

Uniprot

我们能在 uniprot 数据库中查到：酶的来源种属信息（来自哪个物种）、酶蛋白的序列信息，以及可能与之发生相互作用的蛋白质。

所需的输入为：酶的 uniprotID (具有唯一性)、基因名，或者感兴趣的关键字。

相关思路：在比对的时候可以用 uniprot 查询可与用户给出的基团作用的酶的相关信息。（3. 比对阶段）

PubChem

PubChem 数据库是针对化学结构检索有机物信息，也可以通过有机物名称查询其化学式。

可以通过有机物名称查到，包括 PubChem CID、化合物结构、化学安全分类、分子式（MF）、同义词、分子量和数据更新时间等在内的相关有机物的信息概览。

有些有机物可能找不到数据库提供的化学式。

PubChem 的检索词包括：化合物名称、化学式、CAS ID 号、SMILES 和 InChI 表达式，或基因名。

相关思路：用户可以给出化合物名称或者 SMILES，后台通过这些信息查询 PubChem 得到对应有机物的信息。（1. 数据导入阶段）

KEGG

API - 可以使用 Biopython.Bio.KEGG 解析从 KEGG 查到的记录（将查询结果解析成 Python 中特定类的对象，可以通过访问其属性来得到酶的相关属性）。

爬取 - 合法性有待商榷，要爬取的话可以考虑用 ExplorEnz Database 作为替换，提供 sql 格式的下载。

BRENDA

API - 利用 Python API 查询 BRENDA 需要提供邮箱、密码、EC 编码等信息，返回 JSON 格式的数据。

疑问：能不能通过输入有机物和基团来查询能与之反应的酶（？），而不需要输入酶的 EC 编码。

爬取 - 可以爬取，得到 .txt 文件，但含有大量缩写，需要解析器作进一步解析以获得可供 Python 直接处理的数据。

文件传输

可以用 drf 中提供的 FileField 字段传输文件，在定义后端 API 时直接使用即可。（1. 数据导入阶段接收前端的文件，5. 呈现结果时往前端传文件 不知道前端可不可以解析文件）

SMILES

在 Python 中的处理与普通字符串相似，考虑通过字符串切片来表示某种基团，比如：令 `s='CC(=O)O'`，通过 `s[1:7]` 来表示羧基。（1. 数据导入阶段，用于确定基团类型与用户需要的反应类型）

pysmiles 可以解析 SMILES 以获得元素之间的键连关系。（也许用得上）

rdkit

rdkit 可以将 SMILES 和 .mol 文件相互转化，如果前端觉得 .mol 更方便的话可以用 rdkit 加以解析。（1. 数据导入阶段）