

Guida all'uso: La base di dati è costituita da tre file, che possono essere scaricati liberamente e separatamente: *ISSCAN_vvv_nnn_dddddd.xls*, *ISSCAN_vvv_nnn_dddddd.pdf*, *ISSCAN_vvv_nnn_dddddd.sdf*.

Il nome dei file contiene l'indicazione della versione, del numero di sostanze, e della data. Per esempio, *ISSCAN_v1a_774_10Dec04.xls* indica la versione v1a, con 774 sostanze, del 10 dicembre 2004.

Il file *ISSCAN_vvv_nnn_dddddd.xls* è leggibile con il programma Microsoft Excel. Per ogni sostanza, esso contiene il nome, vari sinonimi, il numero CAS, il peso molecolare, la formula bruta, la codifica SMILES della struttura, la mutagenicità in *Salmonella typhimurium* (test di Ames), la potenza cancerogena (TD₅₀), il potenziale cancerogeno nei quattro gruppi sperimentali di roditori comunemente usati (ratto, topo; maschio, femmina), il risultato e la classificazione dalla sperimentazione del National Toxicology Program (NTP) (se esistente), la referenza bibliografica per il dato di cancerogenesi.

Il file *ISSCAN_vvv_nnn_dddddd.pdf* è leggibile con il programma Adobe Acrobat Reader, e contiene le strutture chimiche in forma grafica.

Il file *ISSCAN_vvv_nnn_dddddd.sdf* è leggibile con programmi specializzati. In aggiunta alle informazioni contenute nel file *ISSCAN_v1a_774_10Dec04.xls*, contiene le formule chimiche in formato .sdf (structure-data file).

Per la sua particolare struttura, questa base di dati può essere usata a vari livelli e per scopi diversi.

Il livello di uso più semplice è la ricerca di informazioni nel file .xls, attraverso il nome (e sinonimi) ed il numero CAS. La struttura, se necessario, può essere reperita nel file .pdf secondo la comune convenzione grafica.

Una peculiarità di questa base di dati è la codifica delle strutture chimiche sia in forma SMILES che in forma .sdf. Queste codifiche sono lette da una vasta serie di programmi specializzati (per esempio, Chemoffice, Sybyl, Maestro, Insight II, Tsar, Daylight Toolkits, etc...). Essi permettono una amplissima gamma di operazioni, dalla visualizzazione in 3 dimensioni della molecola, fino al calcolo di descrittori e proprietà molecolari che possono essere usati in studi di relazioni quantitative tra struttura chimica ed attività biologica (Quantitative Structure-Activity Relationships, QSAR).

Particolarmente importante per il ruolo di questa base di dati come supporto alle decisioni di esperti è la possibilità di fare ricerche per sottostrutture chimiche o gruppi funzionali, e di leggerla come base di dati relazionale, con la possibilità quindi di collegare domande sulla struttura chimica a domande sull'informazione biologica. A tale scopo devono essere applicati al file .sdf programmi specializzati, di tipo Basi di Dati Chimiche Relazionali (Chemical Relational Database (CRD)). Tra tali programmi specializzati citiamo: Leadscope, Chemfolder .

Ulteriori informazioni sui programmi citati, sul concetto di Basi di Dati Chimiche Relazionali, sui tipi di file e sulla loro utilizzazione possono essere trovate nel sito del DSSTox (<http://www.epa.gov/nheerl/dsstox/>).

Nota: I programmi commerciali citati non devono in alcun modo essere considerati come i preferiti dai responsabili della presente base di dati. Vengono elencati per pura informazione ed a scopo esemplificativo, senza ordine di preferenze e senza pretesa di esaustività.

Definizione dei campi del file *ISSCAN_vvv_nnn_ddddddd.xls*

ID: Codice identificativo;

ChemName: Nome della sostanza;

Synonyms: Sinonimi chimici e nomi commerciali (derivati da Chemfinder <http://chemfinder.cambridgesoft.com/>);

CAS: Numero di Registro del Chemical Abstract Service;

Reference: Origine (compilazioni) dei dati di cancerogenesi. Sono i seguenti: CPDB (Carcinogenic Potency DataBase, <http://potency.berkeley.edu/cpdb.html>); Toxnet (base di dati CCRIS all'interno della base di dati tossicologica Toxnet, <http://toxnet.nlm.nih.gov/cgi-bin/sis/htmlgen?CCRIS>); NTP (National Toxicology Program, <http://ntp.niehs.nih.gov/>; e' indicato anche il numero del Technical Report); IARC (International Agency for Research on Cancer, <http://monographs.iarc.fr/>);

MolWeight: Peso Molecolare;

Formula: Formula bruta;

SMILES: SMILES e' una notazione chimica semplificata che permette di rappresentare una struttura chimica con una stringa di testo, per applicazioni al calcolatore (per ulteriori informazioni, vedi http://www.daylight.com/smiles/f_smiles.html);

TD50_Rat; TD50_Mouse: Valori di potenza cancerogena, rispettivamente nel ratto e nel topo. La TD₅₀ e' la dose giornaliera (mg/kg di peso corporeo/giorno) necessaria per dimezzare la probabilita' che un animale da esperimento rimanga senza tumori alla fine del suo periodo di vita. I valori presentati sono la media armonica delle TD₅₀ piu' potenti (valori piu' bassi) nei vari esperimenti in cui la sostanza e' risultata positiva. Tutti i valori presentati derivano dalla base di dati Carcinogenic Potency DataBase, <http://potency.berkeley.edu/cpdb.html>.

Per gli studi di struttura-attivita', la potenza va espressa in valori molari usando la trasformazione $\log_{10}(MW/TD_{50})$, dove MW e' il peso molecolare;

Canc: Dato riassuntivo di cancerogenesi: 3 = cancerogeno; 2 = risultato equivoco; 1 = non cancerogeno. Il valore 3 viene attribuito a sostanze che sono cancerogene in almeno 1 dei 4 sistemi sperimentali; il valore 2 a quelle che hanno almeno un risultato equivoco oltre a risultati negativi;

SAL: Mutagenicita' in *Salmonella typhimurium* (test di Ames): 3 = mutageno; 2 = risultato equivoco; 1 = non mutageno. Le sorgenti di dati sono le stesse citate in Reference, anche se per sostanze singole la fonte per i dati di cancerogenesi non necessariamente coincide con quella per i dati di mutagenesi;

Rat_Male_Canc; Rat_Female_Canc; Mouse_Male_Canc; Mouse_Female_Canc:

Risultati nei quattro gruppi sperimentali usati nei saggi a lungo termine di cancerogenesi (Ratto maschio, Ratto femmina, Topo maschio, Topo femmina). Codici: 3 = cancerogeno; 2 = risultato equivoco; 1 = non cancerogeno;

Rat_Male_NTP; Rat_Female_NTP; Mouse_Male_NTP; Mouse_Female_NTP:

Risultati, suddivisi per gruppi sperimentali, della sperimentazione dell' NTP (quando esistente). Codici: CE = Clear Evidence, Evidenza Chiara; SE = Some Evidence, Evidenza Limitata; EE = Equivocal Evidence, Evidenza Equivoca; NE = No Evidence, Nessuna Evidenza. Le quattro categorie di evidenza riportate sono quelle usate dall'NTP nei suoi studi (eccetto quelli relativi alla sperimentazione piu' vecchia) (vedi <http://ntp.niehs.nih.gov/>);

Codici generali: NP = non positivo; ND = dati non esistenti.

Caveat:

Le informazioni contenute in questa base di dati sono state tutte attentamente controllate, e, quando esistenti, piu' sorgenti di dati sono state confrontate. Tuttavia gli autori della base di dati non possono assicurare che essa sia esente da errori, e saranno lieti di accogliere suggerimenti e segnalazioni.

Per usi di particolare rilevanza e delicatezza (per esempio, per scopi regolamentativi), si invita a consultare la sorgente originaria dei dati, in quanto un risultato finale di cancerogenesi e' il riassunto di una mole notevole ed articolata di dati generati dall'esperimento a lungo termine di cancerogenesi nei roditori.

Ringraziamenti

Si ringraziano per i loro contributi allo sviluppo iniziale di questa base di dati: Maria Rosaria Vari', Laura Passerini, Anna Pino, Celestina D'Ascoli. Si ringraziano Ann M. Richard and Yang Chi-hae per gli utili commenti e discussioni.

Referenze principali:

Benigni,R.(Ed.) 2003. *Quantitative Structure-Activity Relationship (QSAR) models of mutagens and carcinogens*. CRC Press. Boca Raton.

Hansch,C., D.Hoekman, A.Leo, D.Weininger, and C.D.Selassie. 2002. "Chem-bioinformatics: comparative QSAR at the interface between chemistry and biology." *Chem.Revs.* 102:783-812.

Huff,J. 1999. "Value, validity, and historical development of carcinogenesis studies for predicting and confirming carcinogenic risks to humans." In K.T.Kitchin, editor, *Carcinogenicity. Testing, predicting, and interpreting chemical effects*. Marcel Dekker, Inc. New York. 21-123.

Tomatis,L. and J.Huff. 2001. "Evolution of cancer etiology and primary prevention." *Environ.Health Perspect.* 109:5-7.

Tomatis,L., J.Huff, I.Hertz-Picciotto, D.P.Sandler, J.Bucher, P.Boffetta, O.Axelson, A.Blair, J.Taylor, L.Stayner, and J.C.Barrett. 1997. "Avoided and avoidable risks of cancer." *Carcinogenesis.* 18:97-105.

Woo,Y.T., D.Y.Lai, J.L.McLain, M.Ko Manibusan, and V.Dellarco. 2002. "Use of mechanism-based structure-activity relationships analysis in carcinogenic potential ranking for drinking water disinfection by-products." *Environ.Health Perspect.* 110:75-87.