# PDBTM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）

<http://pdbtm.enzim.hu/>

## 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）

PDBTM是一个已知结构的跨膜蛋白质数据库。它旨在收集存储在蛋白质结构数据库（Protein Data Bank，PDB）中的所有跨膜蛋白质，并确定它们的跨膜区域。这些分配是基于TMDET算法，该算法仅使用结构信息来定位脂质双层最可能的位点，并区分跨膜的和球状的蛋白质。这个算法被应用到所有PDB条目，结果被收集到PDBTM数据库中。通过使用TMDET算法，PDBTM数据库能够每周自动的更新，保持和最新的PDB更新同步。

PDBTM web 服务器提供了通过下载菜单下载的各种文件。第一组下载提供了各种原始的PDBTM xml文件。有分别为α-螺旋或β-桶跨膜蛋白的 xml 文件 (pdbtmalpha，pdbtmbeta分别), 或两者都在一个文件 (pdbtmall)。文件 “pdbtm”包含在PDB数据库中的所有条目,包括跨膜和非跨膜蛋白。

## 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）

PDBTM 数据库使用xml文件格式存储由TMDET算法生成的数据，下面为每个字段的信息：

|  |  |
| --- | --- |
| 字段 | 描述 |
| 版权记录（COPYRIGHT record） | 版权声明 |
| 创建日期记录（CREATE\_DATE record） | 创建条目的日期 |
| 修改记录（MODIFICATION record） | 对现有项所作修改的说明，该记录包含两个子记录：修改日期和修改说明 |
| RAWARES记录（RAWRES record） | 包含由 TMDET 算法确定的一些主要数据 |
| 生物矩阵记录（BIOMATRIX record） | 包含用于生成“生物分子”的基质转换，即已显示(或被认为)的蛋白质的低聚物结构；  转换应该应用于在<APPLY\_TO\_CHAIN>记录中定义的链，在这个链记录中，新生成链的链标识由<NEW\_ID>属性给出；  转换是在<TMATRIX>记录中定义的。<TMATRIX>是通常的4x3 转换矩阵。通过使用<ROWX> X的表示法，XY 为<ROWX> Y,对于<ROWZ>T,可以通过以下公式从链 ID 的坐标生成链NEW\_ID的原子坐标 |
| 膜记录（MEMBRANE record） | 记录包含了膜相对于分子最可能的位置的信息。这是由一个变换矩阵给出的，它以这样一种方式变换了分子的坐标，使得膜面与XY平面平行，而原点在薄膜的中间。正常的记录包含了膜平面法向量的数据。由于矩阵变换，法向量的X和Y分量应该为零或接近于零。法向量的Z分量是膜宽度的一半 |
| 链记录（CHAIN record） | 使用生物矩阵生成的或在pdb文件中列出的所有蛋白质链都有一个链记录，这个记录包含三种属性：CHAINID，NUM\_TM，TYPE  每个链记录包含一个或多个区域记录，它们在相对于薄膜的空间中定位链段  pdb\_beg和pdb\_end属性包含使用pdb编号的段本地化, 而seq\_beg和seq\_end则使用在<SEQ>记录中找到的序列中的编号。<SEQ>记录中的序列由对齐方式生成 |

## 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）

因考虑到PDBTM数据库中数据量小，且处理的文件只有一个xml格式的文件，所以采用全局更新方式。

## 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。）

<http://pdbtm.enzim.hu/data/pdbtmall>

下载地址没有变化，点击下载后为XML类型格式的文件，PDBTM 数据库将在 PDB 数据库更新后的每个星期四更新。