

Paola Gramatica

Unità di ricerca QSAR in Chimica Ambientale
ed Ecotossicologia, DBSF

Università dell'Insubria (VA)

paola.gramatica@uninsubria.it

www.qsar.it

QSAR E LA NUOVA NORMATIVA EUROPEA REACH

La modellistica predittiva QSAR, della quale vengono qui illustrati i principi essenziali, è un valido strumento di previsione delle proprietà e della tossicità di composti chimici per i quali non sono disponibili dati sperimentali.

Tale approccio viene inquadrato nella nuova normativa europea Reach.

La modellistica QSAR (*Quantitative Structure-Activity Relationship*) è un approccio predittivo denominato *in silico*, perché basato su metodi di chimica computazionale e modelli matematico-statistici che utilizzano i computer per i loro calcoli. Gli studi QSAR si basano sulla conoscenza esistente riguardo ad un certo "endpoint" (effetto biologico, proprietà chimico-fisica, reattività), già noto sperimentalmente per un certo numero di composti chimici, con lo scopo di trarre informazioni da questa conoscenza, per generalizzarla e permettere di fare predizioni per altri composti chimici per i quali non si hanno dati sperimentali, o addirittura prima di sintetizzarne di nuovi (*chemical design*).

La necessità di avere a disposizione efficaci ed attendibili strumenti predittivi, come i modelli a base strutturale QSAR, è particolarmente evidente se si considera che sono oltre 36 milioni le sostanze chimiche, organiche ed inorganiche, attualmente registrate sul Chemical Abstract Service (CAS), con un incremento giornaliero di circa 12.000, pari ad oltre 4 milioni all'anno. Di questo universo chimico (Fig. 1) sono più di 21 milioni i prodotti disponibili in com-

mercio, di cui solo 247.000 molecole regolate ed inventariate dai vari Paesi (ad esempio circa 100.000 nell'EINECS europea, 83.000 nella US-TSCA). Purtroppo però per la maggior parte dei composti chimici esistenti, ed anche ampiamente utilizzati (ad esempio per gli HPV, *High Production Volume*), si conosce ben poco su proprietà ed effetti sia sull'ambiente sia sull'uomo. Da dati dell'EPA americana è noto che circa il 51% delle sostanze che vengono valutate annualmente per l'immissione sul mercato (da 1.500 a 2.000) non presenta nessun tipo di valutazione tossicologica, circa il 45% ha qualche test di tossicità umana e meno del 5% test di ecotossicità.

Questa grave mancanza di dati, in particolare per le sostanze "existing" cioè immesse sul mercato prima del 1981, è stato uno dei motivi principali che hanno spinto la Commissione Europea a varare la nuova normativa Reach (*Registration Evaluation and Authorization of Chemicals*) [1] entrata in vigore nel giugno del 2007, dopo una lunga e difficile gestazione iniziata con la pubblicazione del *White paper* nel 2001.

Il principale obiettivo di questa nuova strategia chimica è di garantire un alto livello di protezione per la salute dell'uomo e dell'ambiente, nello stesso tempo assicurando un efficace funzionamento del mercato e stimolando l'innovazione e la competitività dell'industria chimica. Tale normativa, che interessa tutte le imprese con sede nell'Unione Europea che vi fabbricano o vi importano sostanze chimiche, prevede fino al 1° dicembre 2008 un'iniziale fase di pre-registrazione, che riguarda le sostanze chimiche fabbricate o importate nell'UE in quantitativi annuali pari o superiori a 1 t, sia già esistenti (dette *phase-in*) sia nuove (dette *non phase-in*).

Prevede poi che vengano forniti entro il 1° dicembre 2010 tutti i dati richiesti per le sostanze HPV (>1.000 t) e per quelle che hanno caratteristiche di elevata Persistenza e Bioaccumulo (vPvB), le molecole PBT (Persistenti, Bioaccumulabili e Tossiche) o CMR (Cancerogene Mutagene, tossiche per la Riproduzione). Per le sostanze prodotte in quantitativi minori sono previsti tempi più lunghi: entro il maggio 2013 per quelle tra 100 e 1.000 t ed entro il maggio 2018 per tutte le altre.

Sono immediatamente evidenti gli enormi sforzi richiesti all'industria allo scopo di colmare questa carenza di dati, in tempi relativamente brevi. La sfida per la comunità scientifica e l'industria sarà quella di sviluppare strategie che permettano di valutare e gestire, in modo efficiente e pratico, i pericoli derivanti da composti chimici. Vi è quindi la necessità di aumentare l'efficienza, la gestibilità dei costi e l'attenzione sulle procedure di valutazione del rischio, riducendo nel contempo l'attuale utilizzo di test su animali. La realizzazione di misure sperimentali, sia per la determinazione di proprietà chimico-fisiche e reattività sia per test di tossicità acuta e cronica, richiesti dal Reach (che in un articolo titola: "No data no market"), è particolarmente gravosa in termini di costi, di tempi, di personale, di strutture adeguate ed è quindi impensabile ottenere tutte le informazioni necessarie nei tempi stabiliti solo con la sperimentazione.

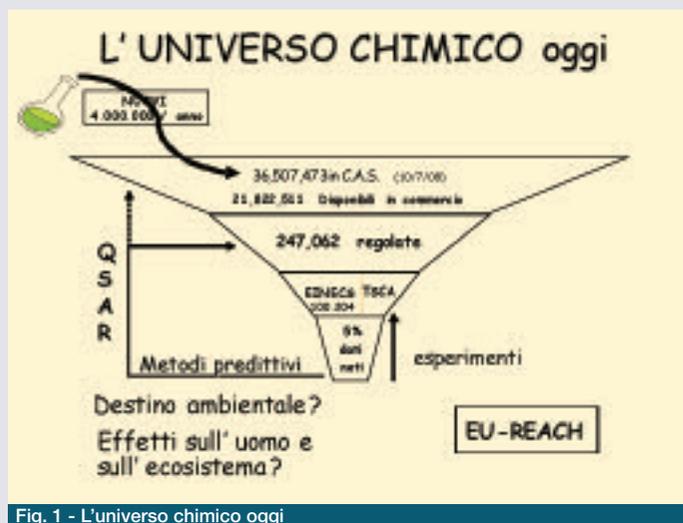


Fig. 1 - L'universo chimico oggi

Per tutti i motivi suddetti è previsto nell'Annesso XI del Reach l'utilizzo di metodi alternativi sia *in vitro* sia *in silico*, modelli QSAR inclusi. Negli Stati Uniti modelli QSAR sono stati utilizzati, già dai primi anni Ottanta, dall'OPPT (*Office of Pollution Prevention and Toxics*) dell'EPA, nell'ambito della procedura PMN (*Pre-Manufacture Notification*) del *New Chemical Program* (NCP), quando il produttore non è in grado di fornire il dato sperimentale. Uno dei software più utilizzati in questo ambito è EPISuite, liberamente utilizzabile da Internet [2], come pure il PBT Profiler [3].

Dalla pubblicazione del White Paper del 2001 ad oggi si sono succedute iniziative a vari livelli per arrivare ad un consenso internazionale sull'applicabilità dei modelli QSAR nell'ambito della regolamentazione dei prodotti chimici nei vari Paesi. Nel 2002 in un workshop dei più autorevoli esperti QSAR sono stati stilati i Principi di Setubal, successivamente discussi e rivisti, fino ad arrivare ai definitivi Principi dell'OECD del 2004 [4].

Per essere utilizzato in regolamentazione un modello QSAR deve avere:

- 1) un *endpoint* definito
- 2) un algoritmo non ambiguo
- 3) un definito dominio di applicabilità
- 4) appropriate misure di riproducibilità, robustezza e predittività
- 5) un'interpretazione meccanicistica, se possibile.

In questi ultimi anni il mondo del QSAR è stato rivitalizzato da un notevole incremento di iniziative, pubblicazioni scientifiche, workshops e congressi, come pure da parecchi progetti europei in corso o di imminente attivazione nel 7° Programma Europeo. Un ampio dibattito si è aperto soprattutto sulle modalità di validazione dei modelli QSAR: solamente quelli validati per la loro predittività esterna [5, 6] possono essere effettivamente considerati predittivi per nuove molecole e quindi utilizzabili in sostituzione alla sperimentazione.

La modellistica QSAR

I QSAR sono modelli statistico-matematici che mettono in relazione la struttura chimica di un composto con le sue proprietà chimico-fisiche o i suoi effetti biologici. Poiché le strutture chimiche sono entità non-quantitative, queste devono essere preventivamente trasformate in valori numerici mediante l'utilizzo dei descrittori molecolari. Il descrittore molecolare è il risultato finale di una procedura logica e matematica che trasforma l'informazione chimica e codifica la rappresentazione simbolica della molecola in un numero utile. In letteratura, si trovano oggi migliaia di descrittori molecolari (ne sono stati contati più di 3.000 in [7]).

La modellistica QSAR è basata su un fondamentale assunto, ampiamente dimostrato da Corvin Hansch [8], considerato il padre di questo approccio, che "la struttura molecolare di una sostanza chimica influenza le sue proprietà chimico-fisiche ed attività biolo-

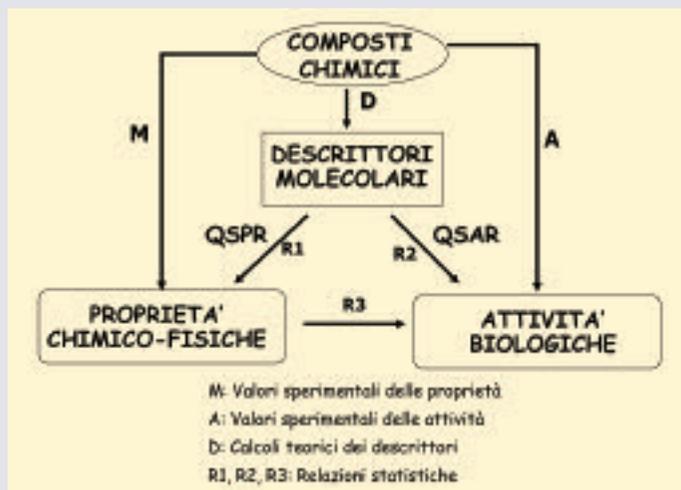


Fig. 2

giche” e che “composti simili si comportano in modo simile”. L’individuazione e l’interpretazione delle caratteristiche strutturali di una molecola organica, correlate ad una particolare proprietà o attività, permette di sviluppare una funzione matematica f che correla tale comportamento alla struttura chimica. Questa funzione matematica f (il modello QSAR) può essere usato per predire dati riguardanti composti non testati o addirittura non ancora sintetizzati. Una classica equazione di Hansch è:

$$\text{Attività Biologica} = f(\text{struttura}) = a + \log P + cE + dS + \dots$$

in cui un’attività biologica di una molecola (ad esempio una risposta di tossicità ad un test) è espressa come funzione di alcune proprietà chimico-fisiche e proprietà strutturali. $\log P$, o $\log K_{ow}$, è il coefficiente di ripartizione ottanolo/acqua, il parametro di idrofobicità che mima l’interazione idrofobica con la membrana cellulare, rappresenta quindi la probabilità di un composto di attraversare tale membrana esplicando un’azione tossica nella cellula. E ed S sono descrittori strutturali, rispettivamente elettronici e sterici, che rappresentano la possibilità di una molecola di interagire con il sito attivo di un sistema enzimatico e di essere quindi attiva. La tossicità di un gran numero di composti è stata modellata e predetta da equazioni simili, ma negli ultimi anni gli studi QSAR si sono notevolmente sviluppati con l’introduzione di nuovi descrittori molecolari teorici, di nuovi metodi di sviluppo e validazione dei modelli QSAR, di strategie per lo studio della similarità strutturale volte alla individuazione di categorie chimiche e all’applicabilità del “read-across”, che permette di stimare dati incogniti da dati noti per molecole simili.

Uno schema generale della procedura QSAR è riportato in Fig. 2.

I dati sperimentali di proprietà chimico-fisiche o attività biologiche, noti per un certo numero di sostanze, sono ottenuti da misure M ed A, rispettivamente. L’informazione strutturale viene tradotta in valori numerici dai descrittori strutturali teorici, che possono essere calco-

lati (D) con software specifici. Le relazioni R1, R2 sono le funzioni matematiche di QSARs (per le Attività) e QSPRs (per le Proprietà), la cui applicazione permette di ottenere i dati predetti.

Quindi è evidente che per ottenere un buon modello QSAR sono necessari:

- dati sperimentali di alta qualità sui quali trovare la relazione con la struttura molecolare, il più possibile numerosi ed omogenei. È importante sottolineare che la qualità dei dati predetti dai modelli QSAR non potrà mai essere superiore a quella dei dati sperimentali modellati, vale infatti il detto: “Garbage in, garbage out!”;
- la miglior rappresentazione numerica dei diversi aspetti (mono-, bi- e tridimensionali) della struttura molecolare mediante i descrittori molecolari;

- metodi statistico-matematici per sviluppare modelli quantitativi con validate capacità predittive (ad esempio metodi chemiometrici).

Una caratteristica dei metodi chemiometrici è l’attenzione posta alla reale predittività del modello ed allo studio del dominio strutturale di applicabilità del modello stesso, perché non si possono applicare modelli QSAR a qualsiasi composto, indipendentemente dall’informazione contenuta nelle molecole utilizzate per sviluppare il modello (*training set*): questi aspetti sono particolarmente importanti quando l’obiettivo è quello di utilizzare i dati predetti dai modelli QSAR in sostituzione di quelli sperimentali. I modelli QSAR sono modelli di regressione, quando il dato modellato e predetto è un valore numerico relativo ad un *endpoint* (ad esempio un valore di EC50), e modelli di classificazione quando si modella e predice l’appartenenza ad una classe (per esempio, mutageno o non). Tale approccio, essendo basato essenzialmente sulla struttura, può anche essere utilizzato per progettare molecole più sicure (*chemical design*) come sostituti di molecole con caratteristiche indesiderate (ad esempio i già citati PBT, i CMR e gli ED: *Endocrine Disruptors* o interferenti endocrini) che necessitano della più rigorosa fase dell’autorizzazione nel



Fig. 3

Reach, ma per le quali è richiesta anche un'azione di sostituzione. Un database di modelli QSAR validati che potrebbero essere utilizzati per la regolamentazione sta per essere pubblicato sul sito web dell'ECB (*European Chemical Bureau*) del JRC (*Joint Research Center*) di Ispra [9].

Oltre all'approccio chemiometrico, un altro filone della modellistica QSAR è quello interessato in particolar modo allo studio dei meccanismi molecolari che sono alla base di un'attività biologica dannosa di interesse (ad esempio sensibilizzazione della pelle, irritazione degli occhi ecc.): con questo approccio è stato sviluppato, nell'ambito dell'OECD, un *QSAR Application Toolbox*, basato sul concetto di raggruppamento delle sostanze in categorie chimiche per similarità [4]. In Fig. 3 sono schematizzati i principali ambiti di applicazione della modellistica QSAR: la predizione di dati mancanti sperimentalmente, l'evidenziazione di molecole pericolose, anche su grandi librerie di composti ed ancor prima della loro sintesi e la possibilità di progettare composti alternativi più sicuri, con un approccio di *Green Chemistry*. Lo scopo finale è quello di costituire delle liste di priorità che permettano quindi la concentrazione delle risorse sulle sostanze prioritarie, che necessitano di adeguate misure sperimentali, ottimizzando quindi le risorse industriali e minimizzando l'uso della sperimentazione sugli animali.

PBT ed ED

I composti chimici PBT (Persistenti Bioaccumulabili e Tossici) ed ED (*Endocrine Disruptors*) sono oggetto di particolare attenzione nel Reach e richiedono un'autorizzazione quando ci siano dimostrate prove dell'impossibilità di sostituirli con prodotti più sicuri. In relazione a questi composti di grande rilevanza, come pure per i CMR, possono quindi essere identificati due livelli di intervento: a) la necessità di

strumenti per una loro veloce identificazione e prioritizzazione e b) lo stimolo per la ricerca e la produzione di alternative più sicure, in accordo con i principi della Green Chemistry.

Per questo motivo sono parecchi gli approcci di modellistica che sono stati e vengono continuamente sviluppati per una loro rapida identificazione (tra questi il PBT profiler dell'US-EPA [3] già utilizzato da alcune agenzie preposte alla regolamentazione, come pure alcuni modelli sviluppati dal gruppo QSAR dell'Insubria [10-13]). In particolare, un approccio QSAR basato sull'applicazione di diversi modelli da noi sviluppati [13], in una logica del *consensus modeling*, ci ha permesso di analizzare 53.000 composti per la loro potenziale attività di interferenti endocrini (ED): circa l'8% (pari a oltre 4.000 molecole) è risultato avere questa attività. Su tali molecole prioritarie sarà quindi necessario concentrare l'indagine sperimentale, mentre circa 40.000 sono state predette negative, in modo concorde dai diversi modelli, permettendo quindi di risparmiare risorse per effettuare test sperimentali su di esse. Le restanti molecole sono state predette con attività diversa da modelli diversi e quindi il principio di precauzione richiederebbe anche per queste una verifica sperimentale.

Conclusioni

Qualsiasi sia l'approccio utilizzato, statistico o meccanicistico, qualsiasi siano i software impiegati, è molto importante che la modellistica QSAR sia trattata da esperti nel campo o con una minima preparazione di base, per evitare di abusare delle sue potenzialità o di applicare modelli in ambiti non consentiti. Una valutazione conclusiva di esperti QSAR è sempre necessaria per riuscire a ricavare informazioni utili anche dall'integrazione di approcci diversi, in una strategia di *consensus modeling*, che permette di ottenere conclusioni più attendibili.

Bibliografia

- [1] <http://europa.eu.int/comm/environment/chemicals/reach.htm>
- [2] <http://epa.gov/opptintr/exposure/pubs/episuite.htm>
- [3] www.pbtprofiler.net/
- [4] www.oecd.org/document/23/0,3343,fr_2649_34379_33957015_1_1_1_1,00.html
- [5] A. Tropsha *et al.*, *QSAR & Comb. Sci.*, 2003, **22**, 69.
- [6] P. Gramatica, *QSAR & Comb.Sci.*, 2007, **26**, 694.

- [7] R. Todeschini, V. Consonni, *Handbook of Molecular Descriptors*. J. Wiley-VCH, Weinheim (GER), 2000, 667.
- [8] C. Hansch, T. Fujita, *J. Am. Chem. Soc.*, 1964, **86**, 1616.
- [9] <http://ecb.jrc.it/qsar/>
- [10] E. Papa, P. Gramatica, *PBTs Screening by Multivariate Analysis and QSAR Modelling*, 17th SETAC Meeting, Portugal, 2007.
- [11] H. Liu *et al.*, *Chem. Res. Toxicol.*, 2006, **19**, 1540.
- [12] H. Liu *et al.*, *J.Mol. Graph. Model.*, 2007, **26**, 135.

ABSTRACT

QSAR and the New European Legislation Reach

The predictive QSAR (Quantitative Structure-Activity Relationships), whose basic principles are here illustrated, is a valid tool for the prediction of physico-chemical properties and biological activities (toxicities, etc.) of chemicals that lack of experimental data. This modelling approach is here presented highlighting the potential applications to fill the data gaps and for hazard assessment in the new European legislation Reach.