大前提：

1. 使用Python3.10.0 IDLE
2. 不import任何module
3. 使用comment
4. 测试时，将所使用到的.fna，.csv（电脑打开为excel）和.py文件放在同一个文件夹内

背景知识

1. .fna文件格式

>fragment sequence of BRCA1.fna

GAGTCCCGGGAAAGGGACAGGGGGCCCAAGTGATGCTCTGGGGTACTGGCGTGGGAGAGTGGATTTCCGAAGCTGACAGATGGGTATTCTTTGACGGGGGGTAGGGGCGGAACCTGAGAGGCGTAAGGCGTTGTGAACCCTGGGGAGGGGGGCAGTTTGTAGGTCGCGAGGGAAGCGCTGAGGATCAGGAAGGGGGCACTG

不超过80个字符

1. 17种核酸代码

图形用户界面, 文本, 应用程序

描述已自动生成

九个分任务

1. 写一个名为*sequence, description* = **load**(*filename*)的function

打开一个.fna文件，以以下格式打印出DNA sequence（序列）和description（描述文字）

图片包含 应用程序

描述已自动生成

1. filename是str，打开.fna文件
2. 输出的sequence是list，其中每一个核酸代码为str。核酸代码是17个核酸代码中的一个
3. 输出的description是str，不包括>，不包括前前后后多余的空格

其他要求：

如果filename不是.fna或者该文件打不开，输出的sequence为空列表，输出的description为‘file was not found’

如果.fna中的核酸代码有小写字母，要全部转换成大写字母

如果.fna中的核酸代码有0-9，A-Z，a-z的除17种之外的代码，也要打印输出，且小写字母全部转化成大写字母

1. 写一个名为*table* = **stats**(*sequence*)的function

输入一个sequence，输出一个17种核酸代码每种出现几次的table

图形用户界面, 文本, 应用程序, Word

描述已自动生成

1. sequence是list，其中每一个核酸代码为str。核酸代码是17个核酸代码中的一个
2. table是dict，一共有18个entries（不知道怎么翻译）。就是17个核酸代码+1个其他（小写，数字blabla）
3. 写一个名为*formatted\_sequences* = **format\_sequence**(*sequence, first\_index, last\_index*)的function

输入一个sequence和两个index，以以下格式输出一个subsequence

图形用户界面, 文本, 应用程序, 电子邮件

描述已自动生成

1. sequence是list，其中每一个核酸代码为str。核酸代码是17个核酸代码中的一个
2. first\_index和last\_index是第一个和最后一个从sequence种提取的核酸代码的index**（仔细数了一下例子里100到300，它输出了201个数字，说明是包括第300个的，注意list里的index是不包括stop的那个值的）**
3. 输出的formatted\_sequence是list，其中每80个（末位不够80的话就不够）核酸代码为一个str
4. 写一个名为**write**(*filename, description, sequence, first\_index, last\_index*)的function

输入DNA sequence（序列）、range（序列范围）和description（描述文字），以以下格式输出一个.fna文件

图形用户界面, 文本, 应用程序

描述已自动生成

1. filename是str， 用于打开.fna文件
2. description是str，包含写进.fna文件的文本（description text）
3. sequence是list，每个核酸代码是str。核酸代码是17个核酸代码中的一个
4. first\_index和last\_index是第一个和最后一个从sequence种提取的核酸代码的index
5. 写一个名为*matches* = **find**(*sequence, sequence\_to\_find*)的function

在一个sequence中查找是否有另一个sequence存在，输出他们的位置（index）

图形用户界面, 文本, 应用程序

描述已自动生成

1. sequence和sequence\_to\_find是str，**看例子的意思是只有一个index，所有的一长串核酸序列全部在一个‘’里面。**核酸代码是17个核酸代码中的一个
2. matches是list，每个index是int。**需要考虑重复出现的情况**（不能第一次重复检测出来以后就结束了）
3. 写一个名为*new\_sequence* = **add**(*sequence, sequence\_to\_add, index*)的function

添加一个sequence到已知sequence的一个指定位置（index）

1. sequence和sequence\_to\_add是list，**看例子的意思是只有一个index，所有的一长串核酸序列全部在一个‘’里面。**核酸代码是17个核酸代码中的一个
2. index是int，sequence\_to\_add应该**被加在index之前**。

其他要求：

如果index指定的位置是在整个sequence之外（如只有4个数，指定index为7之类的），则sequence\_to\_add被加在sequence末尾

1. 写一个名为*new\_sequence =* **delete***(sequence, index, number\_of\_codes)*的function

在已知sequence中删掉一个subsequence，指明起始点（index）和删除的代码长度（*number\_of\_codes*）

1. sequence和sequence\_to\_add是list，**看例子的意思是只有一个index，所有的一长串核酸序列全部在一个‘’里面。**核酸代码是17个核酸代码中的一个
2. index指明开始删除的起始位置
3. number\_of\_codes指明被删除的核酸代码个数

其他要求：

如果index和number\_of\_code组合起来有超出sequence总长度的部分，则最大程度删到末尾（如只有4个数，但指定从第2个开始删除5个数之类的，就是从第二个开始删到底）

如果index指定的位置是在整个sequence之外，**不删除任何东西**

1. 写一个名为*new\_sequence* = **replace**(*sequence, sequence\_to\_add, index, number\_of\_codes*)的function

用一个subsequence替换已知sequence的一部分

1. sequence和sequence\_to\_add是list，**看例子的意思是只有一个index，所有的一长串核酸序列全部在一个‘’里面。**核酸代码是17个核酸代码中的一个
2. index指明开始删除（被替换）的起始位置
3. number\_of\_codes指明被删除的核酸代码个数

其他要求：

如果index和number\_of\_code组合起来有超出sequence总长度的部分，则最大程度删到末尾（如只有4个数，但指定从第2个开始删除5个数之类的，就是从第二个开始删到底）

如果index指定的位置是在整个sequence之外，则sequence\_to\_add被加在sequence末尾

6-7-8的例子

图形用户界面, 文本, 应用程序, 电子邮件

描述已自动生成

1. 写一个名为*protein\_sequence, table* = **dna2protein**(dna\_sequence)的function

把DNA序列转化成对应的蛋白质序列

日历

中度可信度描述已自动生成

1. dna\_sequence是list，**看例子的意思是只有一个index，所有的一长串核酸序列全部在一个‘’里面。**核酸代码是17个核酸代码中的一个
2. 输出的protein\_sequence是list，每一种蛋白质简写是一个str（具体简写参照dna2protein的excel文件）。如果三个三个组队后有多出来的，不要写进protein\_sequence
3. 输出的table是dict，包括3个字母的核酸代码和1个字母的蛋白质简写

其他要求：

如果有3个字母的核酸代码没有在excel文件中出现，则打印’???’：’?’

这里我实在搞不懂为什么同时print了protein\_sequence和table，但是最终输出的只有table。就按例子来吧

\*\*2-3-4都没有说清楚输入的sequence的形式到底是[‘A’, ‘B’, ‘C’]还是[‘A, B, C’]，我的描述都是按第一种来的，因为有不同的index更好操作。到时候要是说是第二种的话再改。

\*\*1是很明确的打开一个现有.fna文件，里面的格式是这样的：

电脑萤幕的截图

描述已自动生成

\*\*5-6-7-8-9的例子里面都明确说了是第二种，[‘A, B, C’]这样，我确实不知道在一个index里面怎么弄