ISSCAN database on chemical carcinogens

Definition of the fields in files *ISSCAN_vvv_nnn_ddddddd.xls* and *ISSCAN_vvv_nnn_ddddddd.sdf*

Substance ID: Identification Code of the chemical;

ChemName: Chemical Name of the tested form;

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

CAS: Registry Number of the Chemical Abstract Service of the parent form;

Structure: The displayed structure refers to the parent form;

CASTestedForm: Registry Number of the Chemical Abstract Service of the tested form. This field is given only when the tested form is not the parent compound. It relates to **ChemName**;

Route: Route of administration: DF = Dosed Feed; DW = Dosed Water; IN = Inhalation; GV = Gavage; DER = Dermal; IP = Intraperitoneal; SC = Subcutaneous; IV = Intravenous; GV-DF = Gavage followed by diet; Other = other routes of administration. The acronyms are preceded by the indication of the four bioassay experimental groups (Rat, Mouse, Male, Female), e.g., MF_DF = Mouse Female, Dosed Feed;

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/); **SOC** (Survey of Compounds, v4.0, NCI/NIEH);

MolWeight: Molecular Weight;

Formula: Chemical Formula;

SMILES: simplified chemical notation that represents a chemical structure as a linear textual string. It is aimed at computer applications (for more information, see 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.
For structure-activity studies, the potency should be transformed into 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. Code 3 is given to chemicals carcinogenic in at least one experimental group; Code 2 is given to chemicals with 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. Usually, the sources of data are those quoted in Reference for carcinogenicity data; however, for some chemicals the sources may be different;

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.