**chembl\_25 Schema Documentation**

**ACTION\_TYPE:**

Table storing the distinct list of action types used in the drug\_mechanism table, together with a higher-level parent action type.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **KEYS** | **COLUMN\_NAME** | **DATA\_TYPE** | **NULLABLE** | **COMMENT** |
| PK | ACTION\_TYPE | VARCHAR2(50) | NOT NULL | Primary key. Type of action of the drug e.g., agonist, antagonist |
|  | DESCRIPTION | VARCHAR2(200) | NOT NULL | Description of how the action type is used |
|  | PARENT\_TYPE | VARCHAR2(50) |  | Higher-level grouping of action types e.g., positive vs negative action |

**ACTIVITIES:**

Activity 'values' or 'end points' that are the results of an assay recorded in a scientific document. Each activity is described by a row.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **KEYS** | **COLUMN\_NAME** | **DATA\_TYPE** | **NULLABLE** | **COMMENT** |
| PK | ACTIVITY\_ID | NUMBER(11,0) | NOT NULL | Unique ID for the activity row |
| FK | ASSAY\_ID | NUMBER(9,0) | NOT NULL | Foreign key to the assays table (containing the assay description) |
| FK | DOC\_ID | NUMBER(9,0) |  | Foreign key to documents table (for quick lookup of publication details - can also link to documents through compound\_records or assays table) |
| FK | RECORD\_ID | NUMBER(9,0) | NOT NULL | Foreign key to the compound\_records table (containing information on the compound tested) |
| FK | MOLREGNO | NUMBER(9,0) |  | Foreign key to compounds table (for quick lookup of compound structure - can also link to compounds through compound\_records table) |
|  | STANDARD\_RELATION | VARCHAR2(50) |  | Symbol constraining the activity value (e.g. >, <, =) |
|  | STANDARD\_VALUE | NUMBER |  | Same as PUBLISHED\_VALUE but transformed to common units: e.g. mM concentrations converted to nM. |
|  | STANDARD\_UNITS | VARCHAR2(100) |  | Selected 'Standard' units for data type: e.g. concentrations are in nM. |
|  | STANDARD\_FLAG | NUMBER(1,0) |  | Shows whether the standardised columns have been curated/set (1) or just default to the published data (0). |
|  | STANDARD\_TYPE | VARCHAR2(250) |  | Standardised version of the published\_activity\_type (e.g. IC50 rather than Ic-50/Ic50/ic50/ic-50) |
|  | ACTIVITY\_COMMENT | VARCHAR2(4000) |  | Previously used to report non-numeric activities i.e. 'Slighty active', 'Not determined'. STANDARD\_TEXT\_VALUE will be used for this in future, and this will be just for additional comments. |
| FK | DATA\_VALIDITY\_COMMENT | VARCHAR2(30) |  | Comment reflecting whether the values for this activity measurement are likely to be correct - one of 'Manually validated' (checked original paper and value is correct), 'Potential author error' (value looks incorrect but is as reported in the original paper), 'Outside typical range' (value seems too high/low to be correct e.g., negative IC50 value), 'Non standard unit type' (units look incorrect for this activity type). |
|  | POTENTIAL\_DUPLICATE | NUMBER(1,0) |  | When set to 1, indicates that the value is likely to be a repeat citation of a value reported in a previous ChEMBL paper, rather than a new, independent measurement. Note: value of zero does not guarantee that the measurement is novel/independent though |
|  | PCHEMBL\_VALUE | NUMBER(4,2) |  | Negative log of selected concentration-response activity values (IC50/EC50/XC50/AC50/Ki/Kd/Potency) |
| FK | BAO\_ENDPOINT | VARCHAR2(11) |  | ID for the corresponding result type in BioAssay Ontology (based on standard\_type) |
|  | UO\_UNITS | VARCHAR2(10) |  | ID for the corresponding unit in Unit Ontology (based on standard\_units) |
|  | QUDT\_UNITS | VARCHAR2(70) |  | ID for the corresponding unit in QUDT Ontology (based on standard\_units) |
|  | TOID | NUMBER |  | The Test Occasion Identifier, used to group together related activity measurements |
|  | UPPER\_VALUE | NUMBER |  | Where the activity is a range, this represents the highest value of the range (numerically), while the PUBLISHED\_VALUE column represents the lower value |
|  | STANDARD\_UPPER\_VALUE | NUMBER |  | Where the activity is a range, this represents the standardised version of the highest value of the range (with the lower value represented by STANDARD\_VALUE) |
| FK | SRC\_ID | NUMBER(3,0) |  | Foreign key to source table, indicating the source of the activity value |
|  | TYPE | VARCHAR2(250) | NOT NULL | Type of end-point measurement: e.g. IC50, LD50, %inhibition etc, as it appears in the original dataset |
|  | RELATION | VARCHAR2(50) |  | Symbol constraining the activity value (e.g. >, <, =), as it appears in the original dataset |
|  | VALUE | NUMBER |  | Datapoint value as it appears in the original dataset. |
|  | UNITS | VARCHAR2(100) |  | Units of measurement as they appear in the original dataset |
|  | TEXT\_VALUE | VARCHAR2(1000) |  | Non-numeric value for measurement as in original dataset |
|  | STANDARD\_TEXT\_VALUE | VARCHAR2(1000) |  | Standardized version of non-numeric measurement |

**ACTIVITY\_PROPERTIES:**

Table storing parameters and properties of Activity\_IDs in the ACTIVITIES table - can be either independent variables that do not apply to the whole assay (e.g., compounds 'DOSE'), or dependent variables/results that are important in interpreting the activity values (e.g., 'HILL\_SLOPE')

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **KEYS** | **COLUMN\_NAME** | **DATA\_TYPE** | **NULLABLE** | **COMMENT** |
| PK | AP\_ID | NUMBER(11,0) | NOT NULL | Unique ID for each record. |
| FK,UK | ACTIVITY\_ID | NUMBER(11,0) | NOT NULL | FK to ACTIVITY\_ID in ACTIVITIES table. |
| UK | TYPE | VARCHAR2(250) | NOT NULL | The parameter or property type |
|  | RELATION | VARCHAR2(50) |  | Symbol constraining the value (e.g. >, <, =) |
|  | VALUE | NUMBER |  | Numberical value for the parameter or property |
|  | UNITS | VARCHAR2(100) |  | Units of measurement |
|  | TEXT\_VALUE | VARCHAR2(1000) |  | Non-numerical value of the parameter or property |
|  | STANDARD\_TYPE | VARCHAR2(250) |  | Standardised form of the TYPE |
|  | STANDARD\_RELATION | VARCHAR2(50) |  | Standardised form of the RELATION |
|  | STANDARD\_VALUE | NUMBER |  | Standardised form of the VALUE |
|  | STANDARD\_UNITS | VARCHAR2(100) |  | Standardised form of the UNITS |
|  | STANDARD\_TEXT\_VALUE | VARCHAR2(1000) |  | Standardised form of the TEXT\_VALUE |
|  | COMMENTS | VARCHAR2(4000) |  | A Comment. |
|  | RESULT\_FLAG | NUMBER(1,0) | NOT NULL | A flag to indicate, if set to 1, that this type is a dependent variable/result (e.g., slope) rather than an independent variable/parameter (0, the default). |

**ACTIVITY\_SMID:**

A join table between ACTIVITY\_SUPP\_MAP and ACTIVITY\_SUPP

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **KEYS** | **COLUMN\_NAME** | **DATA\_TYPE** | **NULLABLE** | **COMMENT** |
| PK | SMID | NUMBER(11,0) | NOT NULL | FK to SMID in ACTIVITY\_SUPP\_MAP, and a FK to SMID in ACTIVITY\_SUPP |

**ACTIVITY\_STDS\_LOOKUP:**

Table storing details of standardised activity types and units, with permitted value ranges. Used to determine the correct standard\_type and standard\_units for a range of published types/units.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **KEYS** | **COLUMN\_NAME** | **DATA\_TYPE** | **NULLABLE** | **COMMENT** |
| PK | STD\_ACT\_ID | NUMBER(9,0) | NOT NULL | Primary key. |
| UK | STANDARD\_TYPE | VARCHAR2(250) | NOT NULL | The standard\_type that other published\_types in the activities table have been converted to. |
|  | DEFINITION | VARCHAR2(500) |  | A description/definition of the standard\_type. |
| UK | STANDARD\_UNITS | VARCHAR2(100) | NOT NULL | The units that are applied to this standard\_type and to which other published\_units are converted. Note a standard\_type may have more than one allowable standard\_unit and therefore multiple rows in this table. |
|  | NORMAL\_RANGE\_MIN | NUMBER(24,12) |  | The lowest value for this activity type that is likely to be genuine. This is only an approximation, so lower genuine values may exist, but it may be desirable to validate these before using them. For a given standard\_type/units, values in the activities table below this threshold are flagged with a data\_validity\_comment of 'Outside typical range'. |
|  | NORMAL\_RANGE\_MAX | NUMBER(24,12) |  | The highest value for this activity type that is likely to be genuine. This is only an approximation, so higher genuine values may exist, but it may be desirable to validate these before using them. For a given standard\_type/units, values in the activities table above this threshold are flagged with a data\_validity\_comment of 'Outside typical range'. |

**ACTIVITY\_SUPP:**

Supplementary / Supporting Data for ACTIVITIES - can be linked via ACTIVITY\_SMID and ACTIVITY\_SUPP\_MAP tables

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **KEYS** | **COLUMN\_NAME** | **DATA\_TYPE** | **NULLABLE** | **COMMENT** |
| PK | AS\_ID | NUMBER(11,0) | NOT NULL | Unique ID for each record. |
| UK | RGID | NUMBER(11,0) | NOT NULL | Record Grouping ID, used to group together related data points in this table |
| FK | SMID | NUMBER(11,0) |  | FK to SMID in ACTIVITY\_SMID. |
| UK | TYPE | VARCHAR2(250) | NOT NULL | Type of end-point measurement: e.g. IC50, LD50, %inhibition etc, as it appears in the original dataset |
|  | RELATION | VARCHAR2(50) |  | Symbol constraining the activity value (e.g. >, <, =), as it appears in the original dataset |
|  | VALUE | NUMBER |  | Datapoint value as it appears in the original dataset. |
|  | UNITS | VARCHAR2(100) |  | Units of measurement as they appear in the original dataset |
|  | TEXT\_VALUE | VARCHAR2(1000) |  | Non-numeric value for measurement as in original dataset |
|  | STANDARD\_TYPE | VARCHAR2(250) |  | Standardised form of the TYPE |
|  | STANDARD\_RELATION | VARCHAR2(50) |  | Standardised form of the RELATION |
|  | STANDARD\_VALUE | NUMBER |  | Standardised form of the VALUE |
|  | STANDARD\_UNITS | VARCHAR2(100) |  | Standardised form of the UNITS |
|  | STANDARD\_TEXT\_VALUE | VARCHAR2(1000) |  | Standardised form of the TEXT\_VALUE |
|  | COMMENTS | VARCHAR2(4000) |  | A Comment. |

**ACTIVITY\_SUPP\_MAP:**

Mapping table, linking supplementary / supporting data from the ACTIVITY\_SUPP table to the main ACTIVITIES table

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **KEYS** | **COLUMN\_NAME** | **DATA\_TYPE** | **NULLABLE** | **COMMENT** |
| PK | ACTSM\_ID | NUMBER(11,0) | NOT NULL | Primary key |
| FK,UK | ACTIVITY\_ID | NUMBER(11,0) | NOT NULL | FK to ACTIVITY\_ID in ACTIVITIES table. |
| FK,UK | SMID | NUMBER(11,0) | NOT NULL | FK to SMID in ACTIVITY\_SMID. |

**ASSAY\_CLASS\_MAP:**

Mapping table linking assays to classes in the ASSAY\_CLASSIFICATION table

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **KEYS** | **COLUMN\_NAME** | **DATA\_TYPE** | **NULLABLE** | **COMMENT** |
| PK | ASS\_CLS\_MAP\_ID | NUMBER | NOT NULL | Primary key. |
| FK,UK | ASSAY\_ID | NUMBER |  | assay\_id is the foreign key that maps to the 'assays' table |
| FK,UK | ASSAY\_CLASS\_ID | NUMBER |  | assay\_class\_id is the foreign key that maps to the 'assay\_classification' table |

**ASSAY\_CLASSIFICATION:**

Classification scheme for phenotypic assays e.g., by therapeutic area, phenotype/process and assay type. Can be used to find standard assays for a particular disease area or phenotype e.g., anti-obesity assays. Currently data are available only for in vivo efficacy assays

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **KEYS** | **COLUMN\_NAME** | **DATA\_TYPE** | **NULLABLE** | **COMMENT** |
| PK | ASSAY\_CLASS\_ID | NUMBER(9,0) | NOT NULL | Primary key |
|  | L1 | VARCHAR2(100) |  | High level classification e.g., by anatomical/therapeutic area |
|  | L2 | VARCHAR2(100) |  | Mid-level classification e.g., by phenotype/biological process |
| UK | L3 | VARCHAR2(1000) |  | Fine-grained classification e.g., by assay type |
|  | CLASS\_TYPE | VARCHAR2(50) |  | The type of assay being classified e.g., in vivo efficacy |
|  | SOURCE | VARCHAR2(50) |  | Source from which the assay class was obtained |

**ASSAY\_PARAMETERS:**

Table storing additional parameters for an assay e.g., dose, administration route, time point

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **KEYS** | **COLUMN\_NAME** | **DATA\_TYPE** | **NULLABLE** | **COMMENT** |
| PK | ASSAY\_PARAM\_ID | NUMBER(9,0) | NOT NULL | Numeric primary key |
| FK,UK | ASSAY\_ID | NUMBER(9,0) | NOT NULL | Foreign key to assays table. The assay to which this parameter belongs |
| UK | TYPE | VARCHAR2(250) | NOT NULL | The type of parameter being described, according to the original data source |
|  | RELATION | VARCHAR2(50) |  | The relation symbol for the parameter being described, according to the original data source |
|  | VALUE | NUMBER |  | The value of the parameter being described, according to the original data source. Used for numeric data |
|  | UNITS | VARCHAR2(100) |  | The units for the parameter being described, according to the original data source |
|  | TEXT\_VALUE | VARCHAR2(4000) |  | The text value of the parameter being described, according to the original data source. Used for non-numeric/qualitative data |
|  | STANDARD\_TYPE | VARCHAR2(250) |  | Standardized form of the TYPE |
|  | STANDARD\_RELATION | VARCHAR2(50) |  | Standardized form of the RELATION |
|  | STANDARD\_VALUE | NUMBER |  | Standardized form of the VALUE |
|  | STANDARD\_UNITS | VARCHAR2(100) |  | Standardized form of the UNITS |
|  | STANDARD\_TEXT\_VALUE | VARCHAR2(4000) |  | Standardized form of the TEXT\_VALUE |
|  | COMMENTS | VARCHAR2(4000) |  | Additional comments describing the parameter |

**ASSAY\_TYPE:**

Description of assay types (e.g., Binding, Functional, ADMET)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **KEYS** | **COLUMN\_NAME** | **DATA\_TYPE** | **NULLABLE** | **COMMENT** |
| PK | ASSAY\_TYPE | VARCHAR2(1) | NOT NULL | Single character representing assay type |
|  | ASSAY\_DESC | VARCHAR2(250) |  | Description of assay type |

**ASSAYS:**

Table storing a list of the assays that are reported in each document. Similar assays from different publications will appear as distinct assays in this table.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **KEYS** | **COLUMN\_NAME** | **DATA\_TYPE** | **NULLABLE** | **COMMENT** |
| PK | ASSAY\_ID | NUMBER(9,0) | NOT NULL | Unique ID for the assay |
| FK | DOC\_ID | NUMBER(9,0) | NOT NULL | Foreign key to documents table |
|  | DESCRIPTION | VARCHAR2(4000) |  | Description of the reported assay |
| FK | ASSAY\_TYPE | VARCHAR2(1) |  | Assay classification, e.g. B=Binding assay, A=ADME assay, F=Functional assay |
|  | ASSAY\_TEST\_TYPE | VARCHAR2(20) |  | Type of assay system (i.e., in vivo or in vitro) |
|  | ASSAY\_CATEGORY | VARCHAR2(20) |  | screening, confirmatory (ie: dose-response), summary, panel or other. |
|  | ASSAY\_ORGANISM | VARCHAR2(250) |  | Name of the organism for the assay system (e.g., the organism, tissue or cell line in which an assay was performed). May differ from the target organism (e.g., for a human protein expressed in non-human cells, or pathogen-infected human cells). |
|  | ASSAY\_TAX\_ID | NUMBER(11,0) |  | NCBI tax ID for the assay organism. |
|  | ASSAY\_STRAIN | VARCHAR2(200) |  | Name of specific strain of the assay organism used (where known) |
|  | ASSAY\_TISSUE | VARCHAR2(100) |  | Name of tissue used in the assay system (e.g., for tissue-based assays) or from which the assay system was derived (e.g., for cell/subcellular fraction-based assays). |
|  | ASSAY\_CELL\_TYPE | VARCHAR2(100) |  | Name of cell type or cell line used in the assay system (e.g., for cell-based assays). |
|  | ASSAY\_SUBCELLULAR\_FRACTION | VARCHAR2(100) |  | Name of subcellular fraction used in the assay system (e.g., microsomes, mitochondria). |
| FK | TID | NUMBER(9,0) |  | Target identifier to which this assay has been mapped. Foreign key to target\_dictionary. From ChEMBL\_15 onwards, an assay will have only a single target assigned. |
| FK | RELATIONSHIP\_TYPE | VARCHAR2(1) |  | Flag indicating of the relationship between the reported target in the source document and the assigned target from TARGET\_DICTIONARY. Foreign key to RELATIONSHIP\_TYPE table. |
| FK | CONFIDENCE\_SCORE | NUMBER(1,0) |  | Confidence score, indicating how accurately the assigned target(s) represents the actually assay target. Foreign key to CONFIDENCE\_SCORE table. 0 means uncurated/unassigned, 1 = low confidence to 9 = high confidence. |
| FK | CURATED\_BY | VARCHAR2(32) |  | Indicates the level of curation of the target assignment. Foreign key to curation\_lookup table. |
| FK | SRC\_ID | NUMBER(3,0) | NOT NULL | Foreign key to source table |
|  | SRC\_ASSAY\_ID | VARCHAR2(50) |  | Identifier for the assay in the source database/deposition (e.g., pubchem AID) |
| FK,UK | CHEMBL\_ID | VARCHAR2(20) | NOT NULL | ChEMBL identifier for this assay (for use on web interface etc) |
| FK | CELL\_ID | NUMBER(9,0) |  | Foreign key to cell dictionary. The cell type or cell line used in the assay |
| FK | BAO\_FORMAT | VARCHAR2(11) |  | ID for the corresponding format type in BioAssay Ontology (e.g., cell-based, biochemical, organism-based etc) |
| FK | TISSUE\_ID | NUMBER(9,0) |  | ID for the corresponding tissue/anatomy in Uberon. Foreign key to tissue\_dictionary |
| FK | VARIANT\_ID | NUMBER(9,0) |  | Foreign key to variant\_sequences table. Indicates the mutant/variant version of the target used in the assay (where known/applicable) |
|  | AIDX | VARCHAR2(200) | NOT NULL | The Depositor Defined Assay Identifier |

**ATC\_CLASSIFICATION:**

WHO ATC Classification for drugs

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **KEYS** | **COLUMN\_NAME** | **DATA\_TYPE** | **NULLABLE** | **COMMENT** |
|  | WHO\_NAME | VARCHAR2(2000) |  | WHO/INN name for the compound |
|  | LEVEL1 | VARCHAR2(10) |  | First level of classification |
|  | LEVEL2 | VARCHAR2(10) |  | Second level of classification |
|  | LEVEL3 | VARCHAR2(10) |  | Third level of classification |
|  | LEVEL4 | VARCHAR2(10) |  | Fourth level of classification |
| PK | LEVEL5 | VARCHAR2(10) | NOT NULL | Complete ATC code for compound |
|  | LEVEL1\_DESCRIPTION | VARCHAR2(2000) |  | Description of first level of classification |
|  | LEVEL2\_DESCRIPTION | VARCHAR2(2000) |  | Description of second level of classification |
|  | LEVEL3\_DESCRIPTION | VARCHAR2(2000) |  | Description of third level of classification |
|  | LEVEL4\_DESCRIPTION | VARCHAR2(2000) |  | Description of fourth level of classification |

**BINDING\_SITES:**

Table storing details of binding sites for a target. A target may have multiple sites defined in this table.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **KEYS** | **COLUMN\_NAME** | **DATA\_TYPE** | **NULLABLE** | **COMMENT** |
| PK | SITE\_ID | NUMBER(9,0) | NOT NULL | Primary key. Unique identifier for a binding site in a given target. |
|  | SITE\_NAME | VARCHAR2(200) |  | Name/label for the binding site. |
| FK | TID | NUMBER(9,0) |  | Foreign key to target\_dictionary. Target on which the binding site is found. |

**BIO\_COMPONENT\_SEQUENCES:**

Table storing the sequences for biotherapeutic drugs (e.g., monoclonal antibodies, recombinant proteins, peptides etc. For multi-chain biotherapeutics (e.g., mAbs) each chain is stored here as a separate component.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **KEYS** | **COLUMN\_NAME** | **DATA\_TYPE** | **NULLABLE** | **COMMENT** |
| PK | COMPONENT\_ID | NUMBER(9,0) | NOT NULL | Primary key. Unique identifier for each of the molecular components of biotherapeutics in ChEMBL (e.g., antibody chains, recombinant proteins, synthetic peptides). |
|  | COMPONENT\_TYPE | VARCHAR2(50) | NOT NULL | Type of molecular component (e.g., 'PROTEIN','DNA','RNA'). |
|  | DESCRIPTION | VARCHAR2(200) |  | Description/name of molecular component. |
|  | SEQUENCE | CLOB |  | Sequence of the biotherapeutic component. |
|  | SEQUENCE\_MD5SUM | VARCHAR2(32) |  | MD5 checksum of the sequence. |
|  | TAX\_ID | NUMBER(11,0) |  | NCBI tax ID for the species from which the sequence is derived. May be null for humanized monoclonal antibodies, synthetic peptides etc. |
|  | ORGANISM | VARCHAR2(150) |  | Name of the species from which the sequence is derived. |

**BIOASSAY\_ONTOLOGY:**

Lookup table providing labels for Bioassay Ontology terms used in assays and activities tables

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **KEYS** | **COLUMN\_NAME** | **DATA\_TYPE** | **NULLABLE** | **COMMENT** |
| PK | BAO\_ID | VARCHAR2(11) | NOT NULL | Bioassay Ontology identifier (BAO version 2.0) |
|  | LABEL | VARCHAR2(100) | NOT NULL | Bioassay Ontology label for the term (BAO version 2.0) |

**BIOTHERAPEUTIC\_COMPONENTS:**

Links each biotherapeutic drug (in the biotherapeutics table) to its component sequences (in the bio\_component\_sequences table). A biotherapeutic drug can have multiple components and hence multiple rows in this table. Similarly, a particular component sequence can be part of more than one drug.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **KEYS** | **COLUMN\_NAME** | **DATA\_TYPE** | **NULLABLE** | **COMMENT** |
| PK | BIOCOMP\_ID | NUMBER(9,0) | NOT NULL | Primary key. |
| FK,UK | MOLREGNO | NUMBER(9,0) | NOT NULL | Foreign key to the biotherapeutics table, indicating which biotherapeutic the component is part of. |
| FK,UK | COMPONENT\_ID | NUMBER(9,0) | NOT NULL | Foreign key to the bio\_component\_sequences table, indicating which component is part of the biotherapeutic. |

**BIOTHERAPEUTICS:**

Stores sequence information/descriptions for protein therapeutics, including recombinant proteins, peptides and antibodies

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **KEYS** | **COLUMN\_NAME** | **DATA\_TYPE** | **NULLABLE** | **COMMENT** |
| PK,FK | MOLREGNO | NUMBER(9,0) | NOT NULL | Foreign key to molecule\_dictionary |
|  | DESCRIPTION | VARCHAR2(2000) |  | Description of the biotherapeutic. |
|  | HELM\_NOTATION | VARCHAR2(4000) |  | Sequence notation generated according to the HELM standard (http://www.openhelm.org/home). Currently for peptides only |

**CELL\_DICTIONARY:**

Table storing information for cell lines in the target\_dictionary (e.g., tissue/species origin). Cell\_name should match pref\_name in target\_dictionary.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **KEYS** | **COLUMN\_NAME** | **DATA\_TYPE** | **NULLABLE** | **COMMENT** |
| PK | CELL\_ID | NUMBER(9,0) | NOT NULL | Primary key. Unique identifier for each cell line in the target\_dictionary. |
| UK | CELL\_NAME | VARCHAR2(50) | NOT NULL | Name of each cell line (as used in the target\_dicitonary pref\_name). |
|  | CELL\_DESCRIPTION | VARCHAR2(200) |  | Longer description (where available) of the cell line. |
|  | CELL\_SOURCE\_TISSUE | VARCHAR2(50) |  | Tissue from which the cell line is derived, where known. |
|  | CELL\_SOURCE\_ORGANISM | VARCHAR2(150) |  | Name of organism from which the cell line is derived. |
| UK | CELL\_SOURCE\_TAX\_ID | NUMBER(11,0) |  | NCBI tax ID of the organism from which the cell line is derived. |
|  | CLO\_ID | VARCHAR2(11) |  | ID for the corresponding cell line in Cell Line Ontology |
|  | EFO\_ID | VARCHAR2(12) |  | ID for the corresponding cell line in Experimental Factory Ontology |
|  | CELLOSAURUS\_ID | VARCHAR2(15) |  | ID for the corresponding cell line in Cellosaurus Ontology |
|  | CL\_LINCS\_ID | VARCHAR2(8) |  | Cell ID used in LINCS (Library of Integrated Network-based Cellular Signatures) |
| FK,UK | CHEMBL\_ID | VARCHAR2(20) |  | ChEMBL identifier for the cell (used in web interface etc) |

**CHEMBL\_ID\_LOOKUP:**

Lookup table storing chembl identifiers for different entities in the database (assays, compounds, documents and targets)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **KEYS** | **COLUMN\_NAME** | **DATA\_TYPE** | **NULLABLE** | **COMMENT** |
| PK | CHEMBL\_ID | VARCHAR2(20) | NOT NULL | ChEMBL identifier |
| UK | ENTITY\_TYPE | VARCHAR2(50) | NOT NULL | Type of entity (e.g., COMPOUND, ASSAY, TARGET) |
| UK | ENTITY\_ID | NUMBER(9,0) | NOT NULL | Primary key for that entity in corresponding table (e.g., molregno for compounds, tid for targets) |
|  | STATUS | VARCHAR2(10) | NOT NULL | Indicates whether the status of the entity within the database - ACTIVE, INACTIVE (downgraded), OBS (obsolete/removed). |

**COMPONENT\_CLASS:**

Links protein components of targets to the protein\_family\_classification table. A protein can have more than one classification (e.g., Membrane receptor and Enzyme).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **KEYS** | **COLUMN\_NAME** | **DATA\_TYPE** | **NULLABLE** | **COMMENT** |
| FK,UK | COMPONENT\_ID | NUMBER(9,0) | NOT NULL | Foreign key to component\_sequences table. |
| FK,UK | PROTEIN\_CLASS\_ID | NUMBER(9,0) | NOT NULL | Foreign key to the protein\_classification table. |
| PK | COMP\_CLASS\_ID | NUMBER(9,0) | NOT NULL | Primary key. |

**COMPONENT\_DOMAINS:**

Links protein components of targets to the structural domains they contain (from the domains table). Contains information showing the start and end position of the domain in the component sequence.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **KEYS** | **COLUMN\_NAME** | **DATA\_TYPE** | **NULLABLE** | **COMMENT** |
| PK | COMPD\_ID | NUMBER(9,0) | NOT NULL | Primary key. |
| FK,UK | DOMAIN\_ID | NUMBER(9,0) |  | Foreign key to the domains table, indicating the domain that is contained in the associated molecular component. |
| FK,UK | COMPONENT\_ID | NUMBER(9,0) | NOT NULL | Foreign key to the component\_sequences table, indicating the molecular\_component that has the given domain. |
| UK | START\_POSITION | NUMBER(5,0) |  | Start position of the domain within the sequence given in the component\_sequences table. |
|  | END\_POSITION | NUMBER(5,0) |  | End position of the domain within the sequence given in the component\_sequences table. |

**COMPONENT\_GO:**

Table mapping protein components in the COMPONENT\_SEQUENCES table to the GO slim terms stored in the GO\_CLASSIFICATION table

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **KEYS** | **COLUMN\_NAME** | **DATA\_TYPE** | **NULLABLE** | **COMMENT** |
| PK | COMP\_GO\_ID | NUMBER(9,0) | NOT NULL | Primary key |
| FK,UK | COMPONENT\_ID | NUMBER(9,0) | NOT NULL | Foreign key to COMPONENT\_SEQUENCES table. The protein component this GO term applies to |
| FK,UK | GO\_ID | VARCHAR2(10) | NOT NULL | Foreign key to the GO\_CLASSIFICATION table. The GO term that this protein is mapped to |

**COMPONENT\_SEQUENCES:**

Table storing the sequences for components of molecular targets (e.g., protein sequences), along with other details taken from sequence databases (e.g., names, accessions). Single protein targets will have a single protein component in this table, whereas protein complexes/protein families will have multiple protein components.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **KEYS** | **COLUMN\_NAME** | **DATA\_TYPE** | **NULLABLE** | **COMMENT** |
| PK | COMPONENT\_ID | NUMBER(9,0) | NOT NULL | Primary key. Unique identifier for the component. |
|  | COMPONENT\_TYPE | VARCHAR2(50) |  | Type of molecular component represented (e.g., 'PROTEIN','DNA','RNA'). |
| UK | ACCESSION | VARCHAR2(25) |  | Accession for the sequence in the source database from which it was taken (e.g., UniProt accession for proteins). |
|  | SEQUENCE | CLOB |  | A representative sequence for the molecular component, as given in the source sequence database (not necessarily the exact sequence used in the assay). |
|  | SEQUENCE\_MD5SUM | VARCHAR2(32) |  | MD5 checksum of the sequence. |
|  | DESCRIPTION | VARCHAR2(200) |  | Description/name for the molecular component, usually taken from the source sequence database. |
|  | TAX\_ID | NUMBER(11,0) |  | NCBI tax ID for the sequence in the source database (i.e., species that the protein/nucleic acid sequence comes from). |
|  | ORGANISM | VARCHAR2(150) |  | Name of the organism the sequence comes from. |
|  | DB\_SOURCE | VARCHAR2(25) |  | The name of the source sequence database from which sequences/accessions are taken. For UniProt proteins, this field indicates whether the sequence is from SWISS-PROT or TREMBL. |
|  | DB\_VERSION | VARCHAR2(10) |  | The version of the source sequence database from which sequences/accession were last updated. |

**COMPONENT\_SYNONYMS:**

Table storing synonyms for the components of molecular targets (e.g., names, acronyms, gene symbols etc.) Please note: EC numbers are also currently included in this table although they are not strictly synonyms and can apply to multiple proteins.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **KEYS** | **COLUMN\_NAME** | **DATA\_TYPE** | **NULLABLE** | **COMMENT** |
| PK | COMPSYN\_ID | NUMBER(9,0) | NOT NULL | Primary key. |
| FK,UK | COMPONENT\_ID | NUMBER(9,0) | NOT NULL | Foreign key to the component\_sequences table. The component to which this synonym applies. |
| UK | COMPONENT\_SYNONYM | VARCHAR2(500) |  | The synonym for the component. |
| UK | SYN\_TYPE | VARCHAR2(20) |  | The type or origin of the synonym (e.g., GENE\_SYMBOL). |

**COMPOUND\_PROPERTIES:**

Table storing calculated physicochemical properties for compounds, now calculated with RDKit and ACD/Labs software (note all but FULL\_MWT and FULL\_MOLFORMULA are calculated on the parent structure)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **KEYS** | **COLUMN\_NAME** | **DATA\_TYPE** | **NULLABLE** | **COMMENT** |
| PK,FK | MOLREGNO | NUMBER(9,0) | NOT NULL | Foreign key to compounds table (compound structure) |
|  | MW\_FREEBASE | NUMBER(9,2) |  | Molecular weight of parent compound |
|  | ALOGP | NUMBER(9,2) |  | Calculated ALogP |
|  | HBA | NUMBER(3,0) |  | Number hydrogen bond acceptors |
|  | HBD | NUMBER(3,0) |  | Number hydrogen bond donors |
|  | PSA | NUMBER(9,2) |  | Polar surface area |
|  | RTB | NUMBER(3,0) |  | Number rotatable bonds |
|  | RO3\_PASS | VARCHAR2(3) |  | Indicates whether the compound passes the rule-of-three (mw < 300, logP < 3 etc) |
|  | NUM\_RO5\_VIOLATIONS | NUMBER(1,0) |  | Number of violations of Lipinski's rule-of-five, using HBA and HBD definitions |
|  | ACD\_MOST\_APKA | NUMBER(9,2) |  | The most acidic pKa calculated using ACDlabs v12.01 |
|  | ACD\_MOST\_BPKA | NUMBER(9,2) |  | The most basic pKa calculated using ACDlabs v12.01 |
|  | ACD\_LOGP | NUMBER(9,2) |  | The calculated octanol/water partition coefficient using ACDlabs v12.01 |
|  | ACD\_LOGD | NUMBER(9,2) |  | The calculated octanol/water distribution coefficient at pH7.4 using ACDlabs v12.01 |
|  | MOLECULAR\_SPECIES | VARCHAR2(50) |  | Indicates whether the compound is an acid/base/neutral |
|  | FULL\_MWT | NUMBER(9,2) |  | Molecular weight of the full compound including any salts |
|  | AROMATIC\_RINGS | NUMBER(3,0) |  | Number of aromatic rings |
|  | HEAVY\_ATOMS | NUMBER(3,0) |  | Number of heavy (non-hydrogen) atoms |
|  | QED\_WEIGHTED | NUMBER(3,2) |  | Weighted quantitative estimate of drug likeness (as defined by Bickerton et al., Nature Chem 2012) |
|  | MW\_MONOISOTOPIC | NUMBER(11,4) |  | Monoisotopic parent molecular weight |
|  | FULL\_MOLFORMULA | VARCHAR2(100) |  | Molecular formula for the full compound (including any salt) |
|  | HBA\_LIPINSKI | NUMBER(3,0) |  | Number of hydrogen bond acceptors calculated according to Lipinski's original rules (i.e., N + O count)) |
|  | HBD\_LIPINSKI | NUMBER(3,0) |  | Number of hydrogen bond donors calculated according to Lipinski's original rules (i.e., NH + OH count) |
|  | NUM\_LIPINSKI\_RO5\_VIOLATIONS | NUMBER(1,0) |  | Number of violations of Lipinski's rule of five using HBA\_LIPINSKI and HBD\_LIPINSKI counts |

**COMPOUND\_RECORDS:**

Represents each compound extracted from scientific documents.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **KEYS** | **COLUMN\_NAME** | **DATA\_TYPE** | **NULLABLE** | **COMMENT** |
| PK | RECORD\_ID | NUMBER(9,0) | NOT NULL | Unique ID for a compound/record |
| FK | MOLREGNO | NUMBER(9,0) |  | Foreign key to compounds table (compound structure) |
| FK | DOC\_ID | NUMBER(9,0) | NOT NULL | Foreign key to documents table |
|  | COMPOUND\_KEY | VARCHAR2(250) |  | Key text identifying this compound in the scientific document |
|  | COMPOUND\_NAME | VARCHAR2(4000) |  | Name of this compound recorded in the scientific document |
| FK | SRC\_ID | NUMBER(3,0) | NOT NULL | Foreign key to source table |
|  | SRC\_COMPOUND\_ID | VARCHAR2(150) |  | Identifier for the compound in the source database (e.g., pubchem SID) |
|  | CIDX | VARCHAR2(200) | NOT NULL | The Depositor Defined Compound Identifier. |

**COMPOUND\_STRUCTURAL\_ALERTS:**

Table showing which structural alerts (as defined in the STRUCTURAL\_ALERTS table) are found in a particular ChEMBL compound. It should be noted some alerts/alert sets are more permissive than others and may flag a large number of compounds. Results should be interpreted with care, depending on the use-case, and not treated as a blanket filter (e.g., around 50% of approved drugs have 1 or more alerts from these sets).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **KEYS** | **COLUMN\_NAME** | **DATA\_TYPE** | **NULLABLE** | **COMMENT** |
| PK | CPD\_STR\_ALERT\_ID | NUMBER(9,0) | NOT NULL | Primary key. |
| FK,UK | MOLREGNO | NUMBER(9,0) | NOT NULL | Foreign key to the molecule\_dictionary. The compound for which the structural alert has been found. |
| FK,UK | ALERT\_ID | NUMBER(9,0) | NOT NULL | Foreign key to the structural\_alerts table. The particular alert that has been identified in this compound. |

**COMPOUND\_STRUCTURES:**

Table storing various structure representations (e.g., Molfile, InChI) for each compound

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **KEYS** | **COLUMN\_NAME** | **DATA\_TYPE** | **NULLABLE** | **COMMENT** |
| PK,FK | MOLREGNO | NUMBER(9,0) | NOT NULL | Internal Primary Key for the compound structure and foreign key to molecule\_dictionary table |
|  | MOLFILE | CLOB |  | MDL Connection table representation of compound |
| UK | STANDARD\_INCHI | VARCHAR2(4000) |  | IUPAC standard InChI for the compound |
| UK | STANDARD\_INCHI\_KEY | VARCHAR2(27) | NOT NULL | IUPAC standard InChI key for the compound |
|  | CANONICAL\_SMILES | VARCHAR2(4000) |  | Canonical smiles, generated using pipeline pilot |

**CONFIDENCE\_SCORE\_LOOKUP:**

Lookup table describing how assay2target confidence scores are assigned depending on the type of target(s) assigned to the assay and the level of confidence in their molecular identity

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **KEYS** | **COLUMN\_NAME** | **DATA\_TYPE** | **NULLABLE** | **COMMENT** |
| PK | CONFIDENCE\_SCORE | NUMBER(1,0) | NOT NULL | 0-9 score showing level of confidence in assignment of the precise molecular target of the assay |
|  | DESCRIPTION | VARCHAR2(100) | NOT NULL | Description of the target types assigned with each score |
|  | TARGET\_MAPPING | VARCHAR2(30) | NOT NULL | Short description of the target types assigned with each score |

**CURATION\_LOOKUP:**

Lookup table for assays.curated\_by column. Shows level of curation that has been applied to the assay to target mapping.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **KEYS** | **COLUMN\_NAME** | **DATA\_TYPE** | **NULLABLE** | **COMMENT** |
| PK | CURATED\_BY | VARCHAR2(32) | NOT NULL | Short description of the level of curation |
|  | DESCRIPTION | VARCHAR2(100) | NOT NULL | Definition of terms in the curated\_by field. |

**DATA\_VALIDITY\_LOOKUP:**

Table storing information about the data\_validity\_comment values used in the activities table.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **KEYS** | **COLUMN\_NAME** | **DATA\_TYPE** | **NULLABLE** | **COMMENT** |
| PK | DATA\_VALIDITY\_COMMENT | VARCHAR2(30) | NOT NULL | Primary key. Short description of various types of errors/warnings applied to values in the activities table. |
|  | DESCRIPTION | VARCHAR2(200) |  | Definition of the terms in the data\_validity\_comment field. |

**DEFINED\_DAILY\_DOSE:**

WHO DDD (defined daily dose) information

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **KEYS** | **COLUMN\_NAME** | **DATA\_TYPE** | **NULLABLE** | **COMMENT** |
| FK | ATC\_CODE | VARCHAR2(10) | NOT NULL | ATC code for the compound (foreign key to ATC\_CLASSIFICATION table) |
|  | DDD\_UNITS | VARCHAR2(200) |  | Units of defined daily dose |
|  | DDD\_ADMR | VARCHAR2(1000) |  | Administration route for dose |
|  | DDD\_COMMENT | VARCHAR2(2000) |  | Comment |
| PK | DDD\_ID | NUMBER(9,0) | NOT NULL | Internal primary key |
|  | DDD\_VALUE | NUMBER |  | None |

**DOCS:**

Holds all scientific documents (journal articles or patents) from which assays have been extracted.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **KEYS** | **COLUMN\_NAME** | **DATA\_TYPE** | **NULLABLE** | **COMMENT** |
| PK | DOC\_ID | NUMBER(9,0) | NOT NULL | Unique ID for the document |
|  | JOURNAL | VARCHAR2(50) |  | Abbreviated journal name for an article |
|  | YEAR | NUMBER(4,0) |  | Year of journal article publication |
|  | VOLUME | VARCHAR2(50) |  | Volume of journal article |
|  | ISSUE | VARCHAR2(50) |  | Issue of journal article |
|  | FIRST\_PAGE | VARCHAR2(50) |  | First page number of journal article |
|  | LAST\_PAGE | VARCHAR2(50) |  | Last page number of journal article |
|  | PUBMED\_ID | NUMBER(11,0) |  | NIH pubmed record ID, where available |
|  | DOI | VARCHAR2(100) |  | Digital object identifier for this reference |
| FK,UK | CHEMBL\_ID | VARCHAR2(20) | NOT NULL | ChEMBL identifier for this document (for use on web interface etc) |
|  | TITLE | VARCHAR2(500) |  | Document title (e.g., Publication title or description of dataset) |
|  | DOC\_TYPE | VARCHAR2(50) | NOT NULL | Type of the document (e.g., Publication, Deposited dataset) |
|  | AUTHORS | VARCHAR2(4000) |  | For a deposited dataset, the authors carrying out the screening and/or submitting the dataset. |
|  | ABSTRACT | CLOB |  | For a deposited dataset, a brief description of the dataset. |
|  | PATENT\_ID | VARCHAR2(20) |  | Patent ID for this document |
|  | RIDX | VARCHAR2(200) | NOT NULL | The Depositor Defined Reference Identifier |
| FK | SRC\_ID | NUMBER | NOT NULL | Foreign key to Source table, indicating the source of this document |

**DOMAINS:**

Table storing a non-redundant list of domains found in protein targets (e.g., Pfam domains).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **KEYS** | **COLUMN\_NAME** | **DATA\_TYPE** | **NULLABLE** | **COMMENT** |
| PK | DOMAIN\_ID | NUMBER(9,0) | NOT NULL | Primary key. Unique identifier for each domain. |
|  | DOMAIN\_TYPE | VARCHAR2(20) | NOT NULL | Indicates the source of the domain (e.g., Pfam). |
|  | SOURCE\_DOMAIN\_ID | VARCHAR2(20) | NOT NULL | Identifier for the domain in the source database (e.g., Pfam ID such as PF00001). |
|  | DOMAIN\_NAME | VARCHAR2(20) |  | Name given to the domain in the source database (e.g., 7tm\_1). |
|  | DOMAIN\_DESCRIPTION | VARCHAR2(500) |  | Longer name or description for the domain. |

**DRUG\_INDICATION:**

Table storing indications for drugs from a variety of sources (e.g., ATC, DailyMed, ClinicalTrials.gov)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **KEYS** | **COLUMN\_NAME** | **DATA\_TYPE** | **NULLABLE** | **COMMENT** |
| PK | DRUGIND\_ID | NUMBER(9,0) | NOT NULL | Primary key |
| FK,UK | RECORD\_ID | NUMBER(9,0) | NOT NULL | Foreign key to compound\_records table. Links to the drug record to which this indication applies |
| FK | MOLREGNO | NUMBER(9,0) |  | Molregno corresponding to the record\_id in the compound\_records table |
|  | MAX\_PHASE\_FOR\_IND | NUMBER(1,0) |  | The maximum phase of development that the drug is known to have reached for this particular indication |
| UK | MESH\_ID | VARCHAR2(20) | NOT NULL | Medical Subject Headings (MeSH) disease identifier corresponding to the indication |
|  | MESH\_HEADING | VARCHAR2(200) | NOT NULL | Medical Subject Heading term for the MeSH disease ID |
| UK | EFO\_ID | VARCHAR2(20) |  | Experimental Factor Ontology (EFO) disease identifier corresponding to the indication |
|  | EFO\_TERM | VARCHAR2(200) |  | Experimental Factor Ontology term for the EFO ID |

**DRUG\_MECHANISM:**

Table storing mechanism of action information for FDA-approved drugs and WHO anti-malarials

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **KEYS** | **COLUMN\_NAME** | **DATA\_TYPE** | **NULLABLE** | **COMMENT** |
| PK | MEC\_ID | NUMBER(9,0) | NOT NULL | Primary key for each drug mechanism of action |
| FK | RECORD\_ID | NUMBER(9,0) | NOT NULL | Record\_id for the drug (foreign key to compound\_records table) |
| FK | MOLREGNO | NUMBER(9,0) |  | Molregno for the drug (foreign key to molecule\_dictionary table) |
|  | MECHANISM\_OF\_ACTION | VARCHAR2(250) |  | Description of the mechanism of action e.g., 'Phosphodiesterase 5 inhibitor' |
| FK | TID | NUMBER(9,0) |  | Target associated with this mechanism of action (foreign key to target\_dictionary table) |
| FK | SITE\_ID | NUMBER(9,0) |  | Binding site for the drug within the target (where known) - foreign key to binding\_sites table |
| FK | ACTION\_TYPE | VARCHAR2(50) |  | Type of action of the drug on the target e.g., agonist/antagonist etc (foreign key to action\_type table) |
|  | DIRECT\_INTERACTION | NUMBER(1,0) |  | Flag to show whether the molecule is believed to interact directly with the target (1 = yes, 0 = no) |
|  | MOLECULAR\_MECHANISM | NUMBER(1,0) |  | Flag to show whether the mechanism of action describes the molecular target of the drug, rather than a higher-level physiological mechanism e.g., vasodilator (1 = yes, 0 = no) |
|  | DISEASE\_EFFICACY | NUMBER(1,0) |  | Flag to show whether the target assigned is believed to play a role in the efficacy of the drug in the indication(s) for which it is approved (1 = yes, 0 = no) |
|  | MECHANISM\_COMMENT | VARCHAR2(2000) |  | Additional comments regarding the mechanism of action |
|  | SELECTIVITY\_COMMENT | VARCHAR2(1000) |  | Additional comments regarding the selectivity of the drug |
|  | BINDING\_SITE\_COMMENT | VARCHAR2(1000) |  | Additional comments regarding the binding site of the drug |

**FORMULATIONS:**

Table linking individual ingredients in approved products (and their strengths) to entries in the molecule dictionary. Where products are mixtures of active ingredients they will have multiple entries in this table

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **KEYS** | **COLUMN\_NAME** | **DATA\_TYPE** | **NULLABLE** | **COMMENT** |
| FK,UK | PRODUCT\_ID | VARCHAR2(30) | NOT NULL | Unique identifier of the product. FK to PRODUCTS |
|  | INGREDIENT | VARCHAR2(200) |  | Name of the approved ingredient within the product |
|  | STRENGTH | VARCHAR2(200) |  | Dose strength |
| FK,UK | RECORD\_ID | NUMBER(9,0) | NOT NULL | Foreign key to the compound\_records table. |
| FK | MOLREGNO | NUMBER(9,0) |  | Unique identifier of the ingredient FK to MOLECULE\_DICTIONARY |
| PK | FORMULATION\_ID | NUMBER(9,0) | NOT NULL | Primary key. |

**FRAC\_CLASSIFICATION:**

Table showing classification of fungicide mechanism of action according to the Fungicide Resistance Action Committee (FRAC): http://www.frac.info/publication/anhang/FRAC%20Code%20List%202013-final.pdf

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **KEYS** | **COLUMN\_NAME** | **DATA\_TYPE** | **NULLABLE** | **COMMENT** |
| PK | FRAC\_CLASS\_ID | NUMBER(9,0) | NOT NULL | Unique numeric primary key for each level5 code |
|  | ACTIVE\_INGREDIENT | VARCHAR2(500) | NOT NULL | Name of active ingredient (fungicide) classified by FRAC |
|  | LEVEL1 | VARCHAR2(2) | NOT NULL | Mechanism of action code assigned by FRAC |
|  | LEVEL1\_DESCRIPTION | VARCHAR2(2000) | NOT NULL | Description of mechanism of action |
|  | LEVEL2 | VARCHAR2(2) | NOT NULL | Target site code assigned by FRAC |
|  | LEVEL2\_DESCRIPTION | VARCHAR2(2000) |  | Description of target provided by FRAC |
|  | LEVEL3 | VARCHAR2(6) | NOT NULL | Group number assigned by FRAC |
|  | LEVEL3\_DESCRIPTION | VARCHAR2(2000) |  | Description of group provided by FRAC |
|  | LEVEL4 | VARCHAR2(7) | NOT NULL | Number denoting the chemical group (number not assigned by FRAC) |
|  | LEVEL4\_DESCRIPTION | VARCHAR2(2000) |  | Chemical group name provided by FRAC |
| UK | LEVEL5 | VARCHAR2(8) | NOT NULL | A unique code assigned to each ingredient (based on the level 1-4 FRAC classification, but not assigned by IRAC) |
|  | FRAC\_CODE | VARCHAR2(4) | NOT NULL | The official FRAC classification code for the ingredient |

**GO\_CLASSIFICATION:**

Table storing the ChEMBL Drug Target GO slim (http://www.geneontology.org/ontology/subsets/goslim\_chembl.obo)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **KEYS** | **COLUMN\_NAME** | **DATA\_TYPE** | **NULLABLE** | **COMMENT** |
| PK | GO\_ID | VARCHAR2(10) | NOT NULL | Primary key. Gene Ontology identifier for the GO slim term |
|  | PARENT\_GO\_ID | VARCHAR2(10) |  | Gene Ontology identifier for the parent of this GO term in the ChEMBL Drug Target GO slim |
|  | PREF\_NAME | VARCHAR2(200) |  | Gene Ontology name |
|  | CLASS\_LEVEL | NUMBER(1,0) |  | Indicates the level of the term in the slim (L1 = highest) |
|  | ASPECT | VARCHAR2(1) |  | Indicates which aspect of the Gene Ontology the term belongs to (F = molecular function, P = biological process, C = cellular component) |
|  | PATH | VARCHAR2(1000) |  | Indicates the full path to this term in the GO slim |

**HRAC\_CLASSIFICATION:**

Table showing classification of herbicide mechanism of action according to the Herbicide Resistance Action Committee (HRAC): http://www.hracglobal.com/Education/ClassificationofHerbicideSiteofAction.aspx

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **KEYS** | **COLUMN\_NAME** | **DATA\_TYPE** | **NULLABLE** | **COMMENT** |
| PK | HRAC\_CLASS\_ID | NUMBER(9,0) | NOT NULL | Unique numeric primary key for each level3 code |
|  | ACTIVE\_INGREDIENT | VARCHAR2(500) | NOT NULL | Name of active ingredient (herbicide) classified by HRAC |
|  | LEVEL1 | VARCHAR2(2) | NOT NULL | HRAC group code - denoting mechanism of action of herbicide |
|  | LEVEL1\_DESCRIPTION | VARCHAR2(2000) | NOT NULL | Description of mechanism of action provided by HRAC |
|  | LEVEL2 | VARCHAR2(3) | NOT NULL | Indicates a chemical family within a particular HRAC group (number not assigned by HRAC) |
|  | LEVEL2\_DESCRIPTION | VARCHAR2(2000) |  | Description of chemical family provided by HRAC |
| UK | LEVEL3 | VARCHAR2(5) | NOT NULL | A unique code assigned to each ingredient (based on the level 1 and 2 HRAC classification, but not assigned by HRAC) |
|  | HRAC\_CODE | VARCHAR2(2) | NOT NULL | The official HRAC classification code for the ingredient |

**INDICATION\_REFS:**

Table storing references indicating the source of drug indication information

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **KEYS** | **COLUMN\_NAME** | **DATA\_TYPE** | **NULLABLE** | **COMMENT** |
| PK | INDREF\_ID | NUMBER(9,0) | NOT NULL | Primary key |
| FK,UK | DRUGIND\_ID | NUMBER(9,0) | NOT NULL | Foreign key to the DRUG\_INDICATION table, indicating the drug-indication link that this reference applies to |
| UK | REF\_TYPE | VARCHAR2(50) | NOT NULL | Type/source of reference |
| UK | REF\_ID | VARCHAR2(4000) | NOT NULL | Identifier for the reference in the source |
|  | REF\_URL | VARCHAR2(4000) | NOT NULL | Full URL linking to the reference |

**IRAC\_CLASSIFICATION:**

Table showing classification of insecticide mechanism of action according to the Insecticide Resistance Action Committee (IRAC): http://www.irac-online.org/documents/moa-classification/?ext=pdf

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **KEYS** | **COLUMN\_NAME** | **DATA\_TYPE** | **NULLABLE** | **COMMENT** |
| PK | IRAC\_CLASS\_ID | NUMBER(9,0) | NOT NULL | Unique numeric primary key for each level4 code |
|  | ACTIVE\_INGREDIENT | VARCHAR2(500) | NOT NULL | Name of active ingredient (insecticide) classified by IRAC |
|  | LEVEL1 | VARCHAR2(1) | NOT NULL | Class of action e.g., nerve action, energy metabolism (code not assigned by IRAC) |
|  | LEVEL1\_DESCRIPTION | VARCHAR2(2000) | NOT NULL | Description of class of action, as provided by IRAC |
|  | LEVEL2 | VARCHAR2(3) | NOT NULL | IRAC main group code denoting primary site/mechanism of action |
|  | LEVEL2\_DESCRIPTION | VARCHAR2(2000) | NOT NULL | Description of site/mechanism of action provided by IRAC |
|  | LEVEL3 | VARCHAR2(6) | NOT NULL | IRAC sub-group code denoting chemical class of insecticide |
|  | LEVEL3\_DESCRIPTION | VARCHAR2(2000) | NOT NULL | Description of chemical class or exemplifying ingredient provided by IRAC |
| UK | LEVEL4 | VARCHAR2(8) | NOT NULL | A unique code assigned to each ingredient (based on the level 1, 2 and 3 IRAC classification, but not assigned by IRAC) |
|  | IRAC\_CODE | VARCHAR2(3) | NOT NULL | The official IRAC classification code for the ingredient |

**LIGAND\_EFF:**

Contains BEI (Binding Efficiency Index) and SEI (Surface Binding Efficiency Index) for each activity\_id where such data can be calculated.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **KEYS** | **COLUMN\_NAME** | **DATA\_TYPE** | **NULLABLE** | **COMMENT** |
| PK,FK | ACTIVITY\_ID | NUMBER(11,0) | NOT NULL | Link key to activities table |
|  | BEI | NUMBER(9,2) |  | Binding Efficiency Index = p(XC50) \*1000/MW\_freebase |
|  | SEI | NUMBER(9,2) |  | Surface Efficiency Index = p(XC50)\*100/PSA |
|  | LE | NUMBER(9,2) |  | Ligand Efficiency = deltaG/heavy\_atoms [from the Hopkins DDT paper 2004] |
|  | LLE | NUMBER(9,2) |  | Lipophilic Ligand Efficiency = -logKi-ALogP. [from Leeson NRDD 2007] |

**MECHANISM\_REFS:**

Table storing references for information in the drug\_mechanism table

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **KEYS** | **COLUMN\_NAME** | **DATA\_TYPE** | **NULLABLE** | **COMMENT** |
| PK | MECREF\_ID | NUMBER(9,0) | NOT NULL | Primary key |
| FK,UK | MEC\_ID | NUMBER(9,0) | NOT NULL | Foreign key to drug\_mechanism table - indicating the mechanism to which the references refer |
| UK | REF\_TYPE | VARCHAR2(50) | NOT NULL | Type/source of reference (e.g., 'PubMed','DailyMed') |
| UK | REF\_ID | VARCHAR2(200) |  | Identifier for the reference in the source (e.g., PubMed ID or DailyMed setid) |
|  | REF\_URL | VARCHAR2(400) |  | Full URL linking to the reference |

**METABOLISM:**

Table storing drug metabolic pathways, manually curated from a variety of sources

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **KEYS** | **COLUMN\_NAME** | **DATA\_TYPE** | **NULLABLE** | **COMMENT** |
| PK | MET\_ID | NUMBER(9,0) | NOT NULL | Primary key |
| FK,UK | DRUG\_RECORD\_ID | NUMBER(9,0) |  | Foreign key to compound\_records. Record representing the drug or other compound for which metabolism is being studied (may not be the same as the substrate being measured) |
| FK,UK | SUBSTRATE\_RECORD\_ID | NUMBER(9,0) |  | Foreign key to compound\_records. Record representing the compound that is the subject of metabolism |
| FK,UK | METABOLITE\_RECORD\_ID | NUMBER(9,0) |  | Foreign key to compound\_records. Record representing the compound that is the result of metabolism |
| UK | PATHWAY\_ID | NUMBER(9,0) |  | Identifier for the metabolic scheme/pathway (may be multiple pathways from one source document) |
|  | PATHWAY\_KEY | VARCHAR2(50) |  | Link to original source indicating where the pathway information was found (e.g., Figure 1, page 23) |
| UK | ENZYME\_NAME | VARCHAR2(200) |  | Name of the enzyme responsible for the metabolic conversion |
| FK,UK | ENZYME\_TID | NUMBER(9,0) |  | Foreign key to target\_dictionary. TID for the enzyme responsible for the metabolic conversion |
|  | MET\_CONVERSION | VARCHAR2(200) |  | Description of the metabolic conversion |
|  | ORGANISM | VARCHAR2(100) |  | Organism in which this metabolic reaction occurs |
| UK | TAX\_ID | NUMBER(11,0) |  | NCBI Tax ID for the organism in which this metabolic reaction occurs |
|  | MET\_COMMENT | VARCHAR2(1000) |  | Additional information regarding the metabolism (e.g., organ system, conditions under which observed, activity of metabolites) |

**METABOLISM\_REFS:**

Table storing references for metabolic pathways, indicating the source of the data

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **KEYS** | **COLUMN\_NAME** | **DATA\_TYPE** | **NULLABLE** | **COMMENT** |
| PK | METREF\_ID | NUMBER(9,0) | NOT NULL | Primary key |
| FK,UK | MET\_ID | NUMBER(9,0) | NOT NULL | Foreign key to record\_metabolism table - indicating the metabolism information to which the references refer |
| UK | REF\_TYPE | VARCHAR2(50) | NOT NULL | Type/source of reference (e.g., 'PubMed','DailyMed') |
| UK | REF\_ID | VARCHAR2(200) |  | Identifier for the reference in the source (e.g., PubMed ID or DailyMed setid) |
|  | REF\_URL | VARCHAR2(400) |  | Full URL linking to the reference |

**MOLECULE\_ATC\_CLASSIFICATION:**

Table mapping drugs in the molecule\_dictionary to ATC codes in the atc\_classification table

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **KEYS** | **COLUMN\_NAME** | **DATA\_TYPE** | **NULLABLE** | **COMMENT** |
| PK | MOL\_ATC\_ID | NUMBER(9,0) | NOT NULL | Primary key |
| FK | LEVEL5 | VARCHAR2(10) | NOT NULL | ATC code (foreign key to atc\_classification table) |
| FK | MOLREGNO | NUMBER(9,0) | NOT NULL | Drug to which the ATC code applies (foreign key to molecule\_dictionary table) |

**MOLECULE\_DICTIONARY:**

Non redundant list of compounds/biotherapeutics with associated identifiers

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **KEYS** | **COLUMN\_NAME** | **DATA\_TYPE** | **NULLABLE** | **COMMENT** |
| PK | MOLREGNO | NUMBER(9,0) | NOT NULL | Internal Primary Key for the molecule |
|  | PREF\_NAME | VARCHAR2(255) |  | Preferred name for the molecule |
| FK,UK | CHEMBL\_ID | VARCHAR2(20) | NOT NULL | ChEMBL identifier for this compound (for use on web interface etc) |
|  | MAX\_PHASE | NUMBER(1,0) | NOT NULL | Maximum phase of development reached for the compound (4 = approved). Null where max phase has not yet been assigned. |
|  | THERAPEUTIC\_FLAG | NUMBER(1,0) | NOT NULL | Indicates that a drug has a therapeutic application (as opposed to e.g., an imaging agent, additive etc). |
|  | DOSED\_INGREDIENT | NUMBER(1,0) | NOT NULL | Indicates that the drug is dosed in this form (e.g., a particular salt) |
|  | STRUCTURE\_TYPE | VARCHAR2(10) | NOT NULL | Indications whether the molecule has a small molecule structure or a protein sequence (MOL indicates an entry in the compound\_structures table, SEQ indications an entry in the protein\_therapeutics table, NONE indicates an entry in neither table, e.g., structure unknown) |
|  | CHEBI\_PAR\_ID | NUMBER(9,0) |  | Preferred ChEBI ID for the compound (where different from assigned) |
|  | MOLECULE\_TYPE | VARCHAR2(30) |  | Type of molecule (Small molecule, Protein, Antibody, Oligosaccharide, Oligonucleotide, Cell, Unknown) |
|  | FIRST\_APPROVAL | NUMBER(4,0) |  | Earliest known approval year for the molecule |
|  | ORAL | NUMBER(1,0) | NOT NULL | Indicates whether the drug is known to be administered orally. |
|  | PARENTERAL | NUMBER(1,0) | NOT NULL | Indicates whether the drug is known to be administered parenterally |
|  | TOPICAL | NUMBER(1,0) | NOT NULL | Indicates whether the drug is known to be administered topically. |
|  | BLACK\_BOX\_WARNING | NUMBER(1,0) | NOT NULL | Indicates that the drug has a black box warning |
|  | NATURAL\_PRODUCT | NUMBER(1,0) | NOT NULL | Indicates whether the compound is natural product-derived (currently curated only for drugs) |
|  | FIRST\_IN\_CLASS | NUMBER(1,0) | NOT NULL | Indicates whether this is known to be the first compound of its class (e.g., acting on a particular target). |
|  | CHIRALITY | NUMBER(1,0) | NOT NULL | Shows whether a drug is dosed as a racemic mixture (0), single stereoisomer (1) or is an achiral molecule (2) |
|  | PRODRUG | NUMBER(1,0) | NOT NULL | Indicates that the molecule is a pro-drug (see molecule hierarchy for active component, where known) |
|  | INORGANIC\_FLAG | NUMBER(1,0) | NOT NULL | Indicates whether the molecule is inorganic (i.e., containing only metal atoms and <2 carbon atoms) |
|  | USAN\_YEAR | NUMBER(4,0) |  | The year in which the application for a USAN/INN name was made |
|  | AVAILABILITY\_TYPE | NUMBER(1,0) |  | The availability type for the drug (0 = discontinued, 1 = prescription only, 2 = over the counter) |
|  | USAN\_STEM | VARCHAR2(50) |  | Where the compound has been assigned a USAN name, this indicates the stem, as described in the USAN\_STEM table. |
|  | POLYMER\_FLAG | NUMBER(1,0) |  | Indicates whether a molecule is a small molecule polymer (e.g., polistyrex) |
|  | USAN\_SUBSTEM | VARCHAR2(50) |  | Where the compound has been assigned a USAN name, this indicates the substem |
|  | USAN\_STEM\_DEFINITION | VARCHAR2(1000) |  | Definition of the USAN stem |
|  | INDICATION\_CLASS | VARCHAR2(1000) |  | Indication class(es) assigned to a drug in the USP dictionary |
|  | WITHDRAWN\_FLAG | NUMBER(1,0) | NOT NULL | Flag indicating whether the drug has been withdrawn in at least one country (not necessarily in the US) |
|  | WITHDRAWN\_YEAR | NUMBER(4,0) |  | Year the drug was first withdrawn in any country |
|  | WITHDRAWN\_COUNTRY | VARCHAR2(1000) |  | List of countries/regions where the drug has been withdrawn |
|  | WITHDRAWN\_REASON | VARCHAR2(1000) |  | Reasons for withdrawl (e.g., safety) |
|  | WITHDRAWN\_CLASS | VARCHAR2(500) |  | High level categories for the withdrawn reason (e.g., Cardiotoxicity, Hepatotoxicity) |

**MOLECULE\_FRAC\_CLASSIFICATION:**

Table showing Fungicide Resistance Action Committee (FRAC) mechanism of action classification for known crop protection fungicides.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **KEYS** | **COLUMN\_NAME** | **DATA\_TYPE** | **NULLABLE** | **COMMENT** |
| PK | MOL\_FRAC\_ID | NUMBER(9,0) | NOT NULL | Primary key. |
| FK,UK | FRAC\_CLASS\_ID | NUMBER(9,0) | NOT NULL | Foreign key to frac\_classification table showing the mechanism of action classification of the compound. |
| FK,UK | MOLREGNO | NUMBER(9,0) | NOT NULL | Foreign key to molecule\_dictionary, showing the compound to which the classification applies. |

**MOLECULE\_HIERARCHY:**

Table storing relationships between parents, salts and active metabolites (for pro-drugs).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **KEYS** | **COLUMN\_NAME** | **DATA\_TYPE** | **NULLABLE** | **COMMENT** |
| PK,FK | MOLREGNO | NUMBER(9,0) | NOT NULL | Foreign key to compounds table. This field holds a list of all of the ChEMBL compounds with associated data (e.g., activity information, approved drugs). Parent compounds that are generated only by removing salts, and which do not themselves have any associated data will not appear here. |
| FK | PARENT\_MOLREGNO | NUMBER(9,0) |  | Represents parent compound of molregno in first field (i.e., generated by removing salts). Where molregno and parent\_molregno are same, the initial ChEMBL compound did not contain a salt component, or else could not be further processed for various reasons (e.g., inorganic mixture). Compounds which are only generated by removing salts will appear in this field only. Those which, themselves, have any associated data (e.g., activity data) or are launched drugs will also appear in the molregno field. |
| FK | ACTIVE\_MOLREGNO | NUMBER(9,0) |  | Where a compound is a pro-drug, this represents the active metabolite of the 'dosed' compound given by parent\_molregno. Where parent\_molregno and active\_molregno are the same, the compound is not currently known to be a pro-drug. |

**MOLECULE\_HRAC\_CLASSIFICATION:**

Table showing Herbicide Resistance Action Committee (HRAC) mechanism of action classification for known herbicidal compounds.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **KEYS** | **COLUMN\_NAME** | **DATA\_TYPE** | **NULLABLE** | **COMMENT** |
| PK | MOL\_HRAC\_ID | NUMBER(9,0) | NOT NULL | Primary key |
| FK,UK | HRAC\_CLASS\_ID | NUMBER(9,0) | NOT NULL | Foreign key to hrac\_classification table showing the classification for the compound. |
| FK,UK | MOLREGNO | NUMBER(9,0) | NOT NULL | Foreign key to molecule\_dictionary, showing the compound to which this classification applies. |

**MOLECULE\_IRAC\_CLASSIFICATION:**

Table showing Insecticide Resistance Action Committee (IRAC) mechanism of action classification for known crop protection insecticides.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **KEYS** | **COLUMN\_NAME** | **DATA\_TYPE** | **NULLABLE** | **COMMENT** |
| PK | MOL\_IRAC\_ID | NUMBER(9,0) | NOT NULL | Primary key. |
| FK,UK | IRAC\_CLASS\_ID | NUMBER(9,0) | NOT NULL | Foreign key to the irac\_classification table showing the mechanism of action classification for the compound. |
| FK,UK | MOLREGNO | NUMBER(9,0) | NOT NULL | Foreign key to the molecule\_dictionary table, showing the compound to which the classification applies. |

**MOLECULE\_SYNONYMS:**

Stores synonyms for a compound (e.g., common names, trade names, research codes etc)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **KEYS** | **COLUMN\_NAME** | **DATA\_TYPE** | **NULLABLE** | **COMMENT** |
| FK,UK | MOLREGNO | NUMBER(9,0) | NOT NULL | Foreign key to molecule\_dictionary |
| UK | SYN\_TYPE | VARCHAR2(50) | NOT NULL | Type of name/synonym (e.g., TRADE\_NAME, RESEARCH\_CODE, USAN) |
| PK | MOLSYN\_ID | NUMBER(9,0) | NOT NULL | Primary key. |
| FK | RES\_STEM\_ID | NUMBER(9,0) |  | Foreign key to the research\_stem table. Where a synonym is a research code, this links to further information about the company associated with that code. |
| UK | SYNONYMS | VARCHAR2(200) |  | Synonym for the compound |

**ORGANISM\_CLASS:**

Simple organism classification (essentially a cut-down version of the NCBI taxonomy for organisms in ChEMBL assay2target and target\_dictionary tables), allowing browsing of ChEMBL data by taxonomic groups

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **KEYS** | **COLUMN\_NAME** | **DATA\_TYPE** | **NULLABLE** | **COMMENT** |
| PK | OC\_ID | NUMBER(9,0) | NOT NULL | Internal primary key |
| UK | TAX\_ID | NUMBER(11,0) |  | NCBI taxonomy ID for the organism (corresponding to tax\_ids in assay2target and target\_dictionary tables) |
|  | L1 | VARCHAR2(200) |  | Highest level classification (e.g., Eukaryotes, Bacteria, Fungi etc) |
|  | L2 | VARCHAR2(200) |  | Second level classification |
|  | L3 | VARCHAR2(200) |  | Third level classification |

**PATENT\_USE\_CODES:**

Table from FDA Orange Book, showing definitions of different patent use codes (as used in the product\_patents table).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **KEYS** | **COLUMN\_NAME** | **DATA\_TYPE** | **NULLABLE** | **COMMENT** |
| PK | PATENT\_USE\_CODE | VARCHAR2(8) | NOT NULL | Primary key. Patent use code from FDA Orange Book |
|  | DEFINITION | VARCHAR2(500) | NOT NULL | Definition for the patent use code, from FDA Orange Book. |

**PREDICTED\_BINDING\_DOMAINS:**

Table storing information on the likely binding domain of compounds in the activities table (based on analysis of the domain structure of the target. Note these are predictions, not experimentally determined. See Kruger F, Rostom R and Overington JP (2012), BMC Bioinformatics, 13(S17), S11 for more details.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **KEYS** | **COLUMN\_NAME** | **DATA\_TYPE** | **NULLABLE** | **COMMENT** |
| PK | PREDBIND\_ID | NUMBER(9,0) | NOT NULL | Primary key. |
| FK | ACTIVITY\_ID | NUMBER(11,0) |  | Foreign key to the activities table, indicating the compound/assay(+target) combination for which this prediction is made. |
| FK | SITE\_ID | NUMBER(9,0) |  | Foreign key to the binding\_sites table, indicating the binding site (domain) that the compound is predicted to bind to. |
|  | PREDICTION\_METHOD | VARCHAR2(50) |  | The method used to assign the binding domain (e.g., 'Single domain' where the protein has only 1 domain, 'Multi domain' where the protein has multiple domains, but only 1 is known to bind small molecules in other proteins). |
|  | CONFIDENCE | VARCHAR2(10) |  | The level of confidence assigned to the prediction (high where the protein has only 1 domain, medium where the compound has multiple domains, but only 1 known small molecule-binding domain). |

**PRODUCT\_PATENTS:**

Table from FDA Orange Book, showing patents associated with drug products.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **KEYS** | **COLUMN\_NAME** | **DATA\_TYPE** | **NULLABLE** | **COMMENT** |
| PK | PROD\_PAT\_ID | NUMBER(9,0) | NOT NULL | Primary key |
| FK,UK | PRODUCT\_ID | VARCHAR2(30) | NOT NULL | Foreign key to products table - FDA application number for the product |
| UK | PATENT\_NO | VARCHAR2(11) | NOT NULL | Patent numbers as submitted by the applicant holder for patents covered by the statutory provisions |
| UK | PATENT\_EXPIRE\_DATE | DATE | NOT NULL | Date the patent expires as submitted by the applicant holder including applicable extensions |
|  | DRUG\_SUBSTANCE\_FLAG | NUMBER(1,0) | NOT NULL | Patents submitted on FDA Form 3542 and listed after August 18, 2003 may have a drug substance flag set to 1, indicating the sponsor submitted the patent as claiming the drug substance |
|  | DRUG\_PRODUCT\_FLAG | NUMBER(1,0) | NOT NULL | Patents submitted on FDA Form 3542 and listed after August 18, 2003 may have a drug product flag set to 1, indicating the sponsor submitted the patent as claiming the drug product |
| FK,UK | PATENT\_USE\_CODE | VARCHAR2(10) |  | Code to designate a use patent that covers the approved indication or use of a drug product |
|  | DELIST\_FLAG | NUMBER(1,0) | NOT NULL | Sponsor has requested patent be delisted if set to 1. This patent has remained listed because, under Section 505(j)(5)(D)(i) of the Act, a first applicant may retain eligibility for 180-day exclusivity based on a paragraph IV certification to this patent for a certain period. Applicants under Section 505(b)(2) are not required to certify to patents where this flag is set to 1 |
|  | SUBMISSION\_DATE | DATE |  | The date on which the FDA receives patent information from the new drug application (NDA) holder. Format is Mmm d, yyyy |

**PRODUCTS:**

Table containing information about approved drug products (mainly from the FDA Orange Book), such as trade name, administration route, approval date. Ingredients in each product are linked to the molecule dictionary via the formulations table.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **KEYS** | **COLUMN\_NAME** | **DATA\_TYPE** | **NULLABLE** | **COMMENT** |
|  | DOSAGE\_FORM | VARCHAR2(200) |  | The dosage form of the product (e.g., tablet, capsule etc) |
|  | ROUTE | VARCHAR2(200) |  | The administration route of the product (e.g., oral, injection etc) |
|  | TRADE\_NAME | VARCHAR2(200) |  | The trade name for the product |
|  | APPROVAL\_DATE | DATE |  | The FDA approval date for the product (not necessarily first approval of the active ingredient) |
|  | AD\_TYPE | VARCHAR2(5) |  | RX = prescription, OTC = over the counter, DISCN = discontinued |
|  | ORAL | NUMBER(1,0) |  | Flag to show whether product is orally delivered |
|  | TOPICAL | NUMBER(1,0) |  | Flag to show whether product is topically delivered |
|  | PARENTERAL | NUMBER(1,0) |  | Flag to show whether product is parenterally delivered |
|  | BLACK\_BOX\_WARNING | NUMBER(1,0) |  | Flag to show whether the product label has a black box warning |
|  | APPLICANT\_FULL\_NAME | VARCHAR2(200) |  | Name of the company applying for FDA approval |
|  | INNOVATOR\_COMPANY | NUMBER(1,0) |  | Flag to show whether the applicant is the innovator of the product |
| PK | PRODUCT\_ID | VARCHAR2(30) | NOT NULL | FDA application number for the product |
|  | NDA\_TYPE | VARCHAR2(10) |  | New Drug Application Type. The type of new drug application approval. New Drug Applications (NDA or innovator) are Â¿NÂ¿. Abbreviated New Drug Applications (ANDA or generic) are Â¿AÂ¿. |

**PROTEIN\_CLASS\_SYNONYMS:**

Table storing synonyms for the protein family classifications (from various sources including MeSH, ConceptWiki and UMLS).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **KEYS** | **COLUMN\_NAME** | **DATA\_TYPE** | **NULLABLE** | **COMMENT** |
| PK | PROTCLASSSYN\_ID | NUMBER(9,0) | NOT NULL | Primary key. |
| FK,UK | PROTEIN\_CLASS\_ID | NUMBER(9,0) | NOT NULL | Foreign key to the PROTEIN\_CLASSIFICATION table. The protein\_class to which this synonym applies. |
| UK | PROTEIN\_CLASS\_SYNONYM | VARCHAR2(1000) |  | The synonym for the protein class. |
| UK | SYN\_TYPE | VARCHAR2(20) |  | The type or origin of the synonym (e.g., ChEMBL, Concept Wiki, UMLS). |

**PROTEIN\_CLASSIFICATION:**

Table storing the protein family classifications for protein targets in ChEMBL (formerly in the target\_class table)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **KEYS** | **COLUMN\_NAME** | **DATA\_TYPE** | **NULLABLE** | **COMMENT** |
| PK | PROTEIN\_CLASS\_ID | NUMBER(9,0) | NOT NULL | Primary key. Unique identifier for each protein family classification. |
|  | PARENT\_ID | NUMBER(9,0) |  | Protein\_class\_id for the parent of this protein family. |
|  | PREF\_NAME | VARCHAR2(500) |  | Preferred/full name for this protein family. |
|  | SHORT\_NAME | VARCHAR2(50) |  | Short/abbreviated name for this protein family (not necessarily unique). |
|  | PROTEIN\_CLASS\_DESC | VARCHAR2(410) | NOT NULL | Concatenated description of each classification for searching purposes etc. |
|  | DEFINITION | VARCHAR2(4000) |  | Definition of the protein family. |
|  | CLASS\_LEVEL | NUMBER(9,0) | NOT NULL | Level of the class within the hierarchy (level 1 = top level classification) |

**PROTEIN\_FAMILY\_CLASSIFICATION:**

WARNING! THIS TABLE IS NOW DEPRECATED, PLEASE USE PROTEIN\_CLASSIFICATION INSTEAD.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **KEYS** | **COLUMN\_NAME** | **DATA\_TYPE** | **NULLABLE** | **COMMENT** |
| PK | PROTEIN\_CLASS\_ID | NUMBER(9,0) | NOT NULL | Primary key. Unique identifier for each classification. |
| UK | PROTEIN\_CLASS\_DESC | VARCHAR2(810) | NOT NULL | Concatenated description of each classification for searching purposes etc. |
| UK | L1 | VARCHAR2(100) | NOT NULL | First level classification (e.g., Enzyme, Transporter, Ion Channel). |
| UK | L2 | VARCHAR2(100) |  | Second level classification. |
| UK | L3 | VARCHAR2(100) |  | Third level classification. |
| UK | L4 | VARCHAR2(100) |  | Fourth level classification. |
| UK | L5 | VARCHAR2(100) |  | Fifth level classification. |
| UK | L6 | VARCHAR2(100) |  | Sixth level classification. |
| UK | L7 | VARCHAR2(100) |  | Seventh level classification. |
| UK | L8 | VARCHAR2(100) |  | Eighth level classification. |

**RELATIONSHIP\_TYPE:**

Lookup table for assay2target.relationship\_type column, showing whether assays are mapped to targets of the correct identity and species ('Direct') or close homologues ('Homologue')

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **KEYS** | **COLUMN\_NAME** | **DATA\_TYPE** | **NULLABLE** | **COMMENT** |
| PK | RELATIONSHIP\_TYPE | VARCHAR2(1) | NOT NULL | Relationship\_type flag used in the assay2target table |
|  | RELATIONSHIP\_DESC | VARCHAR2(250) |  | Description of relationship\_type flags |

**RESEARCH\_COMPANIES:**

Table storing a list of pharmaceutical companies (including current and former names) corresponding to each research code stem in the research\_stem table. A stem can sometimes be used by more than one company.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **KEYS** | **COLUMN\_NAME** | **DATA\_TYPE** | **NULLABLE** | **COMMENT** |
| PK | CO\_STEM\_ID | NUMBER(9,0) | NOT NULL | Primary key. |
| FK,UK | RES\_STEM\_ID | NUMBER(9,0) |  | Foreign key to research\_stem table. |
| UK | COMPANY | VARCHAR2(100) |  | Name of current company associated with this research code stem. |
|  | COUNTRY | VARCHAR2(50) |  | Country in which the company uses this research code stem. |
|  | PREVIOUS\_COMPANY | VARCHAR2(100) |  | Previous name of the company associated with this research code stem (e.g., if the company has undergone acquisitions/mergers). |

**RESEARCH\_STEM:**

Table storing a list of stems/prefixes used in research codes.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **KEYS** | **COLUMN\_NAME** | **DATA\_TYPE** | **NULLABLE** | **COMMENT** |
| PK | RES\_STEM\_ID | NUMBER(9,0) | NOT NULL | Primary key. Unique ID for each research code stem. |
| UK | RESEARCH\_STEM | VARCHAR2(20) |  | The actual stem/prefix used in the research code. |

**SITE\_COMPONENTS:**

Table defining the location of the binding sites in the binding\_sites table. A binding site could be defined in terms of which protein subunits (components) are involved, the domains within those subunits to which the compound binds, and possibly even the precise residues involved. For a target where the binding site is at the interface of two protein subunits or two domains, there will be two site\_components describing each of these subunits/domains.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **KEYS** | **COLUMN\_NAME** | **DATA\_TYPE** | **NULLABLE** | **COMMENT** |
| PK | SITECOMP\_ID | NUMBER(9,0) | NOT NULL | Primary key. |
| FK,UK | SITE\_ID | NUMBER(9,0) | NOT NULL | Foreign key to binding\_sites table. |
| FK,UK | COMPONENT\_ID | NUMBER(9,0) |  | Foreign key to the component\_sequences table, indicating which molecular component of the target is involved in the binding site. |
| FK,UK | DOMAIN\_ID | NUMBER(9,0) |  | Foreign key to the domains table, indicating which domain of the given molecular component is involved in the binding site (where not known, the domain\_id may be null). |
|  | SITE\_RESIDUES | VARCHAR2(2000) |  | List of residues from the given molecular component that make up the binding site (where not know, will be null). |

**SOURCE:**

Table showing source from which ChEMBL activity data is derived (e.g., literature, deposited datasets etc)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **KEYS** | **COLUMN\_NAME** | **DATA\_TYPE** | **NULLABLE** | **COMMENT** |
| PK | SRC\_ID | NUMBER(3,0) | NOT NULL | Identifier for each source (used in compound\_records and assays tables) |
|  | SRC\_DESCRIPTION | VARCHAR2(500) |  | Description of the data source |
|  | SRC\_SHORT\_NAME | VARCHAR2(20) |  | A short name for each data source, for display purposes |

**STRUCTURAL\_ALERT\_SETS:**

Table showing list of sets of structural alerts that have been included in COMPOUND\_STRUCTURAL\_ALERT table. It should be noted some alerts/alert sets are more permissive than others and may flag a large number of compounds. Results should be interpreted with care, depending on the use-case, and not treated as a blanket filter (e.g., around 50% of approved drugs have 1 or more alerts from these sets).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **KEYS** | **COLUMN\_NAME** | **DATA\_TYPE** | **NULLABLE** | **COMMENT** |
| PK | ALERT\_SET\_ID | NUMBER(9,0) | NOT NULL | Unique ID for the structural alert set |
| UK | SET\_NAME | VARCHAR2(100) | NOT NULL | Name (or origin) of the structural alert set |
|  | PRIORITY | NUMBER(2,0) | NOT NULL | Priority assigned to the structural alert set for display on the ChEMBL interface (priorities >=4 are shown by default). |

**STRUCTURAL\_ALERTS:**

Table storing a list of structural features (encoded as SMARTS) that are potentially undesirable in drug discovery context. It should be noted some alerts/alert sets are more permissive than others and may flag a large number of compounds. Results should be interpreted with care, depending on the use-case, and not treated as a blanket filter (e.g., around 50% of approved drugs have 1 or more alerts from these sets).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **KEYS** | **COLUMN\_NAME** | **DATA\_TYPE** | **NULLABLE** | **COMMENT** |
| PK | ALERT\_ID | NUMBER(9,0) | NOT NULL | Primary key. Unique identifier for the structural alert |
| FK,UK | ALERT\_SET\_ID | NUMBER(9,0) | NOT NULL | Foreign key to structural\_alert\_sets table indicating which set this particular alert comes from |
| UK | ALERT\_NAME | VARCHAR2(100) | NOT NULL | A name for the structural alert |
| UK | SMARTS | VARCHAR2(4000) | NOT NULL | SMARTS defining the structural feature that is considered to be an alert |

**TARGET\_COMPONENTS:**

Links molecular target from the target\_dictionary to the components they consist of (in the component\_sequences table). For a protein complex or protein family target, for example, there will be multiple protein components in the component\_sequences table.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **KEYS** | **COLUMN\_NAME** | **DATA\_TYPE** | **NULLABLE** | **COMMENT** |
| FK,UK | TID | NUMBER(9,0) | NOT NULL | Foreign key to the target\_dictionary, indicating the target to which the components belong. |
| FK,UK | COMPONENT\_ID | NUMBER(9,0) | NOT NULL | Foreign key to the component\_sequences table, indicating which components belong to the target. |
| PK | TARGCOMP\_ID | NUMBER(9,0) | NOT NULL | Primary key. |
|  | HOMOLOGUE | NUMBER(1,0) | NOT NULL | Indicates that the given component is a homologue of the correct component (e.g., from a different species) when set to 1. This may be the case if the sequence for the correct protein/nucleic acid cannot be found in sequence databases. A value of 2 indicates that the sequence given is a representative of a species group, e.g., an E. coli protein to represent the target of a broad-spectrum antibiotic. |

**TARGET\_DICTIONARY:**

Target Dictionary containing all curated targets for ChEMBL. Includes both protein targets and non-protein targets (e.g., organisms, tissues, cell lines)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **KEYS** | **COLUMN\_NAME** | **DATA\_TYPE** | **NULLABLE** | **COMMENT** |
| PK | TID | NUMBER(9,0) | NOT NULL | Unique ID for the target |
| FK | TARGET\_TYPE | VARCHAR2(30) |  | Describes whether target is a protein, an organism, a tissue etc. Foreign key to TARGET\_TYPE table. |
|  | PREF\_NAME | VARCHAR2(200) | NOT NULL | Preferred target name: manually curated |
|  | TAX\_ID | NUMBER(11,0) |  | NCBI taxonomy id of target |
|  | ORGANISM | VARCHAR2(150) |  | Source organism of molecuar target or tissue, or the target organism if compound activity is reported in an organism rather than a protein or tissue |
| FK,UK | CHEMBL\_ID | VARCHAR2(20) | NOT NULL | ChEMBL identifier for this target (for use on web interface etc) |
|  | SPECIES\_GROUP\_FLAG | NUMBER(1,0) | NOT NULL | Flag to indicate whether the target represents a group of species, rather than an individual species (e.g., 'Bacterial DHFR'). Where set to 1, indicates that any associated target components will be a representative, rather than a comprehensive set. |

**TARGET\_RELATIONS:**

Table showing relationships between different protein targets based on overlapping protein components (e.g., relationship between a protein complex and the individual subunits).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **KEYS** | **COLUMN\_NAME** | **DATA\_TYPE** | **NULLABLE** | **COMMENT** |
| FK | TID | NUMBER(9,0) | NOT NULL | Identifier for target of interest (foreign key to target\_dictionary table) |
|  | RELATIONSHIP | VARCHAR2(20) | NOT NULL | Relationship between two targets (e.g., SUBSET OF, SUPERSET OF, OVERLAPS WITH) |
| FK | RELATED\_TID | NUMBER(9,0) | NOT NULL | Identifier for the target that is related to the target of interest (foreign key to target\_dicitionary table) |
| PK | TARGREL\_ID | NUMBER(9,0) | NOT NULL | Primary key |

**TARGET\_TYPE:**

Lookup table for target types used in the target dictionary

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **KEYS** | **COLUMN\_NAME** | **DATA\_TYPE** | **NULLABLE** | **COMMENT** |
| PK | TARGET\_TYPE | VARCHAR2(30) | NOT NULL | Target type (as used in target dictionary) |
|  | TARGET\_DESC | VARCHAR2(250) |  | Description of target type |
|  | PARENT\_TYPE | VARCHAR2(25) |  | Higher level classification of target\_type, allowing grouping of e.g., all 'PROTEIN' targets, all 'NON-MOLECULAR' targets etc. |

**TISSUE\_DICTIONARY:**

Table storing information about tissues used in assays.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **KEYS** | **COLUMN\_NAME** | **DATA\_TYPE** | **NULLABLE** | **COMMENT** |
| PK | TISSUE\_ID | NUMBER(9,0) | NOT NULL | Primary key, numeric ID for each tissue. |
| UK | UBERON\_ID | VARCHAR2(15) |  | Uberon ontology identifier for this tissue. |
|  | PREF\_NAME | VARCHAR2(200) | NOT NULL | Name for the tissue (in most cases Uberon name). |
| UK | EFO\_ID | VARCHAR2(20) |  | Experimental Factor Ontology identifier for the tissue. |
| FK,UK | CHEMBL\_ID | VARCHAR2(20) | NOT NULL | ChEMBL identifier for this tissue (for use on web interface etc) |
|  | BTO\_ID | VARCHAR2(20) |  | BRENDA Tissue Ontology identifier for the tissue. |
|  | CALOHA\_ID | VARCHAR2(7) |  | Swiss Institute for Bioinformatics CALOHA Ontology identifier for the tissue. |

**USAN\_STEMS:**

Table storing definitions for stems used in USANs (United States Adopted Names).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **KEYS** | **COLUMN\_NAME** | **DATA\_TYPE** | **NULLABLE** | **COMMENT** |
| PK | USAN\_STEM\_ID | NUMBER(9,0) | NOT NULL | Numeric primary key. |
| UK | STEM | VARCHAR2(100) | NOT NULL | Stem defined for use in United States Adopted Names. |
| UK | SUBGROUP | VARCHAR2(100) | NOT NULL | More specific subgroup of the stem defined for use in United States Adopted Names. |
|  | ANNOTATION | VARCHAR2(2000) |  | Meaning of the stem (e.g., the class of compound it applies to). |
|  | STEM\_CLASS | VARCHAR2(100) |  | Indicates whether stem is used as a Prefix/Infix/Suffix. |
|  | MAJOR\_CLASS | VARCHAR2(100) |  | Protein family targeted by compounds of this class (e.g., GPCR/Ion channel/Protease) where known/applicable. |
|  | WHO\_EXTRA | NUMBER(1,0) |  | Stem not represented in USAN list, but added from WHO INN stem list (where set to 1). |

**VARIANT\_SEQUENCES:**

Table storing information about mutant sequences and other variants used in assays. The sequence provided is a representative sequence incorporating the reported mutation/variant used in the assay - it is not necessarily the exact sequence used in the experiment.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **KEYS** | **COLUMN\_NAME** | **DATA\_TYPE** | **NULLABLE** | **COMMENT** |
| PK | VARIANT\_ID | NUMBER(9,0) | NOT NULL | Primary key, numeric ID for each sequence variant. |
| UK | MUTATION | VARCHAR2(2000) |  | Details of variant(s) used, with residue positions adjusted to match provided sequence. |
| UK | ACCESSION | VARCHAR2(25) |  | UniProt accesion for the representative sequence used as the base sequence (without variation). |
|  | VERSION | NUMBER(9,0) |  | Version of the UniProt sequence used as the base sequence. |
|  | ISOFORM | NUMBER(9,0) |  | Details of the UniProt isoform used as the base sequence where relevant. |
|  | SEQUENCE | CLOB |  | Variant sequence formed by adjusting the UniProt base sequence with the specified mutations/variations. |
|  | ORGANISM | VARCHAR2(200) |  | Organism from which the sequence was obtained. |

**VERSION:**

Table showing release version and creation date for the database

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **KEYS** | **COLUMN\_NAME** | **DATA\_TYPE** | **NULLABLE** | **COMMENT** |
| PK | NAME | VARCHAR2(20) | NOT NULL | Name of release version |
|  | CREATION\_DATE | DATE |  | Date database created |
|  | COMMENTS | VARCHAR2(2000) |  | Description of release version |