DATABASE

"CHEMICAL CARCINOGENS: STRUCTURES AND EXPERIMENTAL DATA"

Description: The present database contains information on chemical compounds tested with the long-term carcinogenicity bioassay on rodents (rat, mouse). Beside being a repository of data, it has been specifically designed as an expert decision support tool.

Historically, this database originates from the experience of researchers of the Environment and Primary Prevention Department in the field of structure-activity relationships, aimed at developing models which theoretically predict the carcinogenicity of chemicals. The use of experimental carcinogenicity data for structure-activity relationship studies amplifies their informative value, and contributes to the reduction and replacement of animal experimentation.

In its present form, this database is a collaboration between the Departments of Environment and Primary Prevention, and Drug Research and Evaluation of the Istituto Superiore di Sanita'.

This database does not contain epidemiological data, but only results from the carcinogenicity bioassay. A portion of the chemicals has been the subject of carcinogenicity classification by various Regulatory Agencies and Scientific Bodies; the classifications are available at the Carcinogenic Agents database (http://www.iss.it/sitp/spps/iscs/index.html).

The structure of this database is inspired by that of the Distributed Structure-Searchable Toxicity (DSSTox) Network of the US Enviromental Protection Agency (EPA) (http://www.epa.gov/nheerl/dsstox/). Similarly to the DSSTox spirit this project wants to contribute to the free diffusion of scientific data in a standardized, easy to read format.

Guidance for use: The database consists of three files, that can be downloaded separately: ***ISSCAN_vvv_nnn_ddddddd.xls***, ***ISSCAN_vvv_nnn_ddddddd.pdf***, ***ISSCAN_vvv_nnn_ddddddd.sdf***.

The file names provide information on the version number, the number of chemicals, and the date. For example, ***ISSCAN_v1a_774_10Dec04.xls*** is version v1a, contains 774 chemicals, and was prepared on December 10, 2004.

The file ***ISSCAN_vvv_nnn_ddddddd.xls*** can be read with the program Microsoft Excel. For each chemical, it reports: name, various synonyms, CAS number, molecular weight, chemical formula, SMILES code, mutagenicity in *Salmonella typhimurium* (Ames test), carcinogenic potency (TD_{50}), carcinogenic potential in the four experimental groups most commonly used (rat, mouse, male, female), carcinogenicity classification from the US National Toxicology Program (NTP) experiments (if available), reference for the carcinogenicity data.

The file ***ISSCAN_vvv_nnn_ddddddd.pdf*** can be read with the program Adobe Acrobat Reader, and reports the 2D chemical structure.

The file ***ISSCAN_vvv_nnn_ddddddd.sdf*** can be read with a series of specialized programs. In addition to the information reported in the file ***ISSCAN_vvv_nnn_ddddddd.xls***, it contains the chemical structures in the format .sdf (structure-data file).

This data base can be used for different purposes. As basic application, the .xls file can be searched through the name (and synonym), and the CAS number. The chemical structures, in the common graphical format can be found in the file .pdf.

A specific characteristic of this database is that the chemical structures are coded both in the SMILES and .sdf formats. These codes can be read by a series of specialized programs (e.g., Chemoffice, Sybyl, Maestro, Insight II, Tsar, Daylight Toolkits, etc…). Their capabilities range from e.g., 3D visualization, to the calculation of molecular properties and descriptors that can be used in the study of the relationships between chemical structure and biological activity (Quantitative Structure-Activity Relationships, QSAR).

Particularly important for the role of this database as decision support system, is the possibility of searching by substructures and functional groups, and of reading it as relational database. This permits the combination of chemical with biological interrogations. For this goal, the file .sdf has to be read with Chemical Relational Database (CRD) programs (e.g., Leadscope, Chemfolder).

More information on the above programs and on the concept of Chemical Relational Databases can be found in the DSSTox site(http://www.epa.gov/nheerl/dsstox/).

Important: The commercial programs cited above should not be considered as endorsed by this project team. They are listed only for information, in no preference order, and as examples of a wider range of programs.

Definition of the fields of the file *ISSCAN_vvv_nnn_ddddddd.xls*

ID: Code of the chemical;

ChemName: Chemical Name;

Synonyms: Chemical synonyms and commercial names (derived from Chemfinder http://chemfinder.cambridgesoft.com/);

CAS: Registry Number of the Chemical Abstract Service;

Reference: Source of carcinogenicity data: CPDB (Carcinogenic Potency DataBase, http://potency.berkeley.edu/cpdb.html); Toxnet (database CCRIS from the cluster of toxicological databases Toxnet, http://toxnet.nlm.nih.gov/cgi-bin/sis/htmlgen?CCRIS); NTP (National Toxicology Program, http://ntp.niehs.nih.gov/; the Technical Report number is also provided); IARC (International Agency for Research on Cancer, http://monographs.iarc.fr/);

MolWeight: Molecular Weight;

Formula: Chemical Formula;

SMILES: SMILES is a simplified chemical notation that represents a chemical structure as a linear textual string, for computer applications (for more information, http://www.daylight.com/smiles/f_smiles.html) ;

TD50_Rat; **TD50_Mouse**:   Carcinogenic potency in rat and mouse. TD_{50} is the rate in mg/kg body wt/day which, if administered chronically for the standard lifespan of the species, will halve the probability of remaining tumorless throughout that period. The TD_{50} value reported is the harmonic mean of the most potent TD_{50} values from each positive experiment in the species. All the values were derived from the Carcinogenic Potency DataBase, http://potency.berkeley.edu/cpdb.html.
Notice that for structure-activity studies, the potency should be transformed to molar values, and expressed as $\log_{10}(MW/TD_{50})$, where MW is the Molecular Weight;

Canc: Summary carcinogenicity data: 3 = carcinogen; 2 = equivocal; 1 = noncarcinogen. The code 3 is given to chemicals which were carcinogenic in at least one experimental group; the code 2 is given to chemicals which gave equivocal results in at least one experimental group, together with negative results in the other experimental groups;

SAL: Mutagenicity in *Salmonella typhimurium* (Ames test): 3 = mutagen; 2 = equivocal; 1 = nonmutagen. Overall, the sources of data are those quoted in Reference; however, for some chemicals the origin of carcinogenicity data may be different from the source of mutagenicity data;

Rat_Male_Canc; **Rat_Female_Canc**; **Mouse_Male_Canc**; **Mouse_Female_Canc**:
        Carcinogenicity results in the four experimental groups most commonly used for the cancer bioassay: 3 = carcinogen; 2 = equivocal; 1 = noncarcinogen;

Rat_Male_NTP; **Rat_Female_NTP**; **Mouse_Male_NTP**; **Mouse_Female_NTP**:
>Carcinogenicity results from the NTP experimentation (when available): CE = Clear Evidence; SE = Some Evidence; EE = Equivocal Evidence; NE = No Evidence. The four evidence categories are those used by NTP /(except in the older experimentation) (see http://ntp.niehs.nih.gov/);

General codes: NP = nonpositive; ND = no data.

Disclaimer and a cautionary word:

   The information contained in this database has been carefully checked, and, when existing, more sources have been considered and compared. However, the project team cannot guarantee accuracy, and will appreciate any suggestion to improve the database.
   For particularly significant uses (e.g., regulatory activity), the user is invited to examine the original source of data, and consider that the summary, final outcome of a carcinogenicity bioassay is a simplification of the large and articulated amount of experimental results provided by the bioassay.

Main references:

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.