SZC的网络药理学研究

1 SZC化合物数据库的建立

通过检索TCMID（http://www.megabionet.org/tcmid/）、上海有机所中药与化学成分库数据库（http://www.chemcpd.csdb.cn/cmpref/main/tcm\_introduce.asp?n Count=6077992）整理SZC及其各组方药材化学成分，查阅复方及单味药材成分鉴别文献，完善化合物信息。结合安捷伦科技有限公司提供的“formula-database-generator”工具，建立SZC化学成分数据库，包括药材来源、化合物中文名称、化合物英文名称、分子式及精确分子质量。

1. 治疗PD靶标数据库建立

从 TTD 和 Drugbank 数据库中选取与子宫收缩、抗炎、镇痛、调节内分泌等相关的治疗原发性痛经的药物作用靶点，并从 ＲCSB 的 PDB 库中下载其复合物晶体结构，导入Discovery Studio 2.5 软件对这些靶蛋白结构进行清除配体、水及金属离子、加氢和加力场等修饰，建立治疗PD的靶标数据库。

1. SZC有效部位中小分子化合物的分子对接

将前期实验所得化学物质组，导入 1 项下数据库，检索得到SZC有效化学成分的分子结构，与 2 项下修饰后的靶蛋白导入 Autodock 4.0 软件进行分子对接计算，以蛋白晶体结构中原配体的中心坐标为活性位点中心坐标，活性位点为40 ×40 ×40 个格点组成的正立方体，其中对每一维度，两相邻格点之间距离为 0.375 ，并利用拉马克遗传算法(Lamarckian genetie algorithm) 进行小分子在蛋白质配体中活性位点的能量优化，其参数设置如下：遗传算法种群数量为 150；能量优化最大迭代次数为 1.75 ×107 ，并以此作为优化终止条件；对接结果化合物构象分析均方根背离容忍度设定为 2.0；每一个化合物，对接过程重复采样次数设置为 20；其余设置均为软件的默认值。

1. “成分-靶标”网络构建与分析

根据分子对接结果，选取化合物与靶点的对接得分大于 6.5 且高于原配体得分的分子与靶蛋白，采用 Cytoscape 3.4.0 软件构建SZC治疗PD的“成分-靶标”网络图。网络图中的节点（Node）代表小分子化合物或潜在靶点，边（edge）代表化合物与靶标间的相互作用。应用软件中的 Network Analysis plugin 对构建的网络图进行统计，分析节点在网络图中的作用。通过 STRING数据库（https://string-db.org/），对预测得到的靶标蛋白进行通路富集分析和蛋白相互作用分析。

1. 网络药理学验证SZC治疗PD的作用机制

分析实验所得SZC化学成分与对应子宫平滑肌收缩靶点的网络图，结合 4 项下的实验结果，进行对比分析，证实网络药理学分析在SZC多成分多靶点治疗PD物质基础分析过程中提供有效的验证以及指导意义。

参考文献

[1]王韵. 茵陈蒿汤网络药理学与药效物质基础研究[D]. 第二军医大学, 2017.

[2]王俨如，张新庄，李娜，等. 基于网络药理学方法探讨双鱼颗粒治疗上呼吸道感染的作用机制[J]. 中国新药杂志. 2015(11): 1222-1227

中药化学成分数据库：TCM、TCMID、TCM Database @ Taiwan、上海有机所中药与化学成分库等。

药物靶点数据库（Drug Bank、TTD、STITCH 和 HIT 等）、蛋白质相互作用数据库（String 和 Bio GRID 等）、蛋白晶体结构数据库（PDB）、通路数据库（KEGG）等

常用的可视化和分析工具有：Cytoscape、Ingenuity Pathway Anaylsis、Pajek、GUESS 和 Network X

绘图工具：cytoscript

3D结构绘制工具：chembioffice

# 研究目标

药物的吸收（Absorption）、分布（Distribution）、代谢（Metabolism）、排泄

（Excretion）的相关性质（ADME）的研究是当今化学药物研发中的重要内容，

ADME 性质决定了药物的药代动力学特征及其治疗的有效性。

上面的介绍是本文重要的研究目标.这个也是本文进行研究的时候关键的一个部分。

研究的对象主要有这么几个,分别是

# 数据处理

## 药品信息

浙贝母、薏苡仁、三七、龙血竭

上面的四种药品是本文重点研究的对象，需要的数据也是这四种药品的数据。

## 药品数据库

TCMID:

<http://www.megabionet.org/tcmid/>

上海有机所中药与化学成分库数据库http://www.chemcpd.csdb.cn/cmpref/main/tcm\_introduce.asp?n Count=6077992

TCMSP,西北农林科技大学开发的中药系统药理学数据库和分析平台

<http://lsp.nwu.edu.cn/tcmsp.php>

chemmapper:

BATMAN-TCM:

CNKI:中国知网文献数据库

<http://cnki.net/>,主要用来进行文献的查阅

PubMed: 是一个提供生物医学方面的论文搜寻以及摘要，并且免费搜寻的数据库

<https://www.nlm.nih.gov/pubs/factsheets/pubmed.html>

# 技术路线与方法

本文需要参考那篇博士论文,然后在博士论文的基础上进行改进,从而得到技术路线。

# 任务安排

## 数据库的操作

### 数据库设计

数据库的操作主要是把图片写入到数据库中,这个处理的方式其实很简单,首先创建数据库,将图片使用BLOB(最大只能64K)或者是MediumBlob(最大16M)是比较合理的。

表的设计,根据数据的来源,本实验可以把数据的字段分成下面的几个方面:

|  |  |  |  |
| --- | --- | --- | --- |
| 字段 | 类型 | 索引 | 含义 |
| id | int(11) | PRI | 主键自增 |
| source | varchar(20) |  | 数据来源 |
| drugs | varchar(20) |  | 药品名称 |
| chemical | varchar(256) |  | 化学成分 |
| photo | MediumBlob |  | 图片数据 |

根据上面的数据库设计完成数据库的初步设计,然后做的是进行数据的读写操作。

### Java数据写入读取

使用函数写入数据和读取数据的操作，前者很简单，后者却是比较容易处理的，后者实际上是将数据从数据库中读取出来,然后写入到相应的文件夹中。

这个使用java进行相关操作就能完成，下面我们需要做的是进行文件夹中的数据的读取。

### 文件夹数据的读取

文件夹数据的读取与前者是类似的,经过相关的设计同样是可以满足数据的读取需求的。

## 数据确定,数据清洗

## 分析路线的确定(基于小数据的分析)

# 数据分析工具

systemsdock  
在线分子对接工具  
3D结构，基因名PDB ID

数据分析

# 分子对接

## Lipinski 规则

氢键供体数目(以NH 和OH 键数目之和计) 大于5 ;

MWT 大于500 ;lg *P* 大于5 (或*M*lg *P* 大于4. 15) ;

氢键受体数目(以N 和O 数目之和计) 大于10 (这些条件对于生物载体底物类化合物不适用) 。

# 现在遇到的问题

### 文本解析

文本解析中遇到的问题主要是文本语义的理解,中文分词技术等相关的问题.

### 其他博士论文的对接处理

### CNKI知网文献爬虫下载

文本地址解析

窗口转换

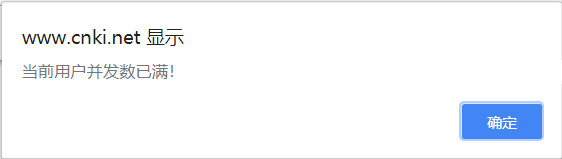
下载的设置,如何直接下来文献到指定的文件夹里

### PUBMED文献数据爬虫

### 分子对接模型的建立

## CNKI知网文献爬虫下载

下面也是经常遇到的问题:



在进行下载的时候,会遇到并发数已满的建议.

通过修改,现在已经可以可以完成这样的文本数据的下载了,同样也能完成本页面的相关的操作信息。

## 建模方案

方案一是使用本文原本计划的方式是使用数据挖掘的方式进行建模,但是根据实际的研究结果,这种方案并不是合理的方案,因为涉及到的数据比较多比较复杂。

方案二是利用现在的处理方式是使用分子对接,也就是3D分子建模的方法进行建模,使用这种

## 分子对接

分子对接涉及到的两个方面,分别是生物大分子和底物。

# 下一步的思路

1. 研究那个论文，在论文研究的基础上,进一步进行相关的实际操作.例如
2. 分子对接继续进行相关的开发推进,使用的方式是**结构对接模型,**但不是基于数据的对接模型。分子对接模型关键是
3. 自然语言处理(NLP)数据分词技术需要推进。
4. 对接工具的使用,分别是discovery studio和AutoDock

步骤:

1)先看论文,阅读论文博士论文中涉及到相关的小论文的部分

2)查看相关的图片转SMILES或者是就简单的获得SMILES的方法.

## 自然语言处理的文献研究

根据查阅到的相关自然语言处理的文献，可以看到,中文的自然语言处理,特别是中文分词技术仍然是一个比较不成熟的研究中的技术,这个技术的使用现在是不现实的。另外,根据中文分词的研究进展,中文自然语言的处理还是一个不是很成熟的技术，实现起来也是一个不很通用的技术。现在直接用来进行解析也是一个很复杂的方案。

## 图片转SMILES的方法

根据前人的研究结果,这里找到了一些研究的进展文献,文献介绍了生成SMILES的方法,方法的介绍都是使用相关的软件来进行的,同时根据研究来看,生成SMILES的方法是使用软件ChemDraw,然后根据这个软件的操作来进行,同时,我们发现某些网站可以根据SMILES来查找具体的化学物质,例如网站[www.chemexper.com](http://www.chemexper.com)。另外，研究使用的是化学式,但不是保证这个是唯一的。这是因为化学式会存在同分异构体等情况。

获得SMILES的软件工程方法有ChemFinder,但是该软件并没有提供自动生动SMILES的控件。根据探究的进展，使用python来解析SMILES（ChemOffice提供了基于Python的ChenScript脚本），但由于ChemScript模块式不正常的,所以该方案不可行。

相关的研究还有一个是使用CAL脚本

参考文献: <http://blog.sciencenet.cn/blog-1213210-1037507.html>

## 对接工具

根据现在的研究进展,现在的对接工具有AUTODOCK，systemsDock，discovery studio这几个可以用来对接的软件,其中, systemsDock是一个在线对接的工具。另外的几个都是经典的分子对接工具,这些工具都有一些比较显著的特点,那就是使用到这些工具必须满足使用3D建模,假如不是3D建模的话,是不能完成对接任务的。

## 分子数据处理工具

现在在进行分子处理的时候,需要对化合物分子进行处理,处理的典型方式是对化合物的结构式进行3D建模,现在处理的软件有ChemDraw和ChemFinder,其中

## 对接工具和分子处理工具

在做分子对接的时候,现在需要注意的是,不同的对接工具和分子处理工具之间是需要注意的。首先,假如使用工具discovery studio来进行对接，那么我们就要使用ChemDraw或者是ChemFinder来进行处理,然后就是另外的AUTODOCK,这个软件需要结合另外的分子数据处理软件来做,但是根据现在的研究进展来看，后者是一个很不合适的方法，软件下载安装都是比较麻烦的,最合适的应该还是前者，使用起来更是很合理,使用Discovery Studio来做对接,使用ChemDraw和ChemFinder来做数据的预处理,这个是很合理合算的.另外,通过相关的研究查看,可以知道,其实也只有这种方式是较为合理的,另外的几种经过试验表明是不可行的，试验的结果宣告失败。

根据初步的探究表明,使用Discovery Studio软件是一个比较合理的方式

## 分子对接实际例子

使用Discovery Studio进行分子对接的时候,

## 分子对接需要的数据库

### NCBI数据库

NCBI是指美国国立生物技术信息中心。理解自然无声但精妙的关于生命细胞的语言是现代分子生物学的要求。