



SIMONA KOVARICH, MANUELA PAVAN, ARIANNA BASSAN, ELENA FIORAVANZO
S-IN SOLUZIONI INFORMATICHE SRL, VICENZA
SIMONA.KOVARICH@S-IN.IT
WWW.S-IN.IT

VALUTAZIONE QSAR DI IMPUREZZE GENOTOSSICHE

La linea guida ICH M7 promuove l'uso delle metodologie (Q)SAR per la valutazione del potenziale genotossico delle impurezze di principi attivi farmaceutici. La valutazione computazionale consente uno screening delle impurezze attuali e potenziali, ottimizzando tempi e costi della sperimentazione.



I metodi computazionali, come gli approcci (Q)SAR ((Quantitative) Structure-Activity Relationships) e la modellistica molecolare, sono ampiamente utilizzati nel settore farmaceutico nelle diverse fasi dello sviluppo e della progettazione di nuovi farmaci, così come nell'ottimizzazione di farmaci già esistenti. La loro applicazione può essere finalizzata ad analisi conformazionali e dello spazio chimico (*drug design*), a screening di proprietà fisico-chimiche, ADME (Assorbimento, Distribuzione, Metabolismo, Eliminazione) e tossicologiche, o, più recentemente, anche per adempiere a requisiti regolatori, come per la valutazione delle impurezze del farmaco.

Le impurezze di prodotti farmaceutici comprendono qualsiasi componente della sostanza attiva o della sua formulazione derivante dalla sintesi o degradazione

del farmaco o da una contaminazione esterna. La caratterizzazione delle impurezze rappresenta una fase fondamentale all'interno del processo di sintesi e produzione di principi attivi farmaceutici, ed è sottoposta a linee guida regolatorie internazionali, come le linee guida definite dall'ICH (*International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use*) per la qualificazione ed il controllo delle impurezze (ICH Q3A e Q3B, per le impurezze nelle sostanze attive e prodotti farmaceutici, rispettivamente). Negli ultimi anni le autorità regolatorie hanno posto l'accento soprattutto sul problema delle impurezze potenzialmente genotossiche, introducendo una nuova linea guida (ICH M7) per l'identificazione, la classificazione, la qualificazione e il controllo delle impurezze genotossiche al fine di limitare il potenziale rischio cancerogeno dei farmaci. La linea guida ICH M7 fornisce raccomandazioni e linee guida per l'utilizzo delle metodologie (Q)SAR per la valutazione del potenziale genotossico delle impurezze, favorendo così l'utilizzo di metodi alternativi alla sperimentazione (*in vitro* e/o *in vivo*) nel contesto regolatorio dei farmaci. In particolare, la linea guida prevede che, in assenza di dati sperimentali di mutagenicità e/o carcinogenicità, necessari per la classificazione delle impurezze come mutagene e/o cancerogene (Classi 1, 2 e 5 - Tab. 1), la valutazione del potenziale mutageno possa essere effettuata mediante l'utilizzo di modelli (Q)SAR. In accordo con la linea guida, la valutazione computazionale deve essere effettuata usando due metodologie (Q)SAR com-



Classe	Definizione	Approccio	Azioni di controllo
1	Nota mutagena cancerogena	Valutazione dati sperimentali (database/letteratura)	Controllare/mantenere sotto limiti di accettabilità (specifici per l'impurezza)
2	Nota mutagena, no dati su cancerogenicità	Valutazione dati sperimentali (database/letteratura)	Controllare/mantenere sotto limiti di accettabilità (TTC)
3	No dati mutagenicità, predetto potenziale mutageno	Valutazione computazionale ((Q)SAR)	Controllare/mantenere sotto limiti di accettabilità (TTC) o fare un test di Ames
4	No dati mutagenicità, predetto potenziale mutageno ma alert strutturale condiviso con API o composti correlati all'API (es: intermedi), per i quali vi sono dati sperimentali e risultano non mutageni.	Valutazione computazionale ((Q)SAR)	Considerare come non-mutagena
5	Non-mutagena e/o non-cancerogena, oppure predetta come non-mutagena	Valutazione dati sperimentali e/o valutazione computazionale	Considerare come non-mutagena

Tab. 1 - Classificazione delle impurezze genotossiche e azioni di controllo (Tab. 1 ICH M7)

plementari, ossia un approccio su base statistica ed un sistema esperto. I modelli (Q)SAR impiegati nella valutazione computazionale devono essere conformi ai **principi definiti dall'OECD** (Organisation for Economic Co-operation and Development) per la validità e l'applicabilità di tali modelli per fini regolatori, ossia essere caratterizzati da:

- 1) un endpoint definito;
- 2) un algoritmo definito;
- 3) un definito dominio di applicabilità;
- 4) una validazione interna ed esterna;
- 5) un'interpretazione meccanicistica, se possibile.

Il ruolo dell'esperto è infine fondamentale per la revisione e valutazione dei risultati ottenuti mediante l'utilizzo di modelli (Q)SAR, al fine di supportare qualsiasi predizione positiva (mutageno) o negativa (non-mutageno), così come per fornire un'adeguata giustificazione in caso di risultati conflittuali o incerti [1].

Metodologie (Q)SAR: modelli statistici e sistemi esperti

Convenzionalmente si possono distinguere due differenti metodologie (Q)SAR: i modelli su base statistica ("*statistical-based methods*") e i sistemi esperti ("*expert rule-based methods*" o "*knowledge-based methods*").

I modelli statistici, propriamente definiti dall'acronimo QSAR, si basano su una relazione quantitativa (statistico-matematica) tra una rappresentazione numerica della struttura chimica (definita tramite descrittori molecolari) ed una specifica proprietà o attività biologica. Ad oggi sono migliaia i descrittori molecolari che sono stati sviluppati sulla base di diverse teorie ed approcci chemiometrici al fine di fornire una descrizione esaustiva della struttura chimica. Tali descrittori codificano proprietà strutturali mono-, bi- o tri-dimensionali delle molecole





(es: contatori di atomi, descrittori topologici, ecc.), proprietà chimico-fisiche (es: LogP) o proprietà elettroniche. Analogamente, molteplici sono gli algoritmi attualmente utilizzati per identificare la relazione quantitativa tra struttura e attività/proprietà, al fine di costruire dei modelli statisticamente robusti e predittivi (es: metodi di regressione multivariata (*Multiple Linear Regression* - MLR; *Partial Least Square Regression* - PLS), reti neurali (*Artificial Neural Networks* - ANN), ecc.). Ne deriva quindi che la maggior parte dei modelli su base statistica sia caratterizzata da robuste tecniche di validazione statistica e alte performance predittive (sebbene con una tendenza a decrescere all'aumentare della complessità della proprietà/attività predetta). I modelli statistici sono inoltre caratterizzati da un dominio di applicabilità ben definito (esistono ormai diversi metodi per la definizione del dominio di applicabilità e di stima del livello di estrapolazione di una predizione) e dalla possibilità di fornire delle predizioni anche quando il meccanismo d'azione non è noto a priori. Tuttavia, non è sempre possibile dare una chiara interpretazione delle predizioni generate da un modello QSAR, specialmente quando il modello utilizzato si basa su algoritmi e descrittori complessi, risultando così poco "trasparenti" all'utente finale.

I sistemi esperti si basano su un sistema di "regole" derivanti dalla conoscenza tossicologica esistente, caratterizzata da forti basi meccanicistiche (i.e., conoscenza del meccanismo d'azione), che viene utilizzata per predire un determinato effetto avverso. Nella maggior parte dei sistemi esperti, le regole (ossia i frammenti strutturali correlati ad un determinato effetto) vengono identificate da esperti tossicologi sulla base dell'analisi approfondita di serie di molecole "attive" per un determinato effetto che condividono uno stesso frammento ("*structural alert*" - SA). Più recentemente, si sono sviluppati algoritmi in grado di identificare automaticamente alert strutturali correlati ad un effetto/attività. I sistemi esperti possono essere basati su semplici regole (es: presenza/assenza di un alert) o su alberi decisionali più complessi. Il punto di forza dei sistemi esperti consiste nella possibilità di fornire agevolmente un'interpretazione meccanicistica della predizione (che viene spesso supportata da riferimenti bibliografici e una valutazione dell>alert identificato). Per tale motivo le

predizioni risultano spesso all'utente finale più trasparenti e facilmente interpretabili. Come i modelli statistici, anche i sistemi esperti presentano alcuni svantaggi, in quanto sono ad oggi limitate le misure per definirne un dominio di applicabilità [2], elemento essenziale per determinare l'attendibilità di una predizione, così come spesso non è possibile discriminare molecole attive e inattive che possiedono uno stesso alert strutturale, rendendo così i sistemi esperti dei metodi predittivi conservativi. Per quanto riguarda le performance predittive, l'accuratezza dei sistemi esperti è spesso paragonabile a quella dei modelli statistici, anche se i sistemi esperti tendono ad essere caratterizzati da una più bassa specificità (ossia comportano una più alta frequenza di falsi positivi) [3].

È importante ricordare che qualsiasi modello predittivo (inclusi i modelli (Q)SAR) è, per definizione, una simulazione della realtà, e, pertanto, non è mai accurato al 100%. Come discusso sopra, ciascun approccio computazionale, sia esso un modello statistico o sistema esperto, presenta dei punti di forza e dei limiti. Allo stesso modo, ciascun modello (Q)SAR è caratterizzato da specifiche capacità predittive (es., sensibilità VS specificità) ed un definito dominio di applicabilità (nessun modello può essere applicato all'intero universo chimico), fornendo così diverse "viste" parziali sulle proprietà/attività predette per una data molecola. Pertanto, l'approccio raccomandato per raccogliere al meglio le informazioni fornite da ciascun modello e ottenere predizioni più accurate consiste nel combinare le predizioni ottenute da diversi modelli e approcci computazionali (statistico ed esperto). Tale combinazione può essere fatta dall'esperto o mediante l'utilizzo di appropriati algoritmi [4]. È infatti dimostrato che l'uso complementare di approcci statistici ed esperti, supportati da una valutazione da parte dell'esperto, determina un incremento dell'accuratezza nelle predizioni [5, 6].

Per agevolare e incentivare l'utilizzo dei modelli (Q)SAR (in ambito regolatorio o per fini di ricerca e sviluppo), nel corso degli anni istituti di ricerca, università e aziende private si sono focalizzati sull'implementazione di tali modelli (spesso disponibili unicamente nella letteratura scientifica) in soluzioni software, rendendo così i modelli facilmente acces-

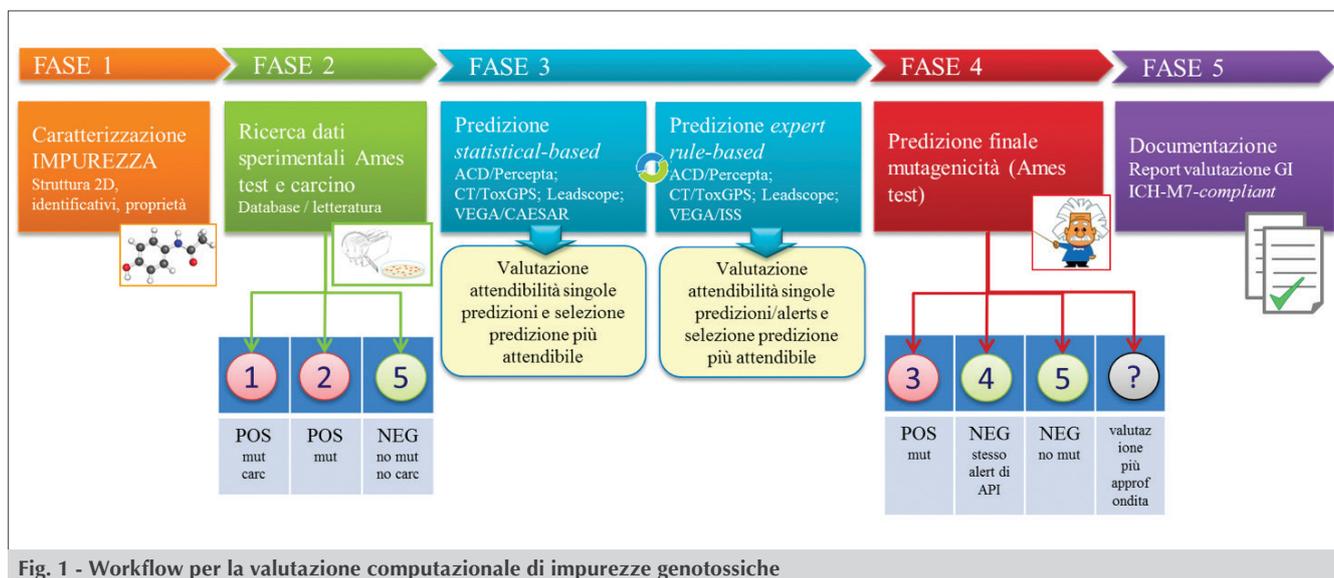


Fig. 1 - Workflow per la valutazione computazionale di impurezze genotossiche

sibili anche a utenti non esperti. Per quanto riguarda i modelli (Q)SAR per la predizione della mutagenicità, sono ad oggi disponibili numerosi software, sia commerciali che di pubblico dominio (per un maggiore approfondimento fare riferimento alle seguenti reviews [7-11]).

Workflow per la valutazione computazionale di impurezze genotossiche

Come descritto inizialmente, la linea guida ICH M7 fornisce delle indicazioni precise su come effettuare la valutazione e la classificazione delle impurezze genotossiche. Sulla base dell'esperienza acquisita nell'ambito dei numerosi servizi di consulenza per industrie farmaceutiche o produttori di principi attivi, descriviamo qui di seguito la procedura da noi utilizzata per la valutazione del potenziale mutageno delle impurezze in conformità con la linea guida ICH M7 [12]. La stessa procedura è illustrata in Fig. 1.

Fase 1

Caratterizzazione dell'impurezza target in termini di struttura chimica, identificativi (nome comune, nome IUPAC, CAS...), descrittori strutturali (SMILES, InChI), peso molecolare e formula molecolare.

Fase 2

Ricerca di eventuali dati sperimentali di mutagenicità (Ames test) e cancerogenicità pubblicamente

disponibili per l'impurezza (database riconosciuti o letteratura scientifica) mediante l'utilizzo di appositi software (commerciali - **ChemTunes**, e gratuiti - **OECD QSAR Toolbox**).

Fase 3

In assenza di dati sperimentali validi che consentano la classificazione dell'impurezza nelle classi 1, 2 o 5 (Tab. 1), viene effettuata la valutazione computazionale della potenziale mutagenicità (Ames test), che prevede l'utilizzo di un approccio predittivo sia su base statistica sia mediante sistemi esperti.

3a

Predizione *statistical-based*, che prevede l'utilizzo di una batteria di modelli QSAR, la valutazione dell'attendibilità delle singole predizioni e la successiva selezione della predizione più robusta per la specifica impurezza target. Tra i numerosi predittori disponibili, noi utilizziamo attualmente tre software commerciali (**ACD/Percepta**, **Leadscope** e **ChemTunes/ToxGPS**) e il modello CAESAR implementato nella piattaforma gratuita **VEGA**. Questi predittori sono tutti applicabili in ambito regolatorio in quanto conformi ai principi OECD, basati su dataset molto ampi, eterogenei e validati, caratterizzati da alte performance predittive e ampi domini di applicabilità. La valutazione dell'attendibilità delle predizioni viene effettuata prendendo in considerazione diversi aspetti, tra cui:



- i) se la molecola rientra nel dominio di applicabilità del modello;
 - ii) se e quanto la molecola è rappresentata nel training set, analizzando gli analoghi strutturali inclusi nel training set;
 - iii) la consistenza dei dati sperimentali degli analoghi, tra loro e rispetto alla predizione dell'impurezza;
 - iv) l'accuratezza in predizione degli analoghi.
- Infine, viene valutata la consistenza tra le predizioni ottenute dai diversi modelli QSAR utilizzati e selezionata la predizione ritenuta più attendibile e robusta per l'impurezza target.

3b

Predizione *expert rule-based*, che prevede l'utilizzo di più sistemi esperti, la valutazione dell'attendibilità delle singole predizioni (basate sull'identificazione o meno di alert strutturali), e la successiva selezione della predizione più robusta per la specifica impurezza target. Tra i diversi sistemi esperti disponibili per predire la mutagenicità, noi utilizziamo i tre software commerciali indicati sopra (**ACD/Percepta**, **Leadscope** e **ChemTunes/ToxGPS**), che includono entrambi gli approcci, e l'albero decisionale dell'Istituto Superiore di Sanità (che implementa le **regole di Benigni-Bossa** per la mutagenicità) incluso nella piattaforma gratuita di **VEGA**. I sistemi esperti selezionati sono stati sviluppati analizzando datasets molto ampi, si basano su numerosi alert strutturali validati e documentati nella letteratura scientifica, e forniscono informazioni utili per valutare la significatività degli alert identificati e, di conseguenza, l'attendibilità delle predizioni ottenute. Sebbene per i sistemi esperti non sia possibile definire un dominio di applicabilità definito, è comunque possibile valutare l'attendibilità delle predizioni da essi generati tenendo in considerazione i vari aspetti, tra cui:

- i) se l'impurezza target è rappresentata nel training set (o "alert reference set"), sia in termini di similarità strutturale, gruppi funzionali o altri frammenti strutturali;
- ii) rilevanza e/o significatività statistica degli alert identificati nell'impurezza (ossia, se l>alert ha una frequenza statisticamente significativa maggiore nei composti genotossici del training rispetto a quelli non-genotossici);

- iii) precisione dell>alert identificato nell'impurezza (es., basso tasso di falsi positivi);
- iv) consistenza tra i dati sperimentali degli analoghi che possiedono lo stesso alert dell'impurezza e la predizione dell'impurezza.

Particolare attenzione viene posta alle predizioni negative basate sull'assenza di alert strutturali. Infatti, l'assenza di un alert strutturale non è un'evidenza sufficiente per dimostrare l'assenza di un effetto mutageno, e resta la possibilità che l'impurezza possa agire mediante meccanismi d'azione non noti, per i quali non sono ancora stati sviluppati specifici alerts. Per tale motivo, diventa fondamentale un'analisi dettagliata degli analoghi strutturali che non possiedono alerts e l'accuratezza del sistema esperto nei confronti delle molecole non-mutagene che non possiedono alerts (sensibilità). Infine, viene valutata la consistenza tra le predizioni ottenute dai diversi sistemi esperti impiegati e selezionata la predizione ritenuta più attendibile e robusta per l'impurezza target.

Fase 4

Confronto delle predizioni ottenute dal modello statistico e dal sistema esperto e derivazione di una predizione finale di mutagenicità (i.e., positivo/negativo per Ames test) per l'impurezza target, con relativo livello di attendibilità. Questa rappresenta la fase più critica dell'intero processo, nella quale l'esperto ha un ruolo fondamentale nella valutazione dell'impurezza e la relativa classificazione. In base alle predizioni ottenute dai due approcci, e alle specifiche considerazioni che ne hanno determinato il livello di attendibilità, si possono aprire diversi scenari, tra cui:

- i) predizione negativa: l'impurezza viene predetta come negativa dall'approccio statistico con una buona attendibilità e non vengono identificati alert strutturali per la mutagenicità. L'impurezza può essere classificata in Classe 5 e considerata come non-mutagena;
- ii) predizione positiva: l'impurezza viene predetta come positiva dall'approccio statistico con una buona attendibilità e vengono identificati uno o più alert strutturali rilevanti per la mutagenicità. L'impurezza può essere classificata in Classe 3, considerata come mutagena e controllata di conseguenza (Tab. 1);

- iii) predizione positiva dovuta solamente ad un alert strutturale che è condiviso con l'API (o con un composto analogo correlato all'API, es: intermedio di processo) che è stato testato e risulta non-mutageno: in questo caso l'impurezza viene classificata in Classe 4 e può essere considerata come non-mutageno, previa dimostrazione che l'impurezza sia sufficientemente simile all'analogo non-mutageno, possieda lo stesso alert strutturale con lo stesso contorno chimico e che non ci siano ulteriori fattori (strutturali e non) che potrebbero indurre una potenziale mutagenicità per l'impurezza;
- iv) predizione incerta/non conclusiva: i diversi modelli e approcci utilizzati portano a conclusioni tra loro discordanti o poco attendibili, che preven- gono una chiara classificazione dell'impurezza come mutagena o non mutagena. In questo caso è necessario un ulteriore approfondimento da parte dell'esperto (es: mediante approfondimento in letteratura, analisi del meccanismo d'azione e confronto con strutture analoghe) per valutare se ci sono sufficienti evidenze che consentano di classificare l'impurezza come non-mutagena; in caso contrario è opportuno valutare la convenienza di effettuare un Ames test o classificare l'impurezza come mutagena (e controllarla come tale).

Fase 5

Documentazione dei risultati, che prevede la stesura di un report in cui vengono illustrati e discussi i metodi e strumenti impiegati nella valutazione computazionale, i risultati ottenuti dall'approccio statistico ed esperto, l'eventuale giustificazione aggiuntiva per una classificazione in Classe 4 e/o qualsiasi ulteriore approfondimento ad opera dell'esperto, e infine le conclusioni ottenute.

A prescindere dalle procedure messe in atto per adempiere alla linea guida ICH M7, appare evidente che tale linea guida apre importanti scenari per l'impiego delle metodologie (Q)SAR in ambito regolatorio per tutti quei settori in cui ad oggi ancora non è possibile prescindere dai test sperimentali (es: farmaci, pesticidi). La valutazione computazionale della potenziale genotossicità delle impurezze di

principi attivi farmaceutici (o ad esempio dei metaboliti di pesticidi) offre infatti notevoli vantaggi in termini di ottimizzazione dei tempi e costi rispetto alla sperimentazione in laboratorio, nonché rappresenta l'unica alternativa per poter valutare strutture anche solo potenzialmente presenti e non ancora sintetizzate.

BIBLIOGRAFIA

- [1] M.W. Powley, *Regul. Toxicol. Pharmacol.*, 2014, **71**, 295.
- [2] C.M. Ellison *et al.*, *Chem. Inf. Model.*, 2011, **51**, 975.
- [3] A. Hillebrecht *et al.*, *Chem. Res. Toxicol.*, 2011, **24**, 843.
- [4] C.M. Ellison *et al.*, *Mol. Inf.*, 2010, **26**, 97.
- [5] A. Sutter *et al.*, *Regul. Toxicol. Pharmacol.*, 2013, **67**, 39.
- [6] M. Nendza *et al.*, *Regul. Toxicol. Pharmacol.*, 2013, **66**, 301.
- [7] R. Serafimova *et al.*, JRC Scientific and Technical Reports, **EUR 24427 EN**, 2010.
- [8] A.P. Worth *et al.*, *Methods Mol. Biol.*, 2013, **930**, 125.
- [9] A. Roncaglioni *et al.*, *Curr. Opin. Pharmacol.*, 2013, **13**, 802.
- [10] E. Fioravanzo *et al.*, *SAR QSAR Environ. Res.*, 2012, **23**, 257.
- [11] A. Cassano *et al.*, *Environ. Sci. Health C Environ. Carcinog. Ecotoxicol. Rev.*, 2014, **32**, 273.
- [12] M. Pavan *et al.*, *Methods in Molecular Biology* (Chapter 21), Emilio Benfenati (Ed.), Springer Science+Business Media, New York, 2016, **1425**, 511.

QSAR Methods for Genotoxic Impurities

The ICH M7 guideline promotes the use of (Q)SAR methods for the evaluation of the mutagenic potential of the drug impurities. The computational assessment allows the screening of actual and potential impurities, optimizing time and costs of experimental testing.

