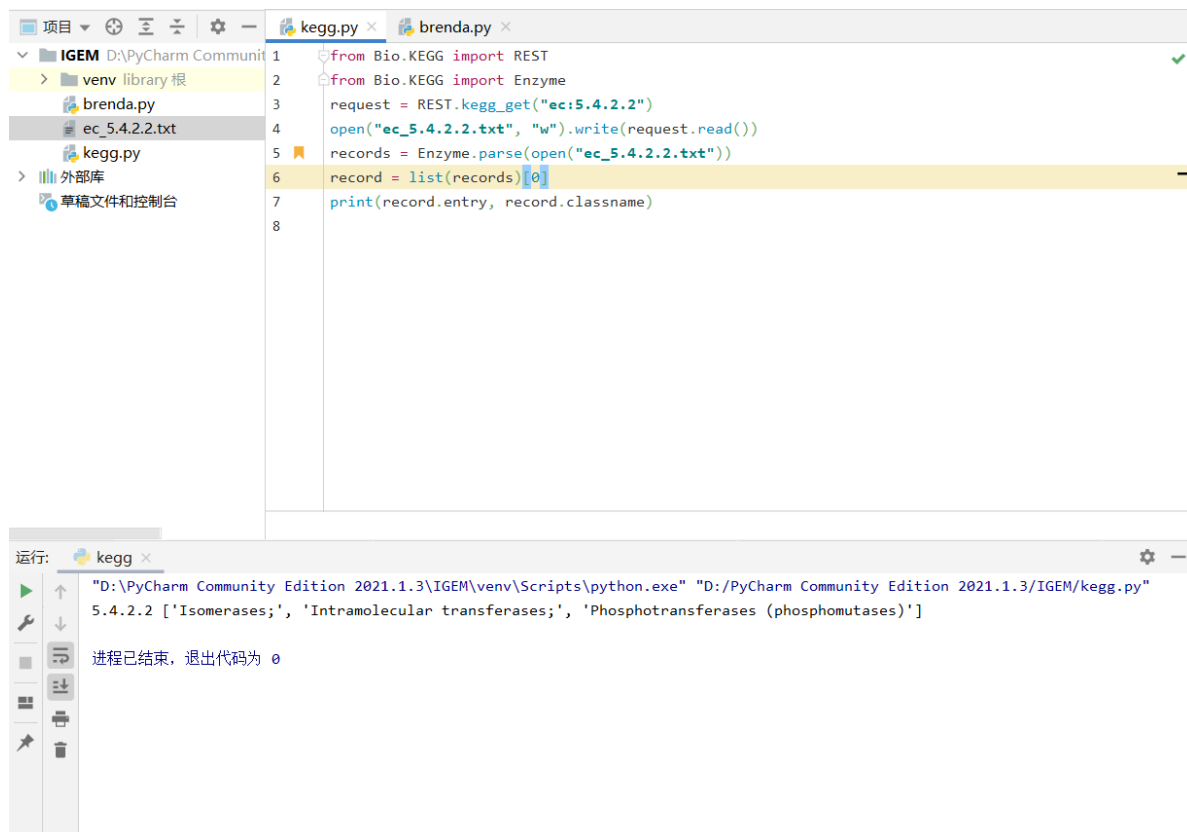


KEGG 调研报告

在 Python 中，可以使用 Biopython 来解析 KEGG 记录。Biopython 中名为 Bio.KEGG 的包中提供了几个处理所得记录的子包：

- Bio.KEGG.Compound：保存来自 KEGG 中配体或化合物的记录。可以获取以下内容：
 - entry：条目编号
 - name：化合物学名
 - formula：分子式
 - mass：质量
 - pathway：通路
 - enzyme：酶
 - structures：结构
 - dblinks：数据库有关的链接
- Bio.KEGG.Enzyme：保存来自 KEGG 中酶的记录。
- Bio.KEGG.Gene：保存来自 KEGG 中基因的记录。
- Bio.KEGG.KGML：保存来自 KEGG 中通路的记录。
- Bio.KEGG.Map：加载 KEGG 中的通路。
- Bio.KEGG.REST：提供使用 REST API 访问 KEGG 数据库的方法。
 - Bio.KEGG.REST.kegg_info：显示数据库发布信息和链接数据库信息
 - Bio.KEGG.REST.kegg_list：获取条目编号和相关的定义
 - Bio.KEGG.REST.kegg_find：通过查询关键字或其他数据找到匹配条目
 - Bio.KEGG.REST.kegg_get：检索给定的数据库条目
 - Bio.KEGG.REST.kegg_conv：将 KEGG 中的编号转化为外部编号
 - Bio.KEGG.REST.kegg_link：通过使用数据库前后参照找到相关条目



```
1 from Bio.KEGG import REST
2 from Bio.KEGG import Enzyme
3 request = REST.kegg_get("ec:5.4.2.2")
4 open("ec_5.4.2.2.txt", "w").write(request.read())
5 records = Enzyme.parse(open("ec_5.4.2.2.txt"))
6 record = list(records)[0]
7 print(record.entry, record.classname)
8
```

运行: kegg ×

```
"D:\PyCharm Community Edition 2021.1.3\IGEM\venv\Scripts\python.exe" "D:/PyCharm Community Edition 2021.1.3/IGEM/kegg.py"
5.4.2.2 ['Isomerases;', 'Intramolecular transferases;', 'Phosphotransferases (phosphomutases)']

进程已结束，退出代码为 0
```

通过我的一些分析，我认为日后主要使用 Bio.KEGG.REST 来进行相关工作，因为它是直接对应到 KEGG 官方的 REST API 中，功能应该比较全面。

BRENDA 调研报告

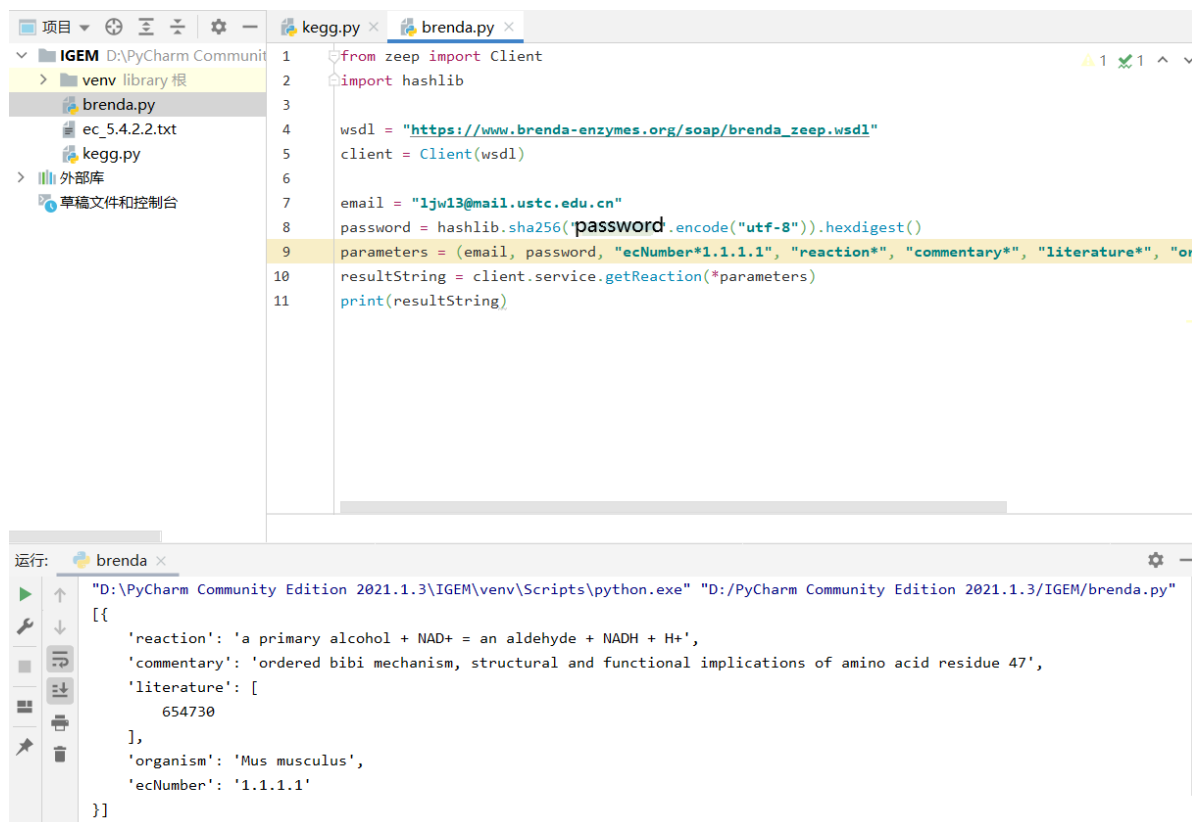
关于如何使用 BRENDA 的博客比较少。主要的参考依据只能通过查阅官方文档使用。BRENDA 的功能目前看起来比较全面，查阅的内容比较多，这是官方文档给出的调用接口：

SOAP methods

Ligand structure Id	Expression	Natural Substrate	Purification	Synonyms
Reference by Id	General Information	Natural Substrates Products	Reaction	Systematic Name
Reference by pubmed Id	General Stability	Organic Solvent Stability	Reaction Type	Temperature Optimum
Activating Compound	IC50 Value	Organism	Recommended Name	Temperature Range
Application	Inhibitors	Oxidation Stability	Reference	Temperature Stability
CAS Registry Number	KCat KM Value	Pathway	Renatured	Turnover Number
Cloned	KI Value	PDB	Sequence	
Cofactor	KM Value	pH Optimum	Source Tissue	
Crystallization	Ligands	pH Range	Specific Activity	
Disease	Localization	pH Stability	Storage Stability	
EC Number	Metals Ions	pl Value	Substrate	
Protein Variants (Engineering)	Molecular Weight	Posttranslational Modification	Substrates Products	
Enzyme Names	Natural Product	Product	Subunits	

对于每一个函数方法，都有一个输出注释和输入示例（不过输入示例看起来怪怪的，可能需要自己摸索）。其中的输入格式比较严格。

据官方文档说明，每个用户会有访问频率的限制，这个对我们用户的使用是否会带来极大的不便？（不过在我的使用中并没有感受到）



个人感觉，BRENDA 中可以获取的信息会比较多，但是感觉 BRENDA 上手又会困难一些。

总结

暂时认为两个数据库都是可以使用的。

目前 Bio.KEGG 的示例比较少，所以需要不断尝试如何处理获取的数据，这可能比较耗费时间。

而 BRENDA 处理数据看起来会方便许多，因为 BRENDA 返回的是 JSON 格式的数据。

参考文献

- [利用Biopython来解析KEGG记录](#)
- [Bio.KEGG package](#)
- [使用biopython解析kegg数据库](#)
- [Biopython Documentation](#)
- [Brenda-利用SOAP API访问Brenda及本地保存](#)
- [SOAP access help - BRENDA Enzyme Database](#)