# DrugBank数据库分析说明书

## 数据库访问网址（以备检查使用，直接填写即可）

<https://www.drugbank.ca/>

## 数据库说明（数据库的介绍，使用范围，包含的数据等内容介绍）

DrugBank 数据库是一个全面、可自由访问的在线数据库, 包含有关毒品和药物目标的信息。作为生物信息学和信息学资源, DrugBank 将详细的药物 (即化学、药理和药物) 数据与综合药物靶 (即序列、结构和通路) 信息结合起来。由于其广泛的范围, 全面的参考和异常详细的数据说明, DrugBank 更类似于药物百科全书比药物数据库。因此, 在维基百科中列出的几乎所有药物都有 DrugBank 链接。DrugBank 广泛应用于制药业、医药化学家、药剂师、内科医师、学生和广大公众。其广泛的药物和药物目标数据使许多现有药物的发现和重新调整用途得以治疗罕见和新发现的疾病。

最新发布的 DrugBank (版本 5.0.11, 发布 2017-12-20) 包含10950药物条目, 包括2367批准的小分子药物, 929 批准的生物技术 (蛋白质/肽) 药物, 109 营养和超过5086的实验药物。此外, 4897 非冗余蛋白 (即药物靶/酶/转运体/载体) 序列与这些药物条目相联系。每个 DrugCard 条目包含超过200数据字段, 其中一半的信息用于药物/化学数据, 另一半用于药物目标或蛋白质数据。

## 数据库具体字段说明（说明数据库每个字段包含的意义，如果是XML文件，那么说明每个XML节点中字段的全部含义，即说明数据库中每个字段的对应含义）

DrugBank数据库采用统一下载模式，下载文件格式为XML。

（1）药物文件

|  |  |  |  |
| --- | --- | --- | --- |
| 字段 | 描述 | 小分子？ | 生物？ |
| 版本 | 入库版本 | Y | Y |
| 创建日期 | 创建条目的日期/时间 | Y | Y |
| 更新日期 | 上次更新条目的日期/时间 | Y | Y |
| DrugBank编号  （主入库编号） | 唯一的DrugBank入库编号,由2字母前缀(DB)和5数字后缀组成。此ID用于通过 URL访问药物输入。如果一个条目被删除,它的DrugBank ID将不会被再次使用 | Y | Y |
| 次要的入库编号 | DrugBank发布1.0的入库编号由4字母前缀和一个5数字后缀组成。每个加入号码是唯一的药物的通用名称。4字母后缀 (APRD, EXPT, BIOD, NUTR) 表明药物的类型(APRD=批准的小分子药物, EXPT=实验药物,BIOD=生物技术药物,NUTR=保健或天然产品) | Y | Y |
| 名称 | 药品制造商提供的药品标准名称 | Y | Y |
| 类型 | 可以是一个小分子或生物技术 | Y | Y |
| 组 | 可以是一个或多个：批准，兽医批准，保健，非法，撤回，研究和实验。 | Y | Y |
| 描述 | 描述一般事实、成分或制剂的药物。 | Y | Y |
| 界 | 有机或无机 | Y | Y |
| 类 | 药物类别是分类系统的主要组成部分，同一类别的药物在结构上被认为是相似的。 | Y | Y |
| 子结构和功能组 | 从药物的结构计算所有的子结构和功能组，这是类的超集 | Y | Y |
| 同义词 | 药物的替代名称 | Y | Y |
| 品牌名称 | 不同制造商的品牌名称 | Y | Y |
| 品牌混合 | 含有特定药物的混合物的品牌名称和成分 | Y | Y |
| 制造商 | 已知的制造商生产的药物 | Y | Y |
| 包装公司 | 包装和销售给定药品的公司 | Y | Y |
| 价格 | 单位药价（以美元计价） | Y | Y |
| 专利 | 药品专利 | Y | Y |
| IUPAC 名称 | 药物的 IUPAC 或标准化学名称 | Y | N |
| 化学式 | 描述原子或元素组成的化学公式 | Y | Y |
| 平均分子量 | 由分子式或序列确定的分子量（g / mol） | Y | Y |
| 单一同位素的分子量 | 一个分子中原子质量的总和，使用每个元素的原子（最丰富）同位素的未结合基态静止质量来代替同位素平均质量 | Y | N |
| 结构 | 对于小分子药物,二维化学结构包括下载和查看各种格式的结构的链接；  对于生物技术药物, 三维结构的图像和链接来查看结构查看器中的三维结构 | Y | Y |
| SMILES | 与药物结构相对应的异构SMILES链 | Y | N |
| InChI | 标准 InChI 标识符 | Y | N |
| InChI 键 | 标准 InChI 键 | Y | N |
| CAS （美国化学文摘社）注册号 | 化学文摘服务标识编号 | Y | Y |
| KEGG 药号 | 京都的基因和染色体组药物识别号 (如果分子在 KEGG中) | Y | Y |
| KEGG复合编号 | 《京都基因与染色体大全》复合识别号 (如果分子在 KEGG中) | Y | Y |
| PubChem 复合编号 | NCBI 的 PubChem 数据库复合识别号 | Y | Y |
| PubChem 物质编号 | NCBI 的 PubChem 数据库物质识别号 | Y | N |
| ChemSpider编号 | ChemSpider 识别号 | Y | N |
| BindingDB编号 | BindingDB 识别号 | Y | Y |
| ChEBI 编号 | EBI 的生物兴趣鉴定编号 (如果药物在ChEBI 中) | Y | N |
| ChEMBL编号 | ChEMBL 识别号 | Y | N |
| Stitch 编号 | Stitch 识别号 | Y | Y |
| 治疗目标数据库 (TTD)编号 | 治疗目标数据库识别号 | Y | Y |
| PharmGKB编号 | 药物基因组学知识库识别编号(如果分子在 PharmGKB) | Y | Y |
| HET编号 | Pdb编号(如果分子在pdb、ChemPDB 或配体集中) | Y | N |
| UniProt编号/名称 | UniProt数据库编号 | N | Y |
| GenBank编号 | GenBank数据库编号 | N | Y |
| 药品编号 | 药品数据库药物识别号 | Y | Y |
| RxList链接 | 链接到给定药物的RxList条目(如果存在) | Y | Y |
| Drugs.com链接 | 链接到Drugs.com为给定的药物(如果它存在) | Y | Y |
| PDRhealth链接 | 链接到给定药物的PDRhealth条目(如果存在) | Y | Y |
| 维基百科链接 | 链接到维基百科词条为特定药物(如果它存在) | Y | Y |
| FDA标签 | 食品药品监督管理局批准标签(如果存在) | Y | Y |
| 材料安全数据表 | 材料安全数据表(MSDS)(如果存在) | Y | Y |
| 合成参考 | 药物合成说明的参考或专利号 | Y | Y |
| 状态 | 物理状态(固体、液体、气体) | Y | Y |
| 熔点 | 熔点（摄氏度） | Y | Y |
| 沸点 | 沸点（摄氏度） | Y | Y |
| 实验的水溶性 | 水溶性（mg/mL或g/L） | Y | Y |
| 预测的水溶性 | 预测的水溶性（mg/ml） | Y | N |
| 实验LogP/疏水性 | 如果是蛋白质/肽，水/辛醇分配系数（如果小分子）或疏水性评分（Gravy 评分） | Y | Y |
| 预测LogP/疏水性 | 预测水/辛醇分配系数 | Y | N |
| 预测日志 | 预测记录(水溶性) | Y | N |
| 实验记录 | 实验记录(水溶性) | Y | Y |
| 实验Caco2渗透性 | Caco-2渗透系数 | Y | Y |
| pKa/等电位点 | 如果小分子，蛋白质没有电荷的pH（如果蛋白质药物）的解离常数（pKa） | Y | Y |
| 质谱 | 药物的EI质谱图像 (如果存在) | Y | N |
| 实验 PDB 条目 | pdb中存在药物结构的pdb标识号 | Y | Y |
| 类别 | 治疗类或一般类药物(抗惊厥药，抗菌药等) | Y | Y |
| ATC编码 | 世卫组织药物分类系统(ATC)标识符 | Y | Y |
| AHFS编码 | AHFS药物信息标识符 | Y | Y |
| 迹象 | 药物用于治疗的疾病的描述或常见名称 | Y | Y |
| 药效 | 药物在临床或生理层面上的工作原理描述 | Y | Y |
| 行动机制 | 药物的工作原理和分子水平的描述 | Y | Y |
| 吸收 | 描述药物的多少或如何容易的药物是由身体采取 | Y | Y |
| 毒性 | 来自测试动物的致死剂量LD 50）值，人类所见的副作用和毒性作用的描述 | Y | Y |
| 蛋白质结合 | 在血浆蛋白中结合的药物的百分比 | Y | Y |
| 代谢 | 药物被中和的机构或器官位置 | Y | Y |
| 半衰期 | 药物在体内的半衰期，以小时或天数计量 | Y | Y |
| 消除路线 | 清除药物的路线，药物主要由肝脏和肾脏清除 | Y | Y |
| 分配量 | 分布的表观体积是流体的理论体积，总药物必须被稀释以产生血浆​​中的浓度 | Y | Y |
| 清除 | 清除是一个描述性的术语，用来评估药物从体内清除的效率 | Y | Y |
| 剂型 | 如何配发药物 (片剂、胶囊、溶液等) | Y | Y |
| 药物相互作用 | 在服用这种药物时，已知的相互作用、干扰或引起不良反应的药物 | Y | Y |
| 食物互动 | 当服用这种药物时，已知的相互作用、干扰或引起不良反应的食物 | Y | Y |
| 途径 | 药物通路中所涉及的药物 | Y | Y |
| 受影响的生物 | 药物最有效的生物体名称 | Y | Y |
| 一般参考 | 一般在线参考有关药物的其他细节 | Y | Y |

（2）目标/酶/转运蛋白/载体文件

|  |  |
| --- | --- |
| 领域 | 描述 |
| 名称 | 蛋白质或大分子（或其他小分子）的名称 |
| 基因名称 | 基因名称 |
| 同义词 | 替代名称（蛋白名称，缩写等） |
| 蛋白质序列 | 氨基酸序列 |
| 残留物数量 | 蛋白质序列中的氨基酸数量 |
| 分子量（道尔顿） | 以道尔顿或g/mol给出的分子量 |
| 理论等点位 | 理论等电位点 |
| 基因本体分类 | 基因本体分类包括功能，细胞过程和定位 |
| 一般功能 | 简短的3-4个字的主要功能总结 |
| 具体功能 | 详细30-40字的具体功能总结 |
| 途径 | 关键路径或过程（来自SMPD）给定分子参与 |
| 反应 | 给定分子参与的反应 |
| Pfam域功能 | PFAM域的名称和ID号 |
| 信号 | 信号肽或其他定位信号在序列中的位置 |
| 跨膜区域 | 跨膜螺旋的数量和位置 |
| 必要性 | 就没有它的生物体的生存能力而言，它的重要性或本质性 |
| GenBank编号蛋白 | GenBank蛋白质编号（如果存在） |
| UniProt编号/名称 | UniProt编号（如果存在） |
| PDB编号 | PDB编号（如果存在） |
| 细胞位置 | 给定的蛋白质或大分子在细胞内或细胞周围的位置（细胞质，细胞核，膜等） |
| 基因序列 | 给定分子的DNA序列（来自cDNA） |
| GenBank编号基因 | GenBank数据库基因标识符和链接 |
| GeneCards编号 | GeneCards数据库标识符和链接 |
| GenAtlas编号 | GenAtlas数据库标识符和链接 |
| HGNC编号 | HGNC数据库标识符和链接 |
| 染色体位置 | 分子在23个人类染色体中的任何一个上的位置（如果分子是细菌，则不给定位置） |
| 轨迹 | 基因染色体位置的更详细的位置 |
| 参考 | Pubmed参考 |
| 药物参考 | 给定的药物协会的具体参考 |

## 数据库进行更新时，数据更新方式（全库更新或增量更新，更新时数据是否包含1、新增数据；2、修改数据、3、删除数据）

数据库采用全库更新方式。

## 数据下载方式，数据更新后的下载地址（下载地址每次有变化或者下载的规律。下载下来的文件是单个文件还是压缩包。数据库更新的频率或每次更新的日期。）

数据下载前需要先进行登录操作，之后可以点击“Download”下载所有文件，下载之后的文件是单个文件，可以在不同的下载界面中下载不同版本的数据库进行使用，数据库对外公开是每天更新，数据库下载每季度发布。

下载网页地址：<https://www.drugbank.ca/releases/latest>

登录账户：313211840@qq.com

密码：bio204