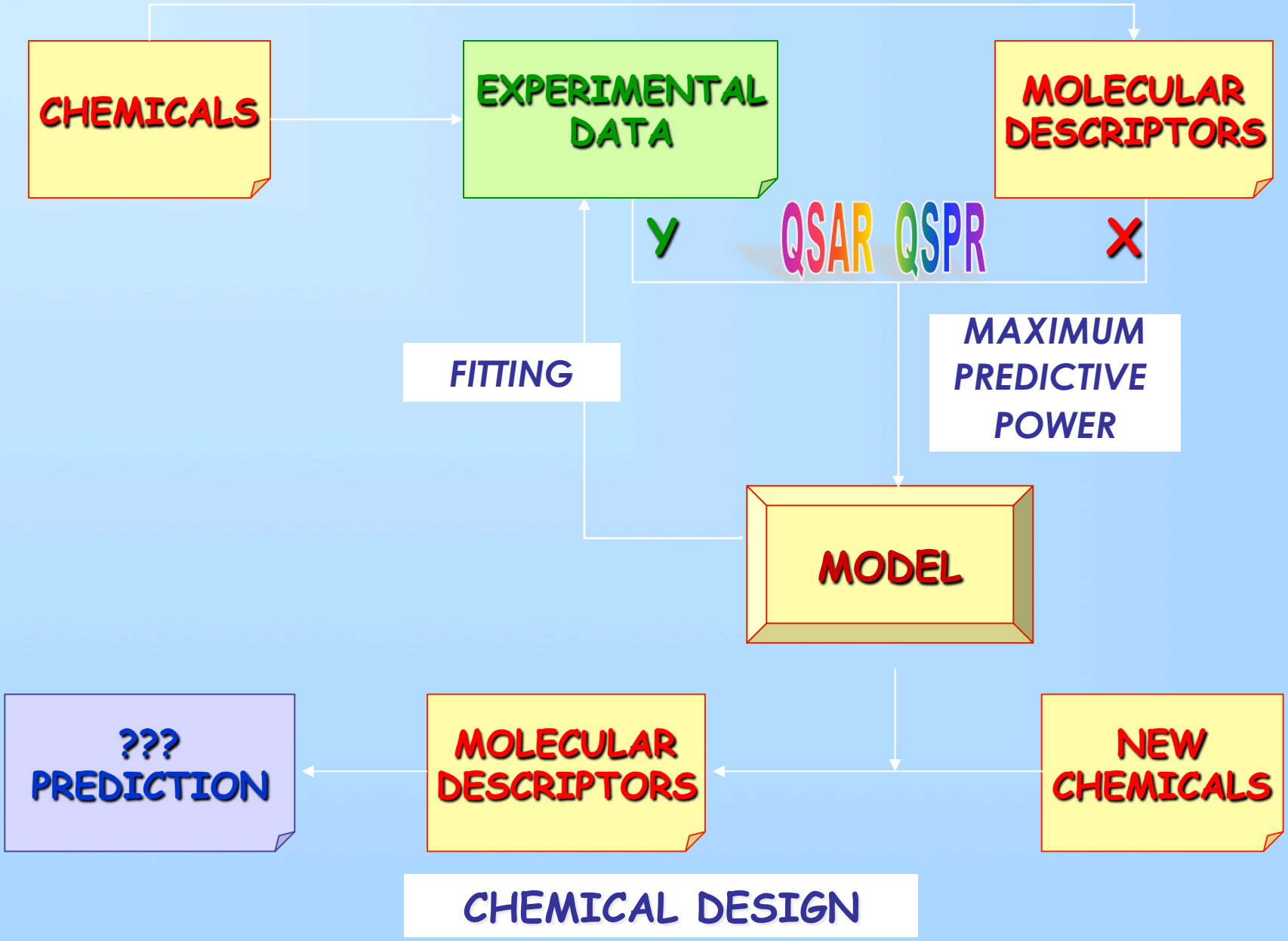
- **È definito dalla copertura strutturale/di risposta delle molecole presenti nel training set (informazione di input)**

- **Mediante apposite tecniche statistiche è possibile valutare se un nuovo composto è incluso in questo spazio**

Predizioni più attendibili solo per le molecole all'interno del dominio (interpolazioni)




DATA SET
SPLITTING
•Dimension
•Chemical composition
TRAINING SET
PREDICTION SET
FITTING
R^2
REGRESSION MODEL
AD
INTERNAL VALIDATION
Q^2_{LOO}
Q^2_{LMO}
Q^2_{Boot}
EXTERNAL VALIDATION
Q^2_{EXT}
PREDICTIVITY
NEW DATA


CHEMICALS
EXPERIMENTAL DATA
MOLECULAR DESCRIPTORS
Y
QSAR QSPR
X
FITTING
MAXIMUM PREDICTIVE POWER
MODEL
??? PREDICTION
MOLECULAR DESCRIPTORS
NEW CHEMICALS
CHEMICAL DESIGN

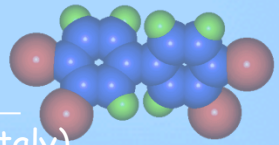
APPLICAZIONI del QSAR

- Predire dati mancanti (REACH)
- Evidenziare i prodotti chimici "pericolosi", anche prima della loro sintesi (HAZARD ASSESSMENT)
- Progettare composti alternativi più sicuri (CHEMICAL DESIGN in GREEN CHEMISTRY)

LISTE di PRIORITA'

Ottimizzare le risorse industriali

Ridurre i tests sugli animali

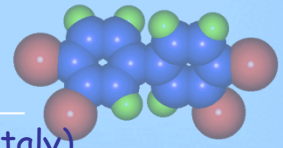


PBT : Persistent, Bioaccumulative and Toxic

Persistenza → Resistenza di una sostanza a degradarsi nell'ambiente

Bioaccumulo → Aumento della concentrazione di un inquinante dall'ambiente ad un organismo vivente mediante tutte le vie possibili (cibo, respirazione, contatto, ecc)

Tossicità → Potenziale di una sostanza di esercitare un effetto pericoloso su uomo o animali (acuta o cronica)



Analisi delle Componenti Principali (PCA)

La PCA è una analisi esplorativa di dati multivariati.

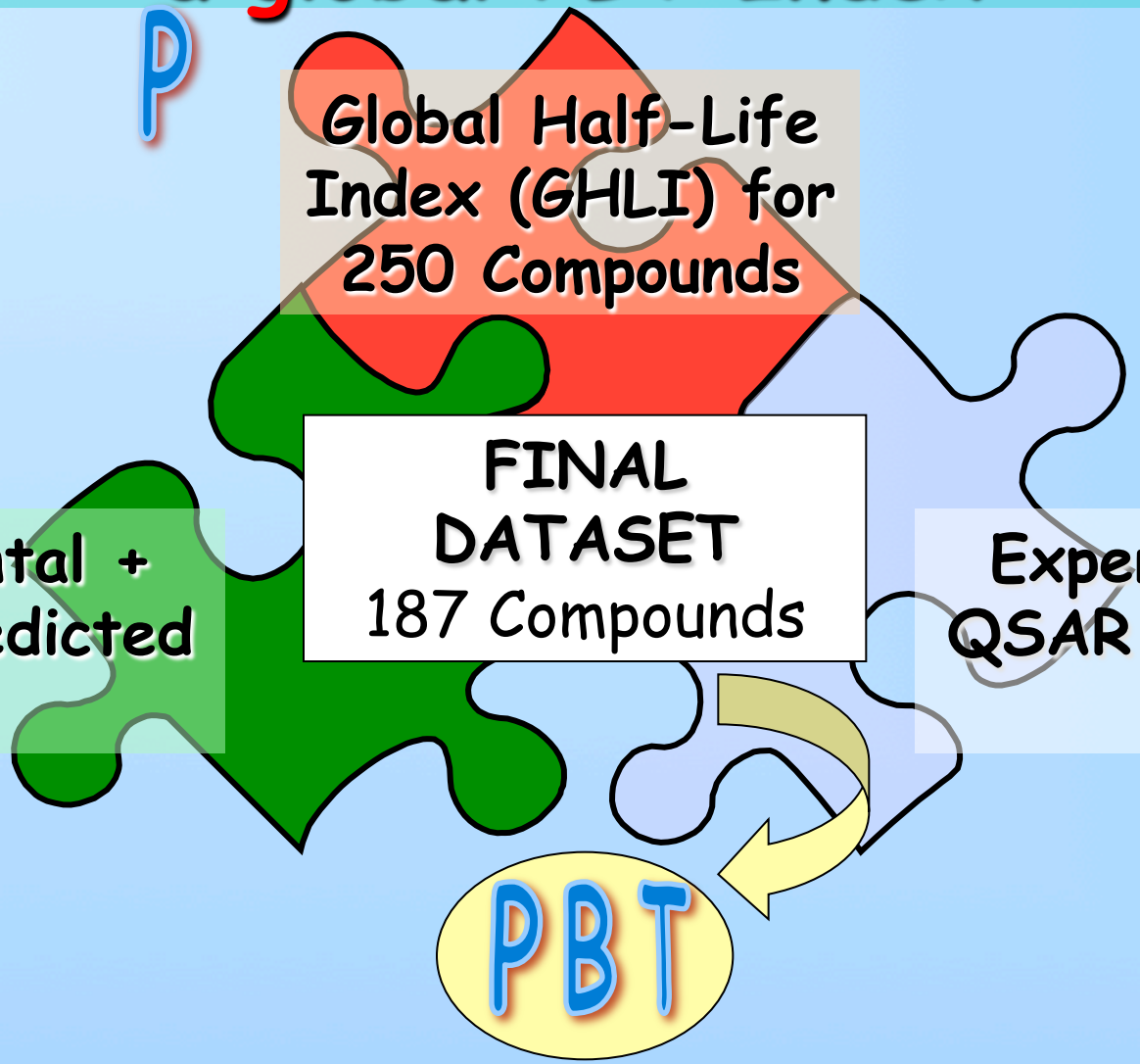
Le Componenti principali sono:

- Combinazioni lineari dei dati originali
- Ordinate secondo le direzioni di massima varianza:PC1, PC2..
- Ortogonali quindi NON correlate
- Sono quindi nuove variabili con le quali si condensa e "pulisce" l'informazione contenuta nei dati originali
- Rappresentano macroproprietà dell'insieme dei dati originali

I dati vengono così "visti" in un diverso sistema di riferimento

Secondo visuali (le diverse PC) controllate per qualità e quantità dell'informazione rappresentata

Rational for a cumulative PBT behaviour: a global PBT Index



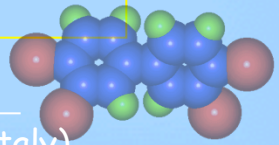
DATA SET for global persistence: screening of POPs

Literature* overall half-life data for degradation
in air, water, sediment and soil
combined by multivariate statistical analysis (PCA)
and correlated to theoretical molecular structure
descriptors by QSAR.

250 ORGANIC COMPOUNDS

pesticides , dioxins, PAHs , PCBs
heterogeneous industrial chemicals

Prediction of the environmental persistence of an **heterogeneous set of organic** chemicals

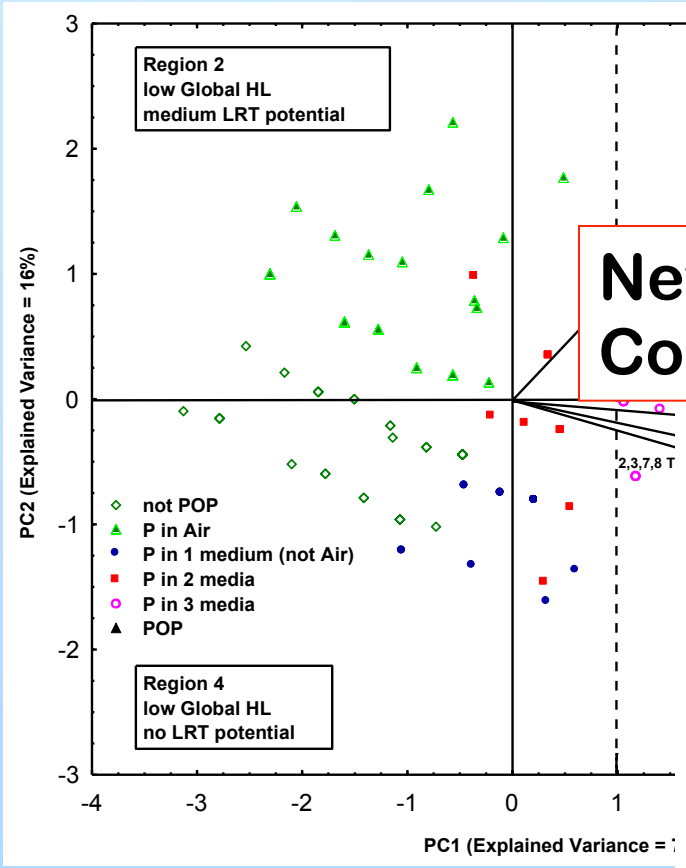


Global Half Life Index - GHLI

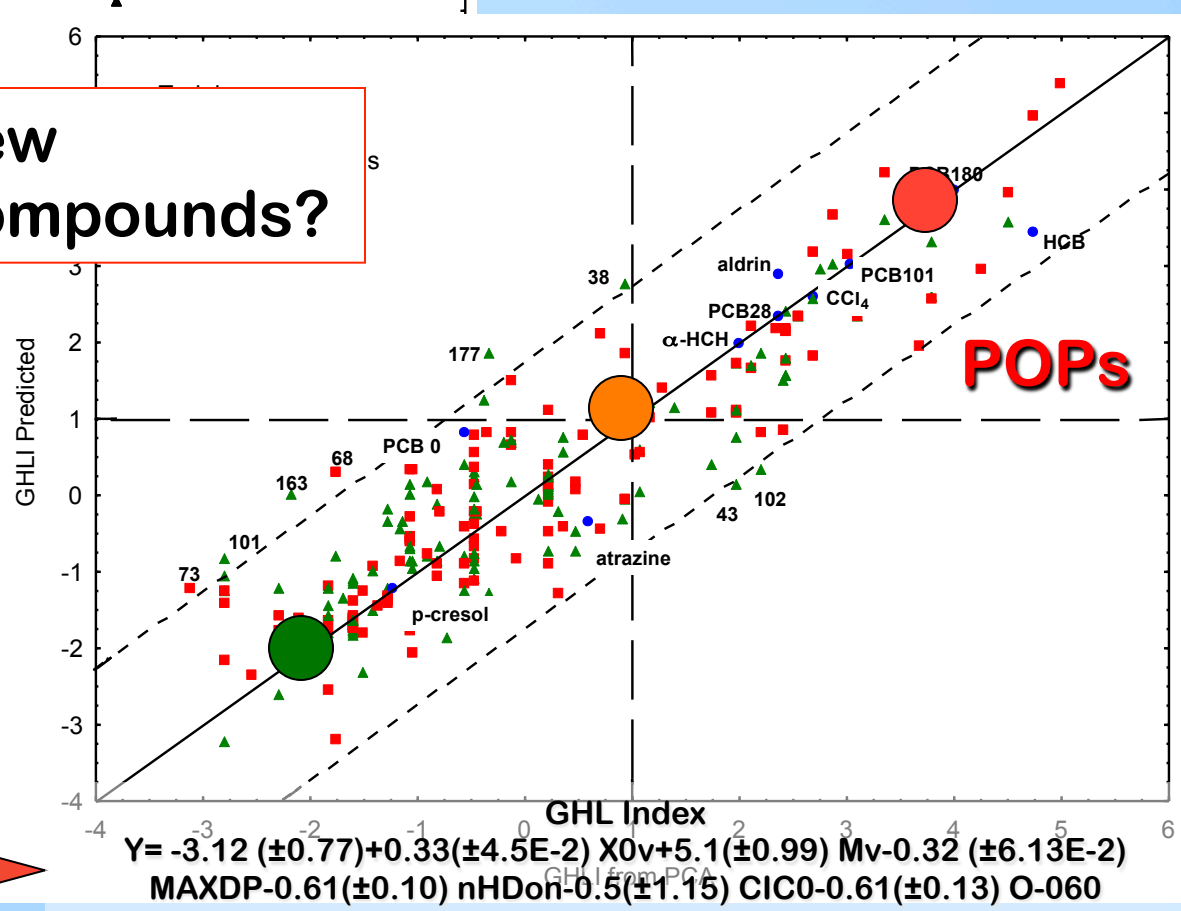
P.Gramatica and E.Papa, *Environ. Sci. Technol. 25*, 755, 2007.

$n_{tr.} = 125$, $R^2 = 0.85$, $Q^2_{LOO} = 0.83$, RMSE=0.76

$n_{pr.} = 125$; $Q^2_{EXT} = 0.79$; RMSEP=0.78



New Compounds?

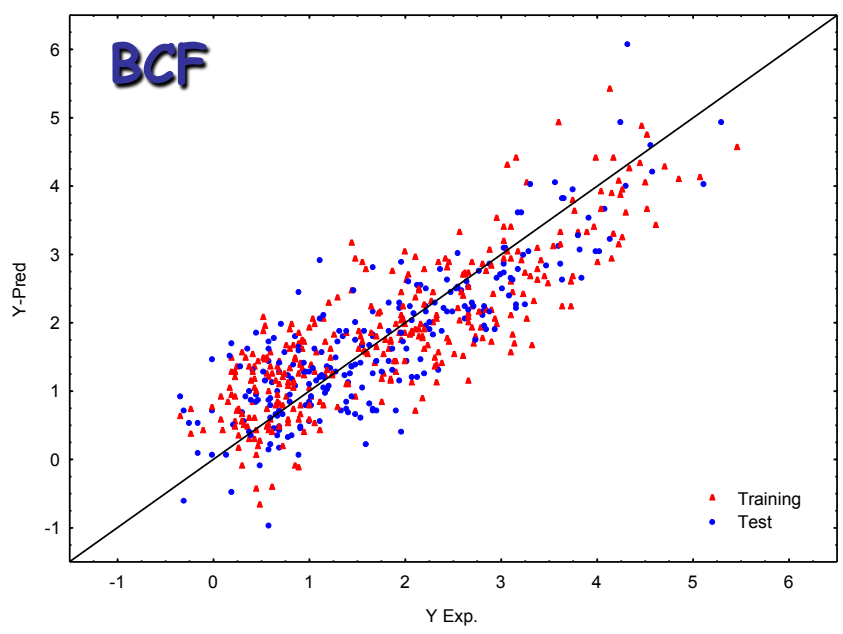


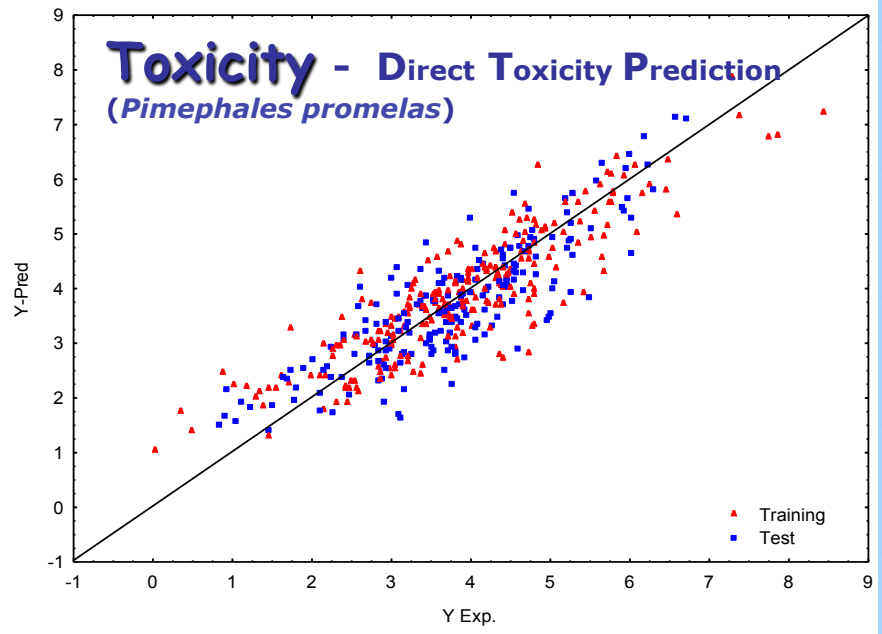
QSAR MODEL

Y= -3.12 (±0.77)+0.33(±4.5E-2) X0v+5.1(±0.99) Mv-0.32 (±6.13E-2) MAXDP-0.61(±0.10) nHDon-0.5(±1.15) CIC0-0.61(±0.13) O-060

GHLI from PCA

Bioconcentration and Acute Fish Toxicity Models



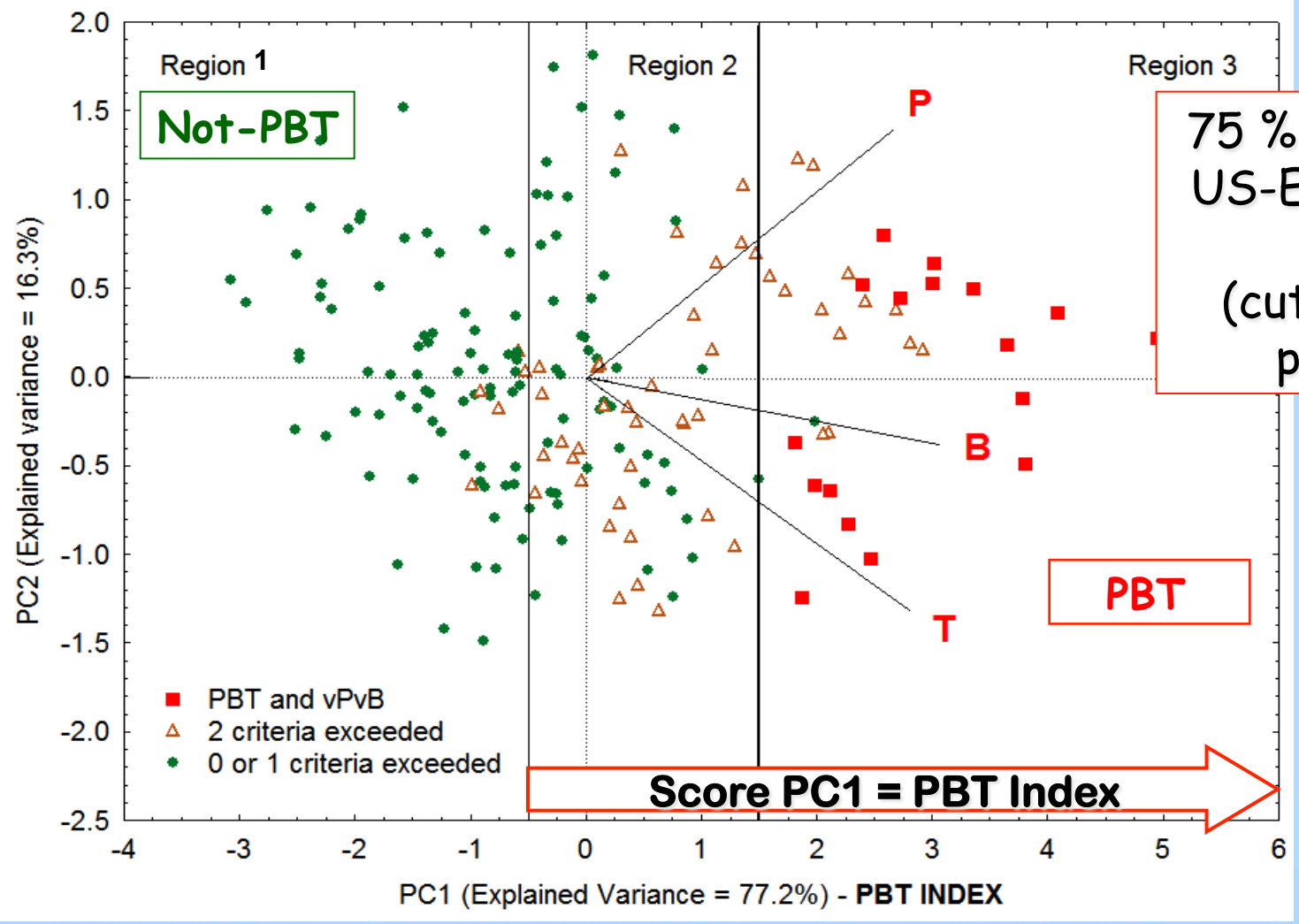


P. Gramatica and E.Papa
QSAR & Comb.Sci.
22, 374, 2003.
24, 953, 2005.
E.Papa et al.
Chemosphere, *67*, 351, 2007.

E.Papa et al.
JCIM, *45*, 1256, 2005.

Models' internal / external predictivity and robustness successfully verified!

Definition of the cumulative PBT Index by PCA: Not Separate P, B, and T criteria



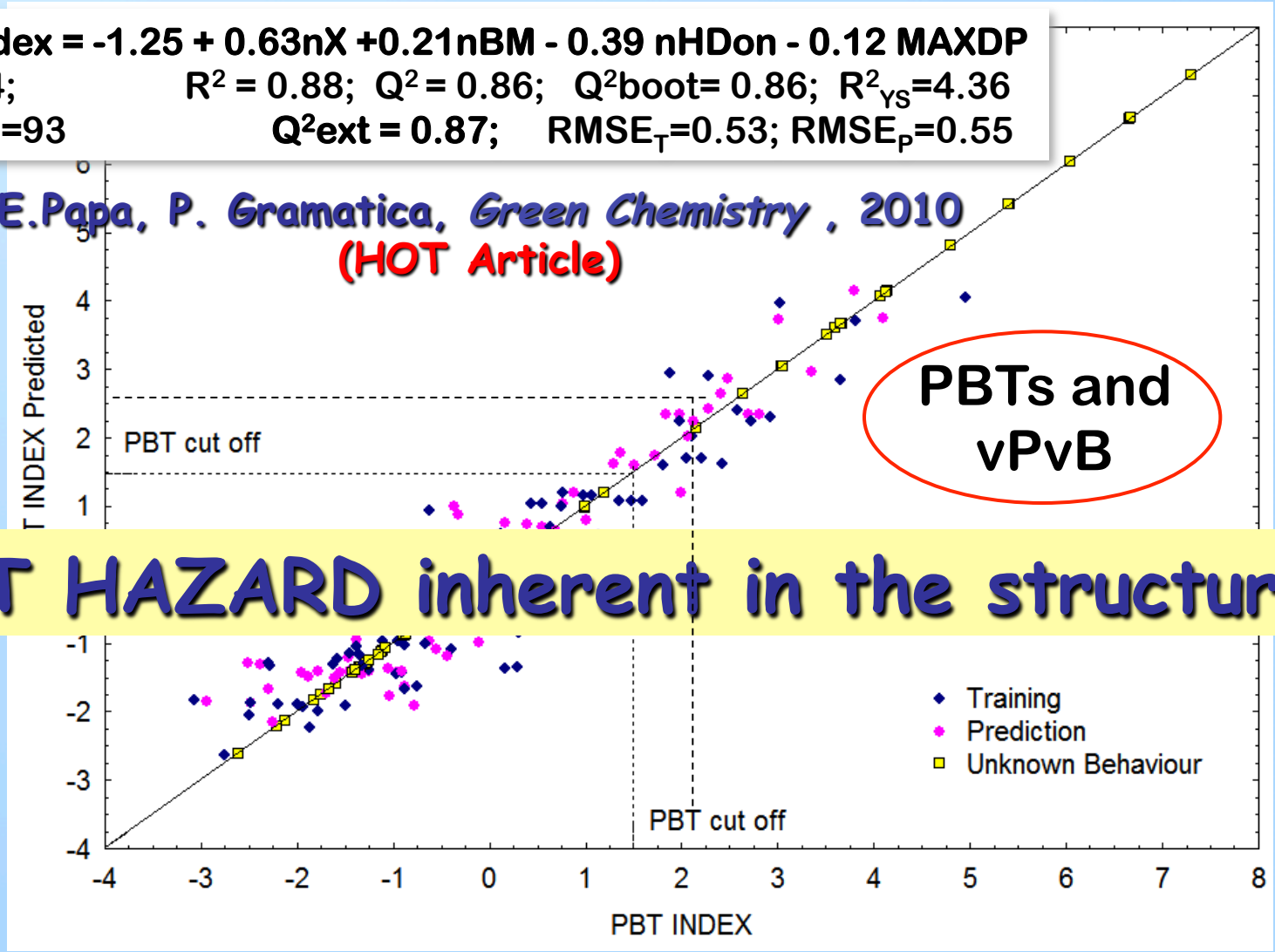
75 % agreement with US-EPA PBT profiler. Our Index (cut-off: 1.5) more precautionary

The cumulative PBT Index was modelled by Molecular Descriptors

QSPR model of PBT Index: Screening of PBTs

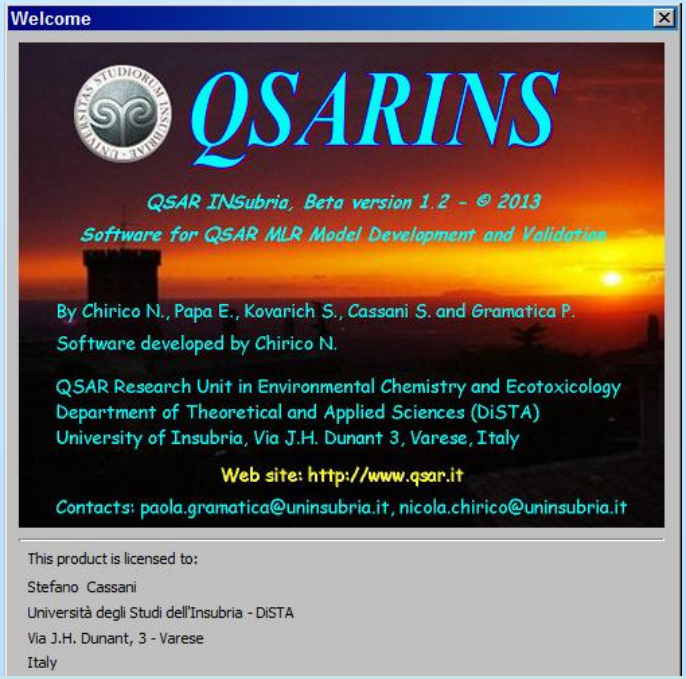
PBT Index = -1.25 + 0.63nX +0.21nBM - 0.39 nHDon - 0.12 MAXDP
ntr=94; R^2 = 0.88; Q^2 = 0.86; Q^2boot= 0.86; R^2_{YS}=4.36
npred=93 Q^2ext = 0.87; $RMSE_T$=0.53; $RMSE_P$=0.55

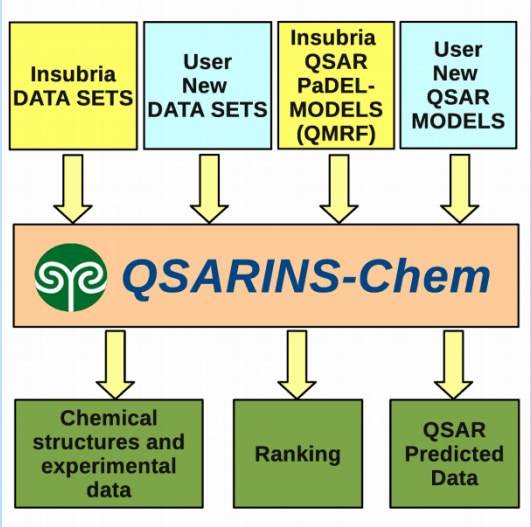
E.Papa, P. Gramatica, *Green Chemistry* , 2010
(HOT Article)



PBT HAZARD inherent in the structure

QSARINS: a new software for QSAR MLR model development and validation





http://www.qsar.it

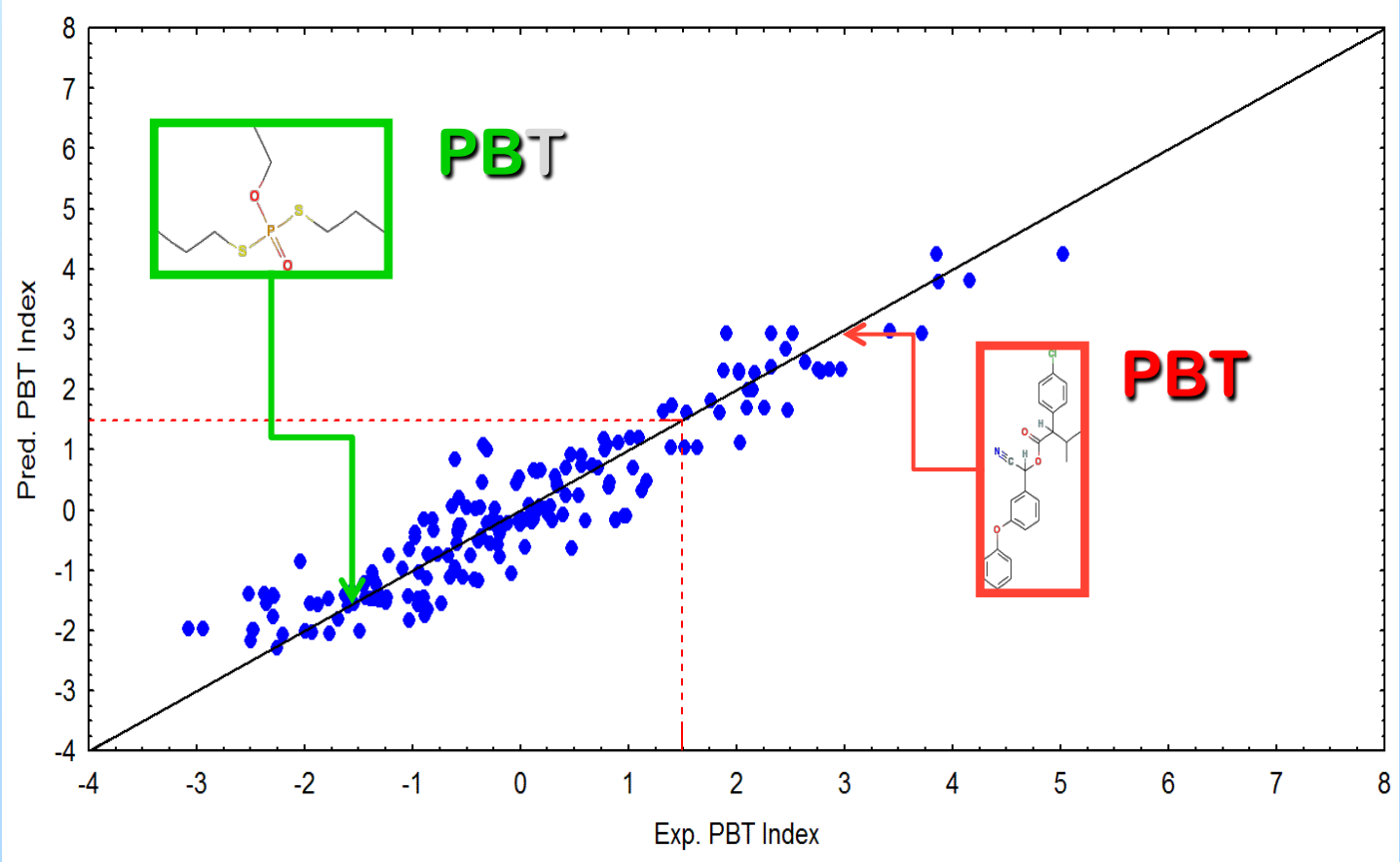
P.Gramatica et al.
J.Comput. Chem.,
2013, *34*, 2121
2014, 35, 1036

The PBT Index model is implemented:

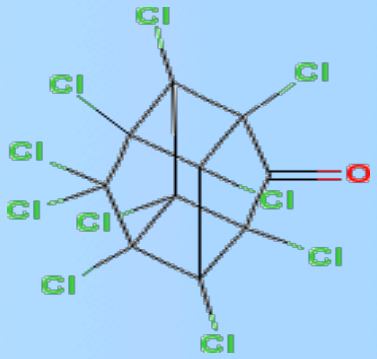
PBT Index=-1.46 +0.64 nX +0.22 nBondsM -0.39 nHBDon_Lip.-0.06 MAXDP2

- **QSARINS predicts the PBT Index score, with AD.**

Screening of alternative chemicals by PBT Index QSARINS model



Chlordecone



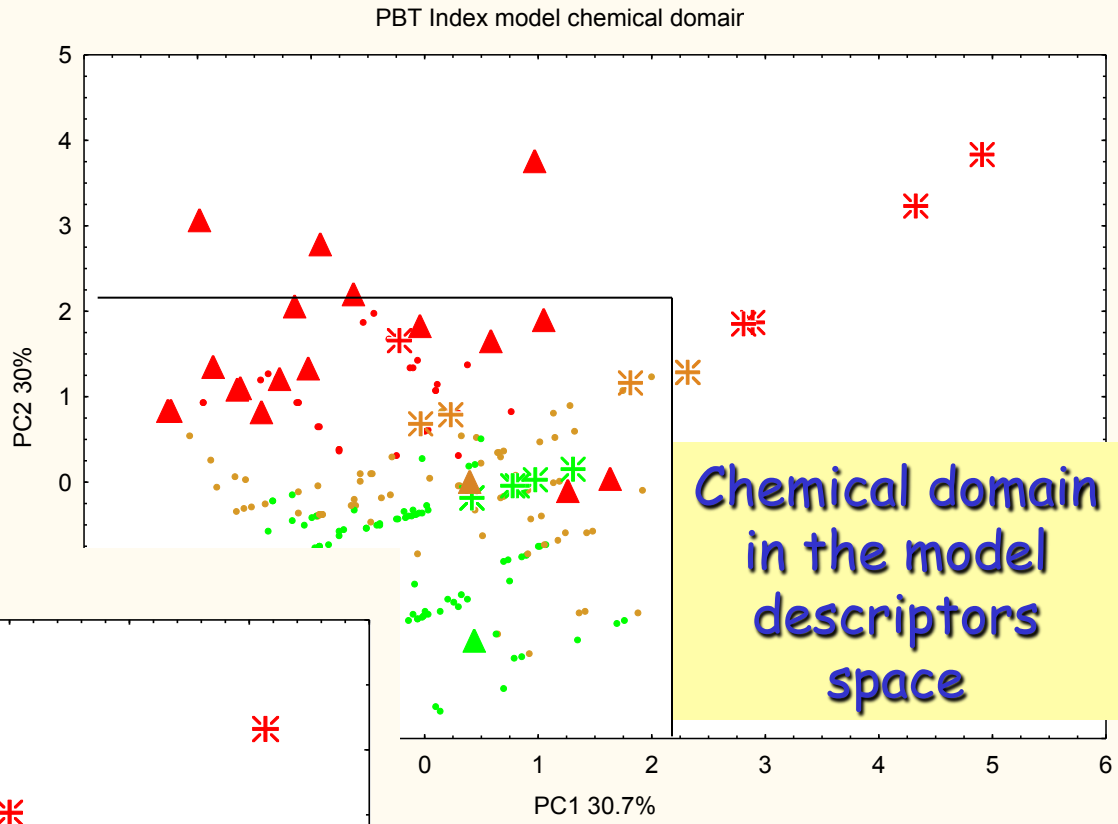
Alternatives:
Ethoprop
Esfenvalerate

E.Papa EU-SETAC Seville 2010, US-SETAC Boston 2011

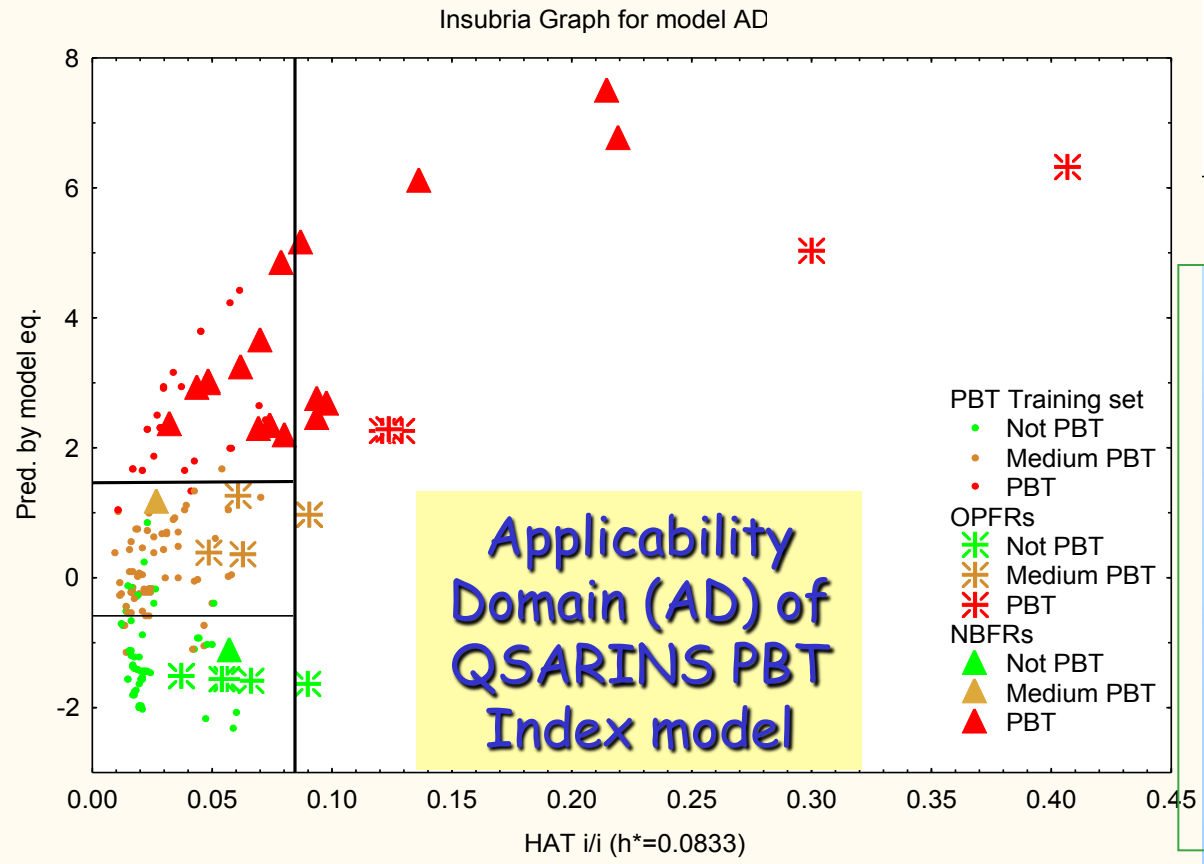
STUDYING ALTERNATIVES TO BROMINATED FLAME RETARDANTS (BFRs)

P. Gramatica – EU-SETAC Glasgow 2013

Screening of New Brominated Flame Retardants (NBRFs) and OrganoPhosphorous FR (OPFRs) by PBT Index



Chemical domain in the model descriptors space



Applicability Domain (AD) of QSARINS PBT Index model

PBT Index highlights, from the design, that:

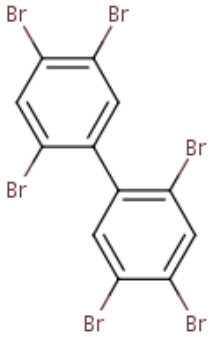
NBRFs are mainly PBTs

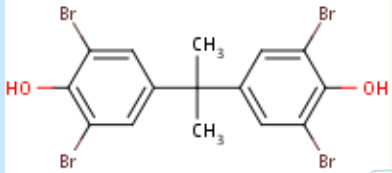
OPFRs are safer

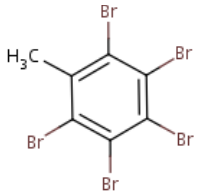
(but probably not all...)

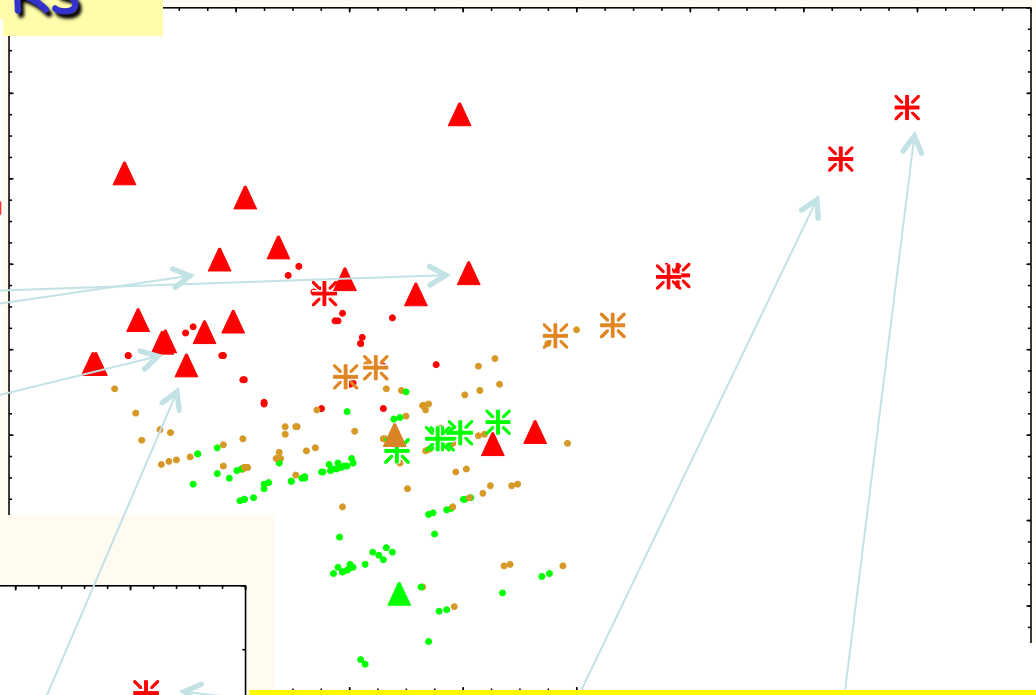
AD check

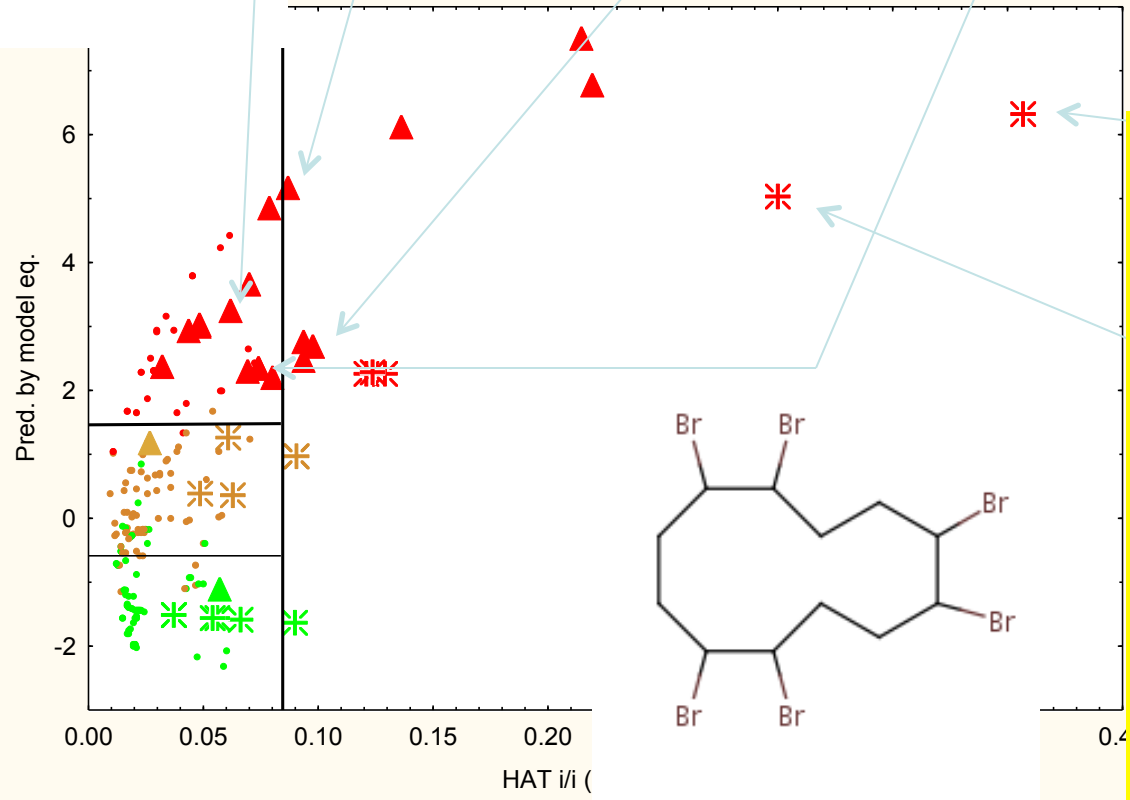
Screening of NBRFs and OPFRs











Some are extrapolations.

BUT…. Some FRs,

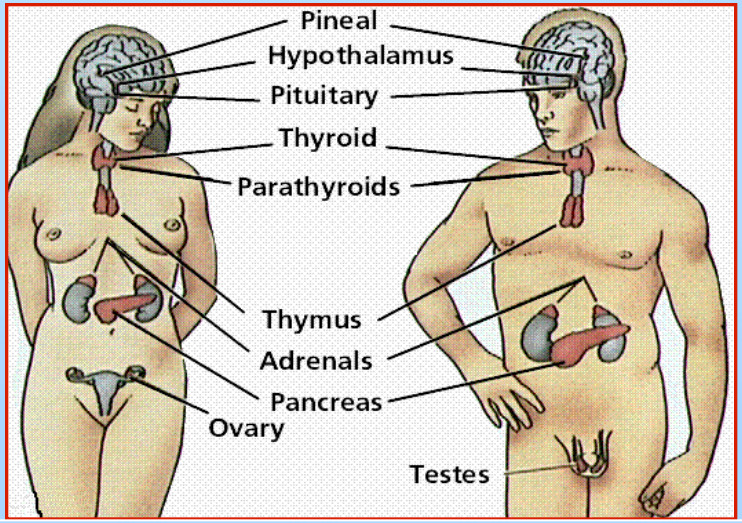
into the model AD,

now on the market as

"safer" alternatives,

are PBTs !!!!!

Endocrine disruptor chemicals (ED o EDC) in the Environment: predictive QSAR modeling



Some serious adverse effects of EDs

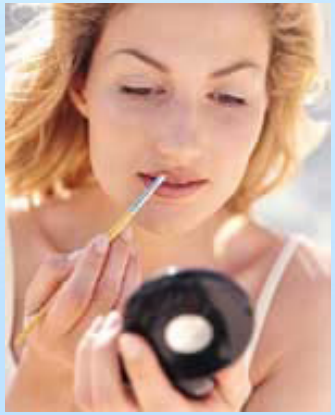
- accelerated puberty in females
- decline in sperm counts (male fertility decrease)
- increase in the incidence of breast cancer and endometriosis
- global increase in testicular cancer
- others.....
- decline and altered sex ratios in some wildlife populations

A large amount of potential EDs (> 57, 000)

Reliable short-term methods are needed

Source and distribution of EDs

Cosmetics



Pesticides

Pharmaceuticals

Natural

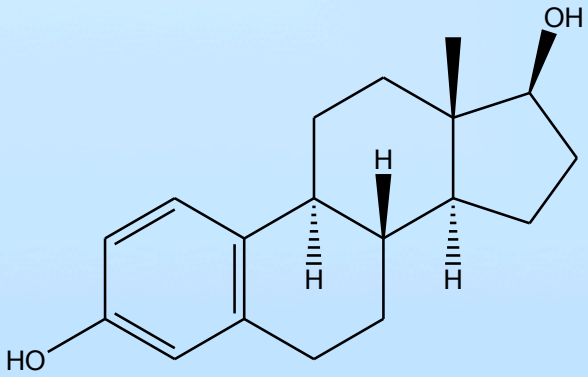
Products

Waste water

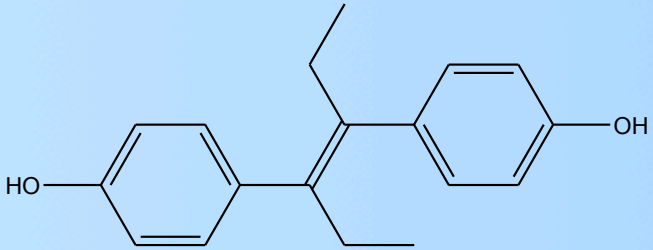
Environmental Pollutants

Plastics

Common Endocrine Disruptors

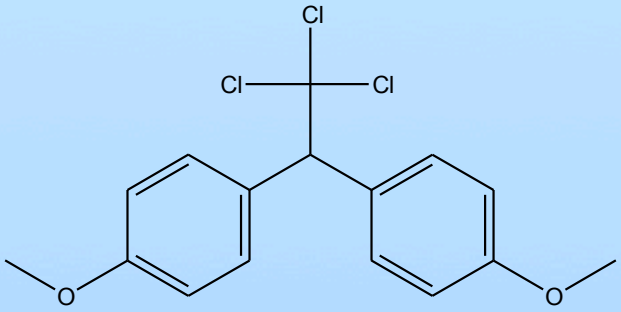


17b-Estradiol

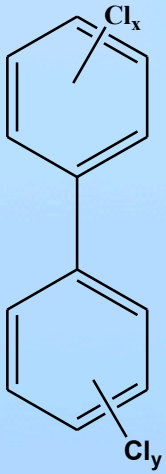


Diethylstilbestrol (DES)

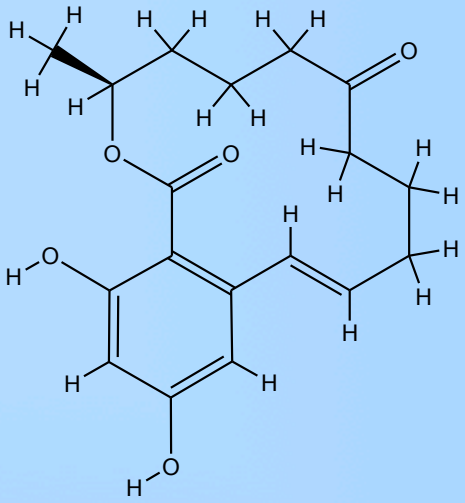
Pharmaceuticals



Methoxychlor

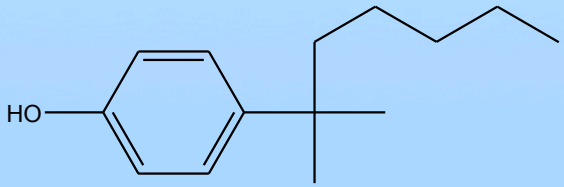


Polychlorinated Biphenyl (PCB)



Zearalenone

Industrial Chemicals



4-t-Octylphenol

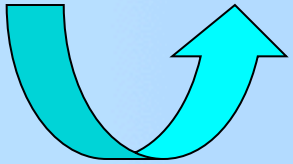
Environmental Pollutants

Natural Products

Methods for Chemical Testing and Assessment
ITS

"In vivo"

"In vitro"



"In silico"

expensive, laborious and time consuming

It is practically impossible to perform thorough toxicological tests on all the potential endocrine disruptors that may need to be evaluated.

Best QSAR Classification models

N Desc.	Descriptor Set	k	% Accuracy
1	nArOH	6	79.7%
2	TIC3, nArOH	3	84.5%
3	Qindex, EEig02r, nArOH	3	87.9%
4	T(Cl, Cl), EEig02r, nArOH, nR=Ct	3	89.2%

Aromatic hydroxyl groups most relevant structural descriptor to discriminate potential ER binders.

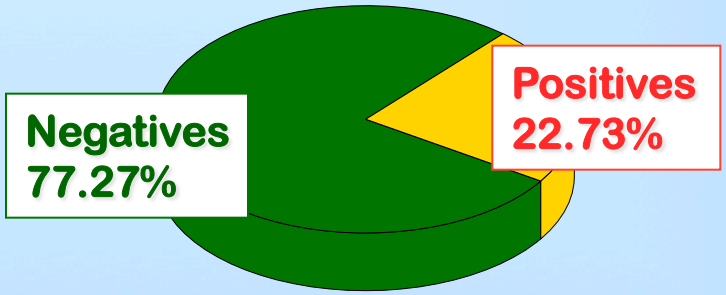
The hydrogen bond interaction between the ligands and receptor play a major role.

K-NN CP-ANN LS-SVM	Accuracy %	False negative %	False positive %
studied set:	87.5-89.7	3.0-6.1	17-24
external set:	82.8-83.9	8.8	30-33.3

Most important for Precautionary Principle: lesser number of EDs predicted as not EDs

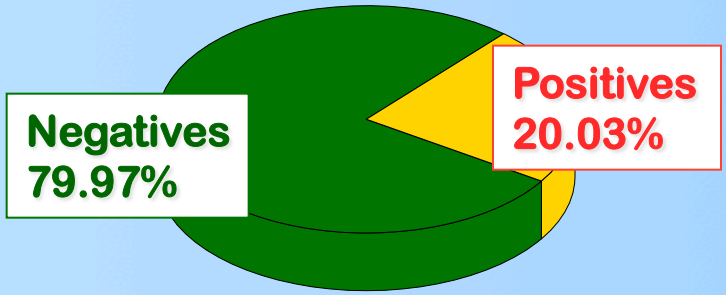
Screening of > 57000 potential EDCs

53.188 compounds (93.12%) compounds are in AD



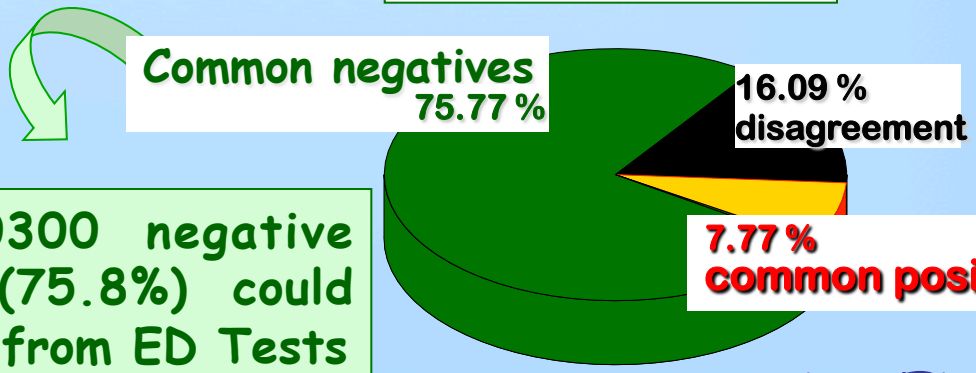
k-NN model

CONSENSUS



LS-SVM model

more Hazardous Chemicals : focus for TESTS



Common 40300 negative compounds (75.8%) could be excluded from ED Tests

(Both models: high accuracy and low false negative rate)

CONSENSUS approach is useful to prioritize chemicals before more expensive assays

SARs, QSARs, sistemi esperti, progetti

Software free:

- Episuite : *www.epa.gov/oppt/exposure/pubs/episuite.htm*
- Toxtree : *ambit.acad.bg/toxTree*
- OECDToolbox: www.oecd.org/env/existingchemicals/qsar.
- OpenToxframework :www.opentox.org

Software commerciale :

- DEREK,MultiCASE,HazardExpert,ToxAlert,ToxBoxes…
- *In silico first consortium (Multicase,Lhasa,MolecularNetworks,Leadscope)*

- QSAR Model Databases(QMDBs): JRC QSAR Model Database http://ecb.jrc.ec.europa.eu/qsar/qsar-tools/

Diversi Progetti EU FP6-FP7 : OSIRIS, CAESAR, OpenTox, CADASTER,….

QSAR TOOLBOX

What is the Toolbox?

The Toolbox is a software application to identify and fill (eco)toxicological data gaps for chemicals hazard assessment. Grouping chemicals into chemical categories is crucial to the workflow of the Toolbox.

What tools are available?

The Toolbox contains:

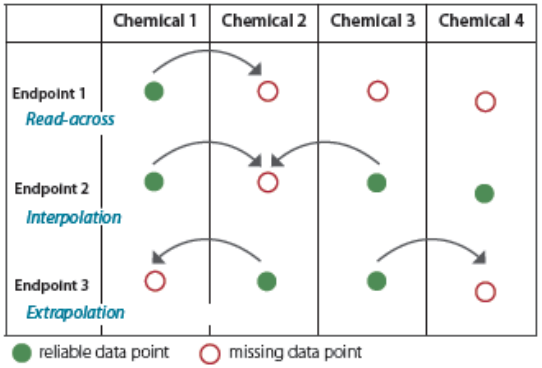
- databases with results from experimental studies;
- accumulated knowledge for structural characteristics (alerts) that can indicate the presence of hazards and other properties, and
- tools to estimate missing experimental values by read-across, by trend analysis (i.e. interpolating (preferred) or extrapolating from a trend (increasing, decreasing, or constant) from tested to untested chemicals within a category) and/or by (Q)SAR models.

Key features of the Toolbox

The Toolbox allows a user to systematically group chemicals into categories according to the presence or potency of a particular effect for all members of the category. It allows a quick evaluation of chemicals for common mechanisms or modes of action as well as for common toxicological behaviour or consistent trends among results related to regulatory endpoints.

What is a chemical category?

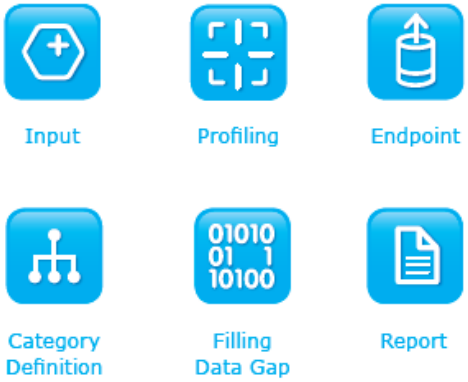
A chemical category is a group of chemicals with physical-chemical, toxicological and/or ecotoxicological properties and/or environmental fate properties that are likely to be similar or follow a regular pattern because of their similar chemical structure. Using this category approach, not every chemical needs to be tested for every endpoint because the available test results for the members of the category allow an estimation of the results for the untested endpoints.



As illustrated above, a chemical category can be represented graphically as a two-dimensional matrix in which category members occupy different columns, and the category endpoints occupy different rows. Data gaps may be filled by read-across from a tested to an untested chemical or by trend analysis.

Why the chemical category approach?

- It uses an identified mechanism or mode of action of chemicals to justify similarity within the category.
- It allows for entire categories of chemicals to be assessed when only a few members have been tested, saving animals and costs.
- It enables robust hazard assessment through mechanistic comparisons.
- It facilitates the optimisation of testing strategies for members in a category.



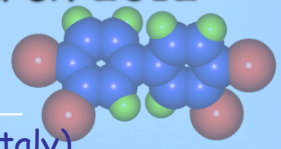
http://www.qsartoolbox.org

(QSAR) Toolbox (version 3.1)

La posizione di ECHA e OECD

-The software helps registrants and authorities to use Quantitative Structure-Activity Relationship ((Q)SAR) methodologies to group chemicals into categories and to fill data gaps by read-across, trend analysis and to assess the (eco)toxicity hazards of chemicals under REACH. This helps to reduce costs and unnecessary testing on vertebrate animals.

Helsinki, 12 March 2012



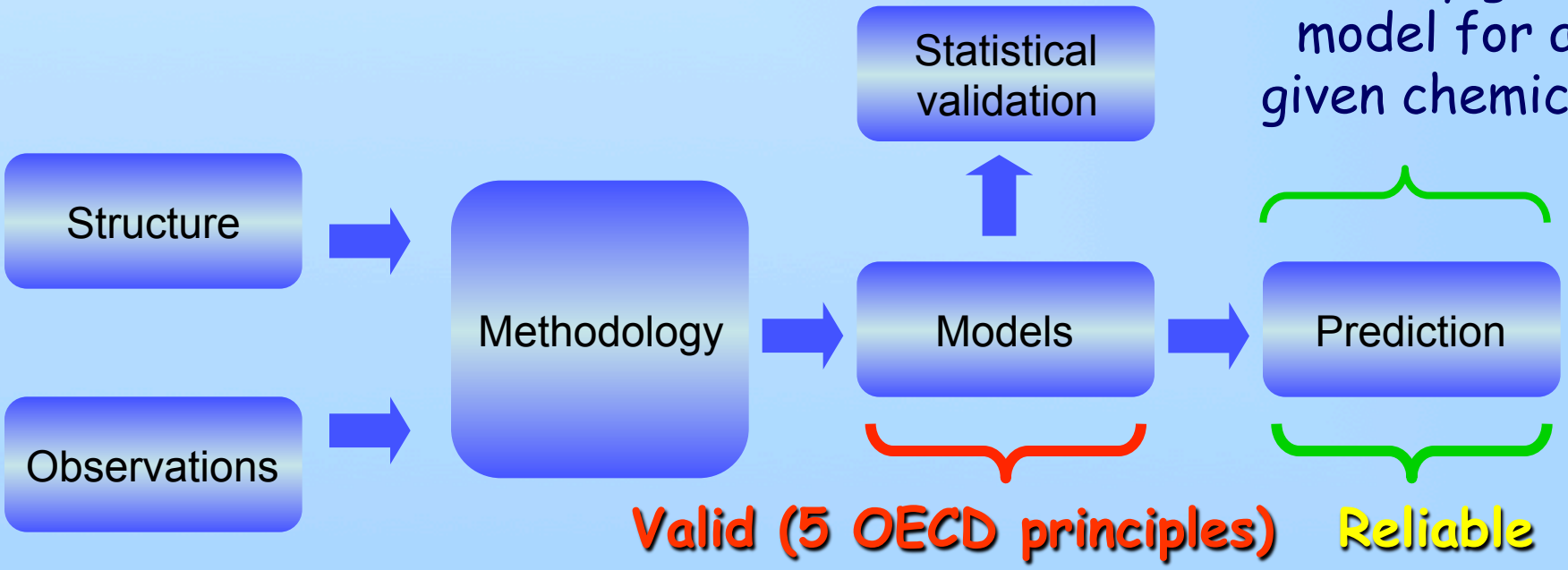
QSAR Model Reporting Formats (QMRF)

QMRF

Robust summary of a (Q)SAR model, which reports key information on the model according to the OECD validation principles.

QPRF

Description and assessment of the prediction made by given model for a given chemical



ECHA guidance

QSAR Reporting Formats (QMRF) and JRC QSAR Model Database

http://qsardb.jrc.ec.europa.eu/qmrf/

Il JRC QSAR Model Database è stato creato per facilitare l'utilizzo di modelli QSAR, sviluppati e verificati in accordo con i principi OECD, ma senza darne giudizi o "validity statement".

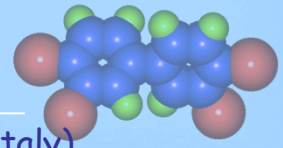
E' accessibile gratuitamente e raccoglie i dati relativi alla descrizione ed alla validazione dei modelli, secondo i reports QMRF.

I reports QPRF (QSAR Prediction Reporting Format) riportano i valori delle predizioni generate dai modelli QSAR.

Utilizzo dei QMRF

- L'inclusione di un modello nel JRC QSAR Model Database non implica la sua indicazione come modello preferenziale da parte dell' autorità competente (i.e. ECHA).

- La responsabilità della scelta e dell'uso dei modelli QSAR è a carico degli utilizzatori, che devono comunque fornire all'ECHA la documentazione QMRF, qualora utilizzino modelli QSAR nella fase di registrazione di una sostanza.



QMRF	***QMRF identifier (ECB Inventory)**:To be entered by ECB* ***QMRF Title**: Statistically Validated QSARs, Based on Theoretical Descriptors, for Modeling Aquatic Toxicity of Organic Chemicals in Pimephales promelas (Fathead Minnow)* ***Printing Date**:Jul 31, 2009*	QMRF

1.QSAR identifier

1.1.QSAR identifier (title):

Statistically Validated QSARs, Based on Theoretical Descriptors, for Modeling Aquatic Toxicity of Organic Chemicals in Pimephales promelas (Fathead Minnow)

1.2.Other related models:

no other information available

1.3.Software coding the model:

MOBY DIGS Software for multilinear regression analysis and variable subset selection by Genetic Algorithm, ver 1 beta for Windows, 2004 Todeschini Roberto, Talete srl, Milan (Italy)

2.General information

2.1.Date of QMRF:

29/05/2009

2.2.QMRF author(s) and contact details:

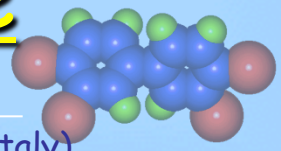
[1]Papa Ester QSAR Research Unit in Environmental Chemistry and Ecotoxicology, Department of Structural and Functional Biology, University of Insubria ester.papa@uninsubria.it

La modellistica QSAR in chimica ambientale sfrutta l'esperienza già acquisita dall'uomo (dati sperimentali noti) e la utilizza per evitare di ripetere errori del passato

(ad es. PBT : persistenza, bioaccumulo, tossicità, Endocrine Disruption, ecc.).

Predire il comportamento nell'ambiente di una molecola dalla sola conoscenza della sua struttura chimica può permettere di

"Prevenire piuttosto che curare"



Grazie a voi per l'attenzione ed ai miei Collaboratori

http://www.qsar.it