

PDB 简介

陈成 (5110809002)

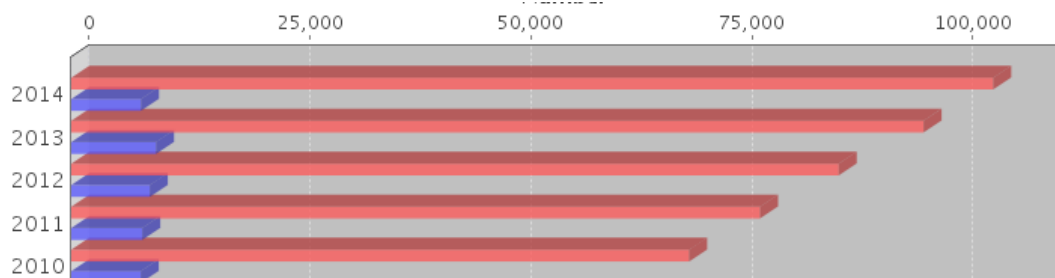
随着现代社会发展,数据库在科学研究和日常生活中有着越来越重要的地位,尤其在计算机辅助药物设计这种计算机相关学科中,其作用更是不言而喻的。在众多数据库中,PDB 是一个对计算机辅助药物设计相当重要的数据库。

PDB 中含有通过实验测定的生物大分子结构,其中主要是蛋白质三维结构,还包括核算、糖类、蛋白质核算复合物等。PDB 中的每条记录都有显示序列信息和隐式序列信息两条信息。SEQRES 最为显示序列标记,以这个关键词大头的每一行都是序列信息。其中隐式序列信息即为立体化学结构。包括原子的名字和三维坐标。因为 PDB 主要记录的三位信息,如何直接返回文本信息的话,不仅不够形象不好理解,而且难以分析,所以可以用分子模拟软件,以图象的形式显示三维结构。现在的主流这种模拟软件都支持 pdb 格式文件的输入,也在一定程度上反应了 PDB 使用范围之广。PDB 中现在有 104371 个结构,分布大概如下:

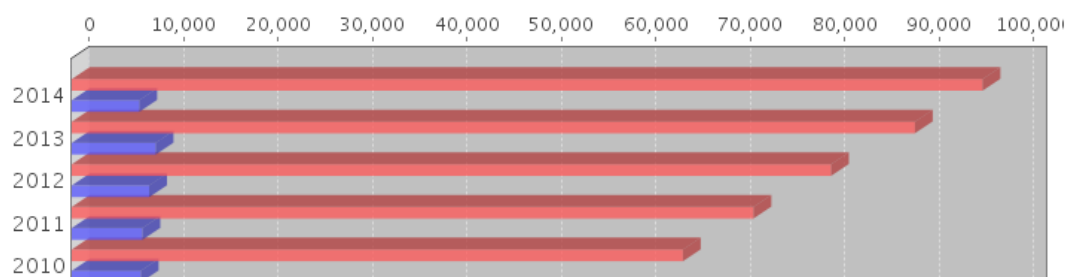
PDB Current Holdings Breakdown

Exp.Method	Proteins	Nucleic Acids	Protein/NA Complexes	Other	Total
X-RAY	86337	1576	4682	5	92600
NMR	9333	1099	220	7	10659
ELECTRON MICROSCOPY	599	67	193	0	859
HYBRID	64	3	2	1	70
other	160	4	6	13	183
Total	96493	2749	5103	26	104371

可以看出,X 光晶体衍射还是得到蛋白质结构的主要方法。尽管 PDB 现在已经给我们提供了丰富的信息,但是数据量每年仍都稳定增长,这里只截取近几年的数据,红色是总量,蓝色是每年的量:



这些增长中大部分是蛋白质结构,蛋白质数据的增长如下:



PDB 中还有很多已经发现,但是由于许多原因尚未公布的结构:

还有两个基于 PDB 的数据库值得一提。一个是 PDBbind,主要功能是为研究各种生物体系中的分子识别过程提供高质量的素材。它系统地收集了 PDB 数据库中各种类型复合物的三维结构信息以及亲合性实验数据。另一个是 MMDB,是一个分子模型数据库,剔除 PDB 中基于理论计算的模型结构,又加了一些补充信息,还提供生物大分子三位结构模型显示和一些结构分析、比较工具。