perl上机练习4

目的：熟悉BioPerl使用和生物信息数据分析编程

输入文件，FASTA序列文件：

/home/shenwk/perl\_practice/test4\_file/3.fasta\_seq.txt（1-5题的输入文件）

1. 计算每条序列的长度，并输出到输出文件。输出形式为两列，第一列为序列名，第二列为序列长度，用tab键分隔。
2. 计算fasta序列文件中每条序列的GC含量（GC数目/序列长度）。
3. 将以上两个输出结果整合成一个三列文件，并在最后一行计算一个汇总值（某列值的总和）。
4. Perl在工作中常用于串流程，现尝试把你上面3个练习写的3个perl脚本分三步将输入文件3.fasta\_seq.txt处理成最终的final.result：

第1步：perl step1.pl 3.fasta\_seq.txt output1

第2步：perl step2.pl 3.fasta\_seq.txt output2

第3步：perl step3.pl output1 output2 final.result

为提高工作效率，现需要写一个脚本使用3.fasta\_seq.txt 作为输入文件，直接得到final.result，中间产生的文件结果不保留。（使用perl编程）

1. 得到每条fasta序列的反向互补序列，输出文件仍是fasta格式。
2. 目的：计算donor位点序列文件

/home/shenwk/perl\_practice/test4\_file/donorseq.txt中各个位点出现四种碱基的频数（每行为一条donor位点的序列，每条序列有21个碱基），要求输出结果为一个矩阵，行名为碱基A、T、C、G。列名为位点号，如可命名为“-9，-8，…………0,1,2…………11”等。