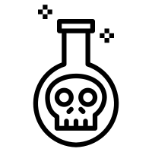
**Toxins and Toxin Target Database**

Source: Kaggle

Link: <https://www.kaggle.com/datasets/ahmedeltom/toxins-and-toxin-target-database>

Description:

The Toxin and Toxin Target Database (T3DB), or, soon to be referred as, the Toxic Exposome Database, is a unique bioinformatics resource that combines detailed toxin data with comprehensive toxin target information. The database currently houses 3,678 toxins described by 41,602 synonyms, including pollutants, pesticides, drugs, and food toxins, which are linked to 2,073 corresponding toxin target records.

The focus of the T3DB is on providing mechanisms of toxicity and target proteins for each toxin. This dual nature of the T3DB, in which toxin and toxin target records are interactively linked in both directions, makes it unique from existing databases. It is also fully searchable and supports extensive text, sequence, chemical structure, and relational query searches. It is both modelled after and closely linked to the Human Metabolome Database (HMDB) and DrugBank.

File format : CSV file .

Number of files: 4 csv files.

**1-all\_toxin\_data.csv**

Description:

All toxins , their description and chemical formula and pathways.

Number of rows: 3480

Number of columns: 44

|  |  |
| --- | --- |
| Title | The title of the toxins |
| Common\_name | Common name of the toxins |
| description | Description of the toxins |
| Cas | Chemical Abstracts Service (CAS) Number: A unique accession number assigned by the Chemical Abstracts Service, a division of the American Chemical Society. |
| pubchem\_id | pubChem is the world's largest collection of freely accessible chemical information |
| chemical\_formula | Chemical formula of the toxins |
| weight | the weight of the toxins |
| appearance | Outward or visible aspect of the toxin. |
| melting\_point | The melting point of the toxins. |
| boiling\_point | The boiling point of the toxins. |
| solubility | Solubility is the amount of solute that dissolves in a solvent |
| route\_of\_exposure | The way people [or other living organisms] come into contact with a hazardous substance(string) |
| mechanism\_of\_toxicity | Mechanism of toxicity is the study of how chemical or physical agents interact with living organisms that may trigger perturbations in cell function(string) |
| metabolism | he chemical reactions in the body's cells that change food into energy.(in this case toxins) |
| toxicity | is the quality or state of being toxic |
| lethaldose | A lethal dose is an amount of a drug or other agent that will kill a human or animal if administered. |
| carcinogenicity | chemical substance or a mixture of chemical substances which induce cancer or increase its incidence. |
| use\_source | Description of how humans commonly encounter the toxin - how the toxin is used or in what products they are found |
| min\_risk\_level | An estimate of the daily human exposure to a hazardous substance that is likely to be without appreciable risk of adverse noncancer health effects over a specified duration of exposure |
| health\_effects | The long term physiological deficiencies, damage or problems caused by the toxin |
| symptoms | The immediate and short term physiological effects of the toxin |
| created\_at | Creation date |
| Treatment | the management and care of a patient to combat disease or disorder. |
| symptoms | any feeling of illness or physical or mental change that is caused by a disease |
| updated\_at | Last update |
| wikipedia | Wikipedia link to get information about toxin |
| kegg\_compound\_id | Toxin id in kegg |
| chebi\_id | The Chemical Entities of Biological Interest identification number for the toxin |
| biocyc\_id | The BioCyc identification number for the toxin |
| ctd\_id | The Comparative Toxicogenomics Database identification number for the toxin |
| actor\_id | The Aggregated Computational Toxicology Resource identification number for the toxin |
| Stitch ID | The Search Tool for Interactions of Chemicals identification name to search with it |
| actor\_id | The Aggregated Computational Toxicology Resource identification number for the toxin |
| moldb\_smiles | Canonical SMILES string corresponding to toxin structure |
| moldb\_formula | Chemical formula describing atomic or elemental composition |
| moldb\_inchi | IUPAC approved InCHI identifier |
| moldb\_inchikey | refers to the InChIKey identifier for a specific molecule or chemical compound., which is a unique identifier for chemical substances that provides a standard and machine-readable representation of a molecule's structure. |
| moldb\_average\_mass | moldb\_average\_mass is a field that is commonly used in chemical databases and refers to the average mass of a specific molecule or chemical compound. The average mass is calculated as the sum of the masses of all atoms in a molecule, divided by the number of atoms in the molecule. |
| moldb\_mono\_mass | is a field that is commonly used in chemical databases and refers to the monoisotopic mass of a specific molecule or chemical compound. |
| State | The state of the toxin (gas , solid or liquid) |
| chemspider\_id | ID is a unique identifier for chemical compounds in the ChemSpider database, which is a free chemical structure database provided by the Royal Society of Chemistry. |
| synonyms\_list | synonyms\_list is a field that is commonly used in chemical databases to store a list of alternative names or synonyms for a specific chemical compound or molecule. Synonyms are names that refer to the same chemical compound, but may differ in spelling, pronunciation, or language |
| Types | The category of the toxin |
| cellular\_locations | Location of the toxin target inside or around the cell (cytoplasm, nucleus, membrane, etc.) |

**2-all\_toxin\_target.csv**

Description:

Substances that the toxin chemically binds to.

Number of rows: 2108

Number of columns: 23

|  |  |
| --- | --- |
| UniProt ID | Id of protein of uniport |
| Name | Name of the protein or macromolecule (or other small molecule) that the toxin is supposed to act upon. Some toxins act on multiple targets, so these fields may be repeated several times, reflecting the number of toxin targets a specific toxin may have |
| UniProt Name | Name of the organism |
| Organism | Organism type |
| NCBI Taxonomy Id | system for naming and organizing things, especially plants and animals ID in NCBI |
| Gene Name | Gene name in NCBI |
| Molecular Weight | Molecule weight of the protein |
| General Function | Function of the protein |
| Theoretical Pi | Theoretical isoelectric point of the toxin target |
| General Function | Short 3-4 word summary of the toxin target's primary functions |
| Specific Function | Detailed 30-40 word summary of the toxin target's specific functions |
| Pdb IDs | The Protein Data Bank identification number for the toxin |
| Genbank Gene ID | GenBank gene ID for the toxin target |
| Genbank Protein ID | GenBank protein ID for the toxin target (if it exists) |
| Chromosome Location | refers to the specific position or location of a gene or DNA sequence on a chromosome |
| locus | More detailed location of the chromosomal position of the toxin target |
| Genatlas ID | GenAtlas ID for the toxin target |
| HGNC ID | HGNC ID for the toxin target |
| KEGG ID | The Kyoto Encyclopedia of Genes and Genomes identification number for the toxin |
| Cellular Location | Location of the toxin target inside or around the cell (cytoplasm, nucleus, membrane, etc.) |
| Amino Acid Sequence | Sequence (protein or DNA) of the toxin target |
| Number of Residues | Number of amino acids (usually) in the toxin target |
| Gene Sequence | DNA sequence (from cDNA) of the toxin target |
| Synonyms | alternative names or terms that refer to the same chemical substance. |

3- Relation\_table1.csv

Description:

contains full records about each toxin and its targets with the mechanisms used and connect data with each other.

Number of rows: 31761

Number of columns: 11

|  |  |
| --- | --- |
| Toxin T3DB ID | Toxin T3DB ID refers to the unique identifier assigned to a specific toxin in the Toxin Target Database (T3DB). T3DB is a comprehensive database that provides information on the molecular and chemical properties, as well as the biological and toxicological effects, of a wide range of toxins. |
| Toxin Name | The Name of the toxin |
| Target Name | Name of the protein or macromolecule (or other small molecule) that the toxin is supposed to act upon. Some toxins act on multiple targets, so these fields may be repeated several times, reflecting the number of toxin targets a specific toxin may have. |
| Target BioDB ID | Target BioDB ID refers to the unique identifier assigned to a specific molecular target in the Target Central Resource Database (TCRD), also known as Target BioDB. |
| Target UniProt ID | Target UniProt ID refers to the unique identifier assigned to a specific protein target in the Universal Protein Resource (UniProt) database. |
| Mechanism of Action | Description of how the toxin works or what it binds to at a molecular level |
| References | Pubmed references for the toxin target |
| Creation Date | Date the ToxinCard entry was created |
| Update Date | Last update |

4- toxin\_structures.csv

Description:   
  
FASTA files Protein Sequences, Gene Sequences available in the link.

Number of rows: 3421

Number of columns: 19

|  |  |
| --- | --- |
| DATABASE\_ID | The title of the toxins |
| DATABASE\_NAME | Common name of the database |
| SMILES | Canonical SMILES string corresponding to toxin structure |
| INCHI\_IDENTIFIER | IUPAC approved InCHI identifier |
| InChIKey | (International Chemical Identifier Key) is a hashed version of the International Chemical Identifier (InChI) for a specific molecule or chemical compound |
| FORMULA | Chemical formula describing atomic or elemental composition |
| MOLECULAR\_WEIGHT | the weight of the toxins |
| EXACT\_MASS | refers to the precise mass of a specific molecule or chemical compound, expressed in atomic mass units (amu). The exact mass is calculated as the sum of the masses of all atoms in a molecule, taking into account the specific isotopes of the atoms in the molecule. |
| JCHEM\_AVERAGE\_POLARIZABILITY | is a field that is commonly used in chemical databases and refers to the average polarizability of a specific molecule or chemical compound. Polarizability is a measure of how easily the electron cloud of a molecule can be distorted by an external electric field |
| JCHEM\_IUPAC | is a field that is commonly used in chemical databases to store the International Union of Pure and Applied Chemistry (IUPAC) name of a specific molecule or chemical compound. |
| JCHEM\_LOGP | is a field that is commonly used in chemical databases to store the calculated algorithm of the partition coefficient of a specific molecule or chemical compound. The partition coefficient is a measure of the relative solubility of a compound in two immiscible liquids, such as water and octanol. |
| JCHEM\_REFRACTIVITY | is a field that is commonly used in chemical databases to store the calculated refractivity of a specific molecule or chemical compound. Refractivity is a measure of how much a molecule bends light as it passes through a medium, such as air or water. |
| JCHEM\_TRADITIONAL\_IUPAC | Is a field that is commonly used in chemical databases to store the traditional or common name of a specific molecule or chemical compound, according to the International Union of Pure and Applied Chemistry (IUPAC) nomenclature rules. |
| T3DB\_ID | T3DB\_ID refers to the unique identifier assigned to a specific toxin in the Toxin Target Database (T3DB). T3DB is a comprehensive database that provides information on the molecular and chemical properties, as well as the biological and toxicological effects, of a wide range of toxins. |
| NAME | is the name of toxin |
| CAS | Chemical Abstracts Service (CAS) Number: A unique accession number assigned by the Chemical Abstracts Service, a division of the American Chemical Society. |
| SYNONYMS | alternative names or terms that refer to the same chemical substance |
| TYPES | The category of the toxin |
| smiles | Canonical SMILES string corresponding to toxin structure. |

4- Mechanism.csv

Description:   
  
Contains all the toxin techniques to poison targets.

Number of rows: 43358

Number of columns: 9

|  |  |
| --- | --- |
| Id | The title of the toxins |
| Mechanism of action | Description of how the toxin works or what it binds to at a molecular level |
| References | Pubmed references for the toxin target |
| Creation Date | Date the ToxinCard entry was created |
| Update Date | Update Date |