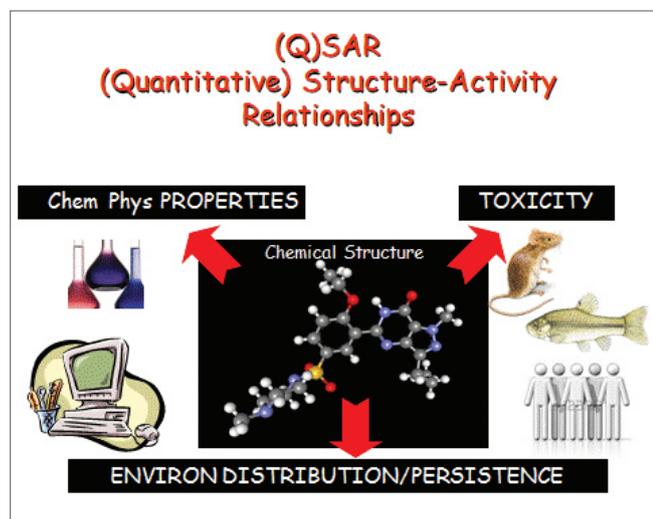


# MODELLI QSAR ED APPLICAZIONI PER LA *GREEN CHEMISTRY*

La modellistica QSAR è di grande utilità sia nell'evidenziare e prioritizzare composti potenzialmente pericolosi, anche senza dati sperimentali, sia nella possibilità di progettare molecole più sicure come alternative a composti messi al bando, nell'approccio "benign by design" della *green chemistry*.



La *green chemistry* è stata definita da Paul Anastas, che ne ha proposto i famosi 12 principi, come "la progettazione, la preparazione e l'applicazione di composti e processi chimici con lo scopo di ridurre o eliminare l'uso e la generazione di sostanze chimiche pericolose" [1-3]. Il più importante aspetto della *green chemistry* è, a mio parere, quello riguardante la progettazione: il "design", perché affronta alla radice ed intenzionalmente l'obiettivo della sostenibilità. Il "design" non può essere casuale, ma si deve basare su precedenti conoscenze ed esperienze per poter al meglio pianificare sintesi chimiche più sicure, non solo trovando ed applicando condizioni sperimentali ottimali (come solventi, temperatura, catalizzatori ecc.), ma anche avendo *a priori* come obiettivo la sintesi di composti con strutture molecolari intrinsecamente non pericolose ("benign by molecular design").

Tre dei 12 principi alla base della *green chemistry* sono incentrati sulla progettazione della sintesi di molecole con caratteristiche di non peri-

colosità: il 3°, il 4° ed il 10° auspicano la sintesi di molecole meno o per nulla tossiche per l'uomo e l'ambiente, ed il 10° include anche la necessità di una non persistenza dei composti stessi o dei loro prodotti di degradazione. Il disegno molecolare è quindi il punto di partenza di una sintesi "green".

Per progettare nuovi composti che abbiano intrinseche proprietà di non tossicità e persistenza, è però necessario conoscere quali caratteristiche strutturali dei composti esistenti siano determinanti per conferire tali proprietà pericolose alle molecole. È quello che possono fare le metodologie di modellamento QSAR (*Quantitative Structure-Activity Relationship*).

La modellistica QSAR, trovando la relazione esistente tra una determinata proprietà o attività, misurata sperimentalmente, di alcune molecole note, permette di estrarre l'informazione intrinseca nella struttura molecolare e di prevedere tale proprietà/attività anche per molecole per le quali non sia disponibile il corrispondente dato sperimentale o addirittura di ipotizzare e disegnare la struttura molecolare di composti più sicuri di quelli individuati come potenzialmente pericolosi ("safer alternatives").

Buona parte dei 122 milioni di composti attualmente registrati nel Chemical Abstract Service (CAS) sono sostanze organiche di sintesi; ogni giorno se ne aggiungono 15.000, di cui una buona parte sono ancora composti organici di sintesi. È quindi evidente quanto sia fondamentale comprendere quali siano le caratteristiche strutturali che conferiscono varie tipologie di intrinseca pericolosità a composti con struttura organica. Comprendere, sulla base delle conoscenze esistenti, quali siano le proprietà molecolari di un composto chimico, che hanno un potenziale impatto negativo sulla salute umana e/o sull'ambiente ed il destino che può avere nell'ambiente o nel biota, è essenziale per la sostenibilità, per evitare di perpetuare errori del passato. *Prevenire è meglio che curare!*

La modellistica QSAR può essere proficuamente usata in questo senso, ma, come tutti i modelli che danno comunque un risultato (un numero si ottiene sempre da un modello QSAR!), va sviluppato secondo regole precise per dare risultati attendibili ed utilizzato con cognizione di causa:

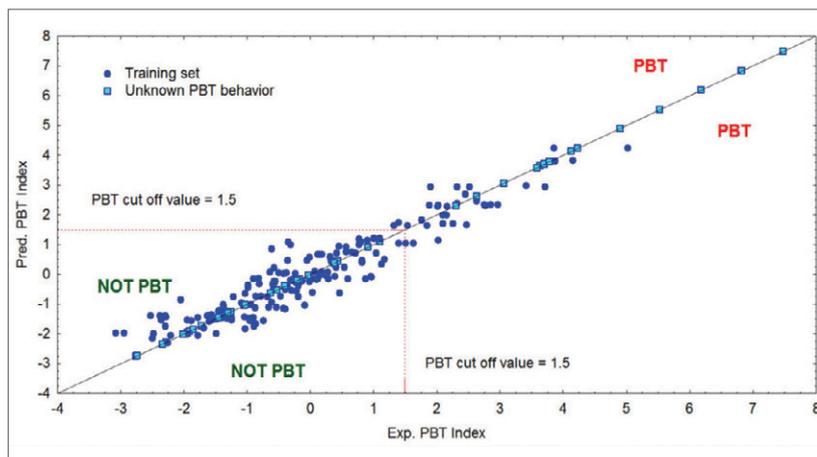


Fig. 1 - Modello QSAR dell'Insubria PBT Index e suo plot sperimentali-predetti

non deve essere usato da inesperti come un “push a button and find a correlation” [4]. Un valido modello QSAR deve soddisfare i famosi 5 principi di validazione OECD (Organisation for Economic Co-operation and Development) [5], cioè avere: 1) un chiaro end-point: la proprietà in esame, 2) un definito algoritmo: il modello deve essere riproducibile, 3) un definito dominio di applicabilità: non si può applicare qualsiasi modello QSAR a qualsiasi molecola, 4) validazione interna ed esterna: deve essere stabile nel riprodurre dati noti e predittivo per molecole nuove, e se possibile 5) un'interpretazione meccanicistica.

Il primo e fondamentale aspetto, che deve essere seriamente considerato ed è incluso nel Principio 1, è relativo ai dati sperimentali da cui si parte per estrarre l'informazione utile: è quello della costruzione del cosiddetto *training set*, che è il vero proprio allenamento del QSAR. La ricerca e attenta selezione di dati sperimentali attendibili e omogenei è un passaggio cruciale che un buon modellista QSAR deve affrontare per poter garantire che da un input sperimentale attendibile possano essere derivati dati predetti dai modelli con altrettanta attendibilità. È la “cura” dei dati oggetto del modellamento, che garantisce di evitare l' indesiderato e potenzialmente possibile: “Spazzatura entra, spazzatura esce”, meglio noto come “*Garbage In, Garbage Out* (GIGO)”.

Il modello QSAR deve essere chiaramente espresso nella sua formulazione matematica e reso comprensibile agli utilizzatori (Principio 2), non deve essere una “*black box*”, in quanto deve garantire di poter essere applicato, senza problemi, anche da chi non l'ha sviluppato. Per questo motivo è anche fondamentale che ogni potenziale utilizzatore sia consapevole della ragionevolezza di applicabilità di un certo modello alla/e molecola/e di suo interesse, per cui ogni modello deve essere sempre corredato dalle indicazioni relative al suo Dominio di Applicabilità (AD) (Principio 3). L'attendibilità di ogni dato predetto da un modello può essere verificata e garantita solo da modelli validi, in termini di capacità di riprodurre bene i dati del training set (*fitting*), di stabilità o robustezza (validazione interna) e di capacità di predire bene nuove molecole non incluse nel training set (validazione esterna) (Principio 4) [6, 7].

Alcune interessanti e recenti applicazioni della modellistica QSAR sono state sviluppate dal gruppo di ricerca QSAR in Insubria negli ultimi anni in modo particolare per lo *screening*, il *ranking* e la prioritizzazione di alcune importanti classi di inquinanti emergenti: lo scopo principale è sempre quello di “imparare” dai pochi dati sperimentali le informazioni

strutturali legate a determinate caratteristiche di pericolosità e di dare quindi indicazioni utili per prevedere *a priori* potenziale pericolo insito in determinate strutture molecolari.

In particolare, è stato sviluppato [8] ed applicato a numerose classi chimiche di composti di grande interesse ambientale [9] e di ampio utilizzo nella vita quotidiana, come i prodotti per la cura personale [10], i ritardanti di fiamma [11] e i farmaci [12], che sono attualmente riconosciuti come inquinanti ambientali emergenti, un modello globale che permette di prevedere il potenziale comportamento di Persistenza, Bioaccumulo e Tossicità (PBT).

Le predizioni ottenute dal modello QSAR dell'Insubria PBT-Index, che, utilizzando 4 descrittori della struttura molecolare, predice cumulativamente le caratteristiche PBT sono state confrontate con quelle ottenute dal famoso ed ampiamente utilizzato PBT Profiler dell'USA-EPA [13], che fornisce predizioni separate per le tre proprietà. Le predizioni in accordo per i due diversi metodi hanno permesso di stilare liste di priorità di composti potenzialmente PBT, sui quali concentrare la successiva sperimentazione. Questo filtro, operato su migliaia di sostanze senza dati sperimentali, porta quindi ad un risparmio notevole in termini di tempi, costi e vite animali sacrificate, anche in linea con quanto richiesto dalla normativa europea REACH (*Registration Evaluation Authorization and restriction of Chemicals*). L'approccio predittivo, basato su descrittori della struttura molecolare, permette anche di ipotizzare la struttura di nuove molecole per verificarne *a priori* la potenziale caratteristica di PBT, per progettare la sintesi di nuovi composti che, come richiede la fase di autorizzazione del REACH, dovrebbero essere alternative più sicure a composti messi al bando. In particolare, sulla base dei descrittori più significativi nell'equazione del modello, le alternative più sicure dovrebbero avere un nullo o basso numero di alogeni (nX) e doppi legami (soprattutto aromatici) (nBondsM) e un numero più elevato di gruppi polari (OH, NH) in grado di formare legami ad idrogeno (nHBDOn\_Lipinski). È interessante citare il caso dello screening dei ritardanti di fiamma [11] in cui sono stati riconosciuti come PBT non solo i composti che sono già stati messi al bando come i Poli Bromo Difenil Eteri (PBDE), ma anche alcuni, principalmente alogenati, che sono stati immessi sul mercato come presunte alternative più sicure, ma a *posteriori* sono già stati inseriti nelle liste di priorità, in quanto anch'essi probabili PBT sulla base di evidenze sperimentali raccolte dopo la loro introduzione nell'ambiente. Da questo screening è stato invece possibile evidenziare la minore tendenza ad essere PBT dei derivati organofosforici a struttura più semplice. Se la modellistica QSAR fosse stata applicata *a priori*, secondo i principi 3, 4 e 10 della *green chemistry*, sarebbe stato possibile evidenziare già in fase di progettazione che la struttura di alcuni sostituti, presunte “*safer alternatives*” richieste dal REACH, non garantiva sicurezza per potenziali proprietà PBT.

Lo screening di oltre 500 prodotti per la cura personale ha portato ad evidenziare la pericolosità come potenziali PBT di solo 8 composti, in particolare filtri delle radiazioni UV a struttura benzotriazolica [10] utilizzati nelle creme solari (Fig. 2), mentre lo screening effettuato su circa 1.200 ingredienti di prodotti farmaceutici ha permesso di stilare una lista di priorità di 35 farmaci potenzialmente PBT [12].

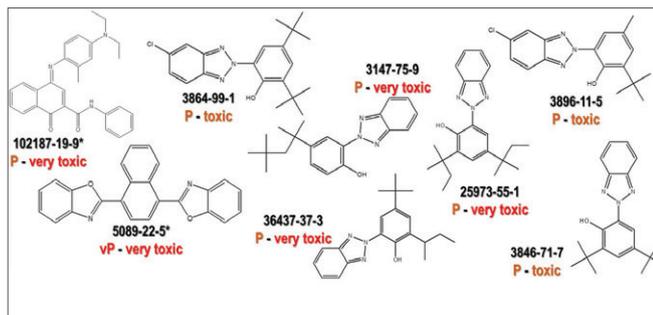


Fig. 2 - Composti individuati come potenziali PBT e loro grado di tossicità verso organismi acquatici

Il modello PBT Index è inserito nel modulo QSARINS-Chem del software QSARINS (Fig. 3) per un possibile ampio utilizzo in fase di disegno molecolare per la progettazione di nuovi composti nella *green chemistry* [14]. Altra utile applicazione della modellistica QSAR è quella di prevedere *a priori* la potenziale tossicità acuta globale su diversi organismi acquatici della catena trofica (alghe, Dafnia e pesci). Dopo aver sviluppato modelli specifici per la tossicità di prodotti per la cura personale sui vari organismi acquatici, sulla base dei pochi dati sperimentali esistenti, integrati dai dati predetti dai suddetti modelli, è stato infatti recentemente proposto un Indice di Tossicità Acquatica (ATI) [15]. Tale indice è stato modellato con descrittori molecolari per poter prevedere, già dal semplice disegno della struttura molecolare, se una sostanza può essere o meno tossica per gli organismi del comparto acqua. Approccio simile è stato sviluppato e proposto per i farmaci, individuandone 35 con potenziali effetti tossici, tra cui quelli indicati come esempi in Fig. 4.

Gli esempi illustrati dimostrano come la modellistica QSAR possa sfruttare i dati sperimentali noti e la pur scarsa informazione esistente sulla pericolosità delle molecole per sviluppare modelli rigorosamente validati per la loro predittività ed applicabili anche a composti non ancora sintetizzati. Questo approccio *a priori*, se opportunamente applicato, è di grandissima utilità per la progettazione di composti chimici sempre più sicuri.

Fig. 3 - Software QSARINS per lo sviluppo e la validazione di modelli QSAR

**BIBLIOGRAFIA**

- [1] P.T. Anastas, J.C. Warner, *Green Chemistry: Theory and Practice*, Oxford Univ. Press, New York, 1998.
- [2] I. Horvath, P.T. Anastas, *Chem. Rev.*, 2007, **107**, 2167.
- [3] P.T. Anastas, T.C. Williamson, in *Green Chemistry: Designing Chemistry for the Environment*, ACS Books, Washington, DC, 1996, 1.
- [4] P. Gramatica *et al.*, *Molecular Informatics*, 2012, **31**, 817.
- [5] <http://www.oecd.org/dataoecd/33/37/37849783.pdf>
- [6] A. Tropsha *et al.*, *QSAR & Combinatorial Science*, 2003, **22**, 69
- [7] a) P. Gramatica, *QSAR & Combinatorial Science*, 2007, **26**, 694; b) *Molecular Informatics*, 2014, **33**, 311.
- [8] E. Papa, P. Gramatica, *Green Chemistry*, 2010, **12**, 836.
- [9] P. Gramatica *et al.*, *Environment International*, 2015, **77**, 25.
- [10] S. Cassani, P. Gramatica, *Sustainable Chemistry and Pharmacy*, 2015, **1**, 19.
- [11] P. Gramatica *et al.*, *J. Hazardous Materials*, 2016, **306**, 237.
- [12] A. Sangion, P. Gramatica, *Environmental Research*, 2016, **147**, 297.
- [13] Persistent, Bioaccumulative, and Toxic Profiles Estimated for Organic Chemicals On-Line, 2006. <http://www.pbtprofiler.net/>
- [14] P. Gramatica *et al.*, *J. Computational Chemistry*, a) 2013, **34**, 2121; b) 2014, **35**, 1036.
- [15] P. Gramatica *et al.*, *Green Chemistry*, 2016, **18**, 4393.

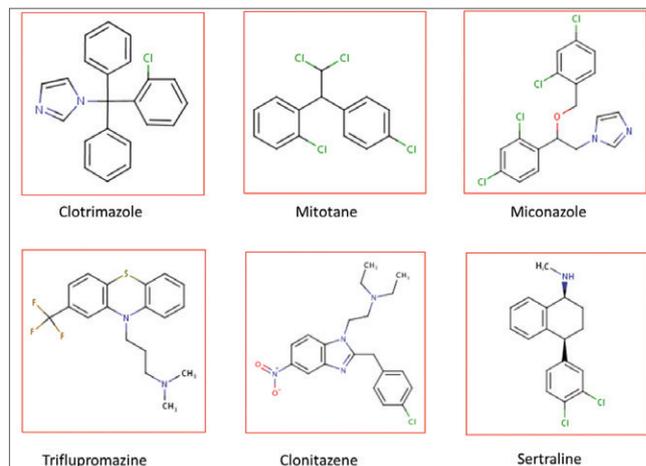


Fig. 4 - Alcuni esempi di farmaci identificati dai modelli QSAR come potenziali PBT ed in particolare come tossici per gli organismi acquatici

**QSAR Modeling in Green Chemistry**

QSAR modeling can be usefully applied either in highlighting and prioritizing chemicals potentially hazardous, even among those without experimental data, and in designing safer compounds as alternatives to those already banned, in the “benign by design” approach of *green chemistry*. Examples of QSAR modeling for screening flame retardants, personal care products and pharmaceuticals as PBT (Persistent Bioaccumulative Toxic) by the PBT Index in QSARINS are presented.