

第三次作业

要求：1. 请根据氨基酸残基的物化性质预测以下蛋白质序列的可能二级结构：

(1) GlyGluLeuGluGluLeuLeuLysLysLeuLysGluLeuLeuLysGly

(2) GlyValThrIleGlnValThrTyrGly

2. 在蛋白质全新设计研究中拟设计一个包含16个残基的稳定两亲性螺旋：

(1) 请参考氨基酸残基的二级结构倾向性设计一个合理的序列，为获得稳定性的螺旋结构需要考虑带电残基的分布以及螺旋偶极的稳定。

(2) N-端第一个残基与C-端第一个残基是否在螺旋的同侧，相差多少度？

(3) 请选用一个蛋白质二级结构预测在线服务器预测所设计序列的二级结构。

3. 蛋白质分子的能量分为键能、键角能、二面角扭转能、非键相互作用能等，你认为对于蛋白质结构稳定性贡献大的主要为哪些项？如果蛋白质中有二硫键，那么二硫键是否会影响蛋白质的三维结构？目前的分子动力学模拟是否可以模拟速度很慢（如秒数量级）的变化？

4. 为什么无法利用能量极小化算法找到蛋白质的能量最低构象？分子动力学或蒙特卡洛模拟退火等方法在分子构象计算中主要解决什么问题？

5. 蛋白质与配体结合过程的主要驱动力包括哪些？在已经形成的复合物中，两者间的主要作用力是什么？

某个蛋白质P在实验条件下可以与配体L结合，但实验测出来的 $\Delta H > 0$ ，你认为驱动结合的主要因素是什么？

问题分析：1. (1) 由Chou-Fasman构象倾向性因子可知，Glu和Leu是强的 α 螺旋形成残基（Glu和Leu具有长的烷基链条，Glu还具有带负电荷的羧基负离子，可以与带正电荷的残基相互作用），Lys是普通的 α 螺旋形成残基（Lys具有长的烷基链条，还具有带正电荷的质子化氨基，可以与带负电荷的残基相互作用），Gly是强的 α 螺旋中断残基（Gly的R基极小，导致Gly构象柔性极大，容易破坏 α 螺旋结构），且N端Glu分布较多，C端Lys分布较多，它们的电荷能与螺旋偶极相抵，因此该序列倾向于形成 α 螺旋。

(2) 由Chou-Fasman构象倾向性因子可知，Val、Ile及Tyr是强的 β 折叠形成残基（Val和Ile具有分叉的烷基链条，Tyr具有芳香环，它们均具有较强的疏水性），Thr和Gln是普通的 β 折叠形成残基（Thr具有羟基，Gln具有酰胺基，它们具有一定亲水性），Gly是普通的 β 折叠中断残基（Gly的R基极小，导致Gly构象柔性极大，容易破坏 β 折叠结构），且疏水残基与亲水残基交替排列，因此该序列倾向于形成 β 折叠。

2. (1) 根据设计要求，我们总结出如下设计要点：(i) 由Chou-Fasman构象倾向性因子可知，Glu、Met、Ala、Leu是强的 α 螺旋形成残基，除此之外，Lys、Phe、Gln、Trp、Ile、Val也是比较容易形成 α 螺旋的残基，在构建稳定的两亲性螺旋时，这些残基可以用于构筑，而Gly是强的 α 螺旋中断残基，可用于在N端和C端中断 α 螺旋；(ii) 两亲性螺旋的螺旋轮应该满足一边全为疏水残基，一边全为亲水残基；(iii) 如果亲水残基带电，通常的设计思路为，尽可能避免带电残基处于相邻位置，以免因为静电排斥作用而导致 α 螺旋不稳定；(iv) 带正电的亲水残基应尽可能在C端，带负电的亲水残基应尽可能在N端，这样能抵消螺旋偶极的影响，使螺旋更稳定。

由此，我们结合以往文献，设计出如下多肽片段：(i) GLKKLKKLLKKLKKLG；(ii)

GKKLKKLLKKLLKLLG；(iii) GLKELKELLKELKELG；(iv) GLEKLEKLLEKLEKLG。此处(i)和(ii)只用到中性和带负电残基，似乎有违上述规则，但查阅文献，发现国外和国内均仅有用Leu和Lys构建稳定两亲性螺旋的实例（见Blondelle, S. E.; Houghten, R. A. *Biochemistry* **1992**, 31, 12688，以及王良，马清泉，单安山，董娜，吕银凤. *微生物学通报* **2014**, 41(2), 312），故(i)和(ii)的设计应该是合理的。同时，利用ExPASy的多肽性质分析工具ProtParam (<https://web.expasy.org/protparam/>)，可以得知

(i) ~ (iv) 的不稳定系数 (Instability Index, 数值越大越不稳定)

为-33.08, -33.08, -11.85, -11.85；疏水性系数 (Grand average of hydropathicity, 数值越负越亲水) 为-0.575, -0.094, -0.475, -0.475，这表明设计的 α 螺旋应该为稳定的两亲性螺旋；多肽的等电点

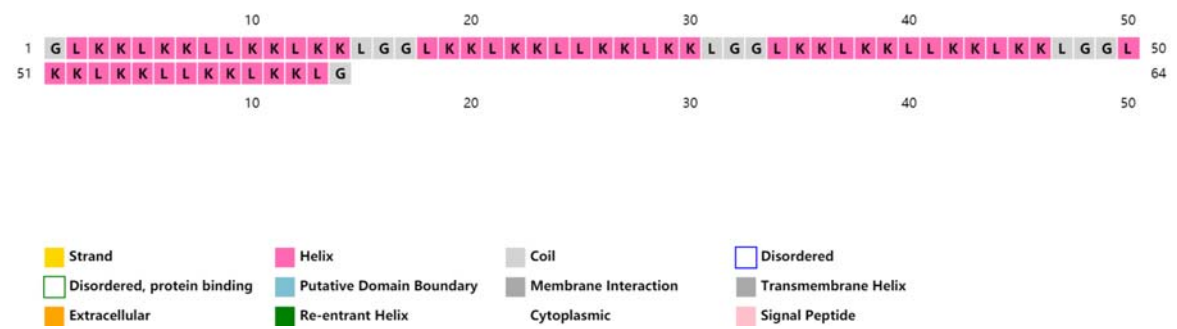
(pI) 分别为10.85, 10.78, 6.29, 6.29，表明前两个螺旋在负电荷较多的条件下容易形成 α 螺旋（如细胞膜表面），后两个螺旋可以在中性略偏酸性的条件下形成 α 螺旋。

(2) N-端第一个残基与C-端第一个残基并不对齐（不在同一条竖直线上），事实上，从N端起到C端，整个螺旋旋转了 $100^\circ \times (16-1) = 1500^\circ = 4 \times 360^\circ + 60^\circ$ ，因此N-端第一个残基与C-端第一个残基相差 60° 。但是，N-端第一个残基与C-端第一个残基是否在同一个疏水侧（或亲水侧），则与最初设计时多

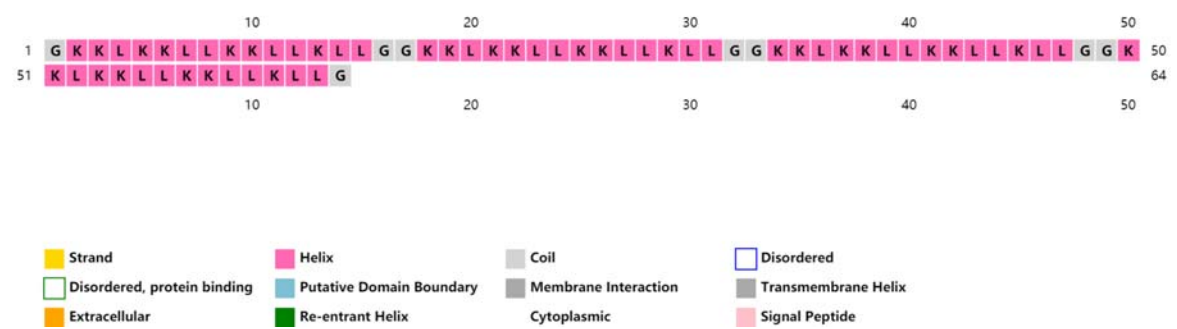
肽的序列有关。

(3) 这里我们采用PSIPRED进行预测，一方面是因为其预测准确度比较高（尽管不如基于深度学习的预测工具），一方面是因为其预测速度快，而且PSIPRED不需要注册即可使用。不过PSIPRED只能对不少于30个残基的多肽进行预测，因此笔者只能将待预测的多肽序列倍增至四倍，然后进行预测。以下是倍增后的多肽序列的预测结果。

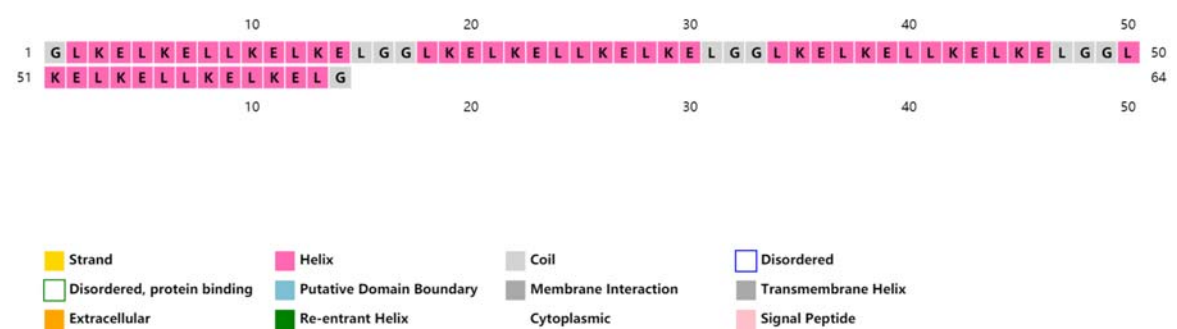
(i) GLKKLKKLLKKLKKLGGLKKLKKLLKKLKKLGGLKKLKKLLKKLKKLG（对应的原始序列为GLKKLKKLLKKLKKLG）



(ii) GKKLKKLLKKLLKLLGGKKLKKLLKKLLKLLGGKKLKKLLKKLLKLLGGKKLKKLLKKLLKLLG（对应的原始序列为GKKLKKLLKKLLKLLG）



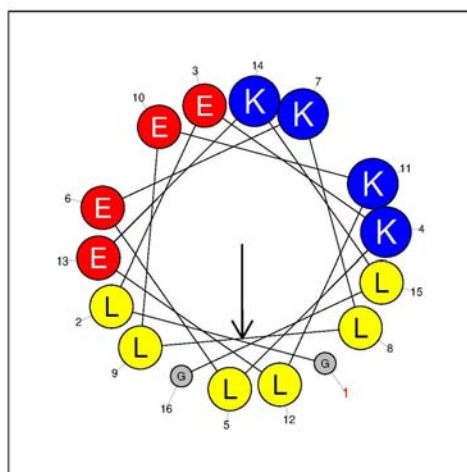
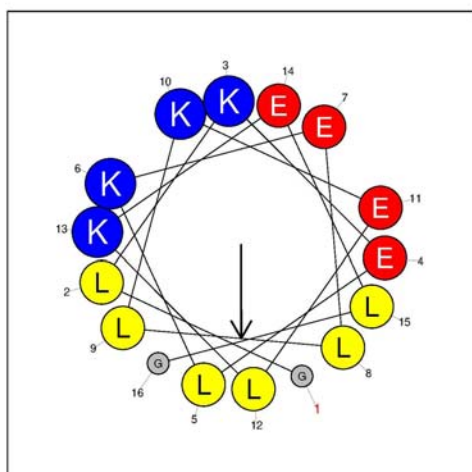
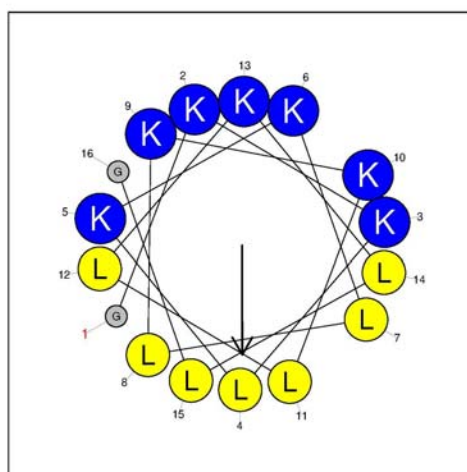
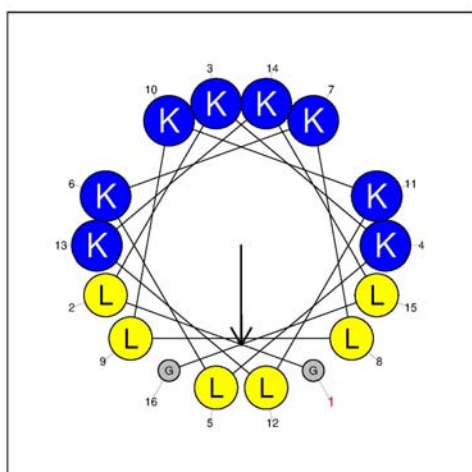
(iii) GLKELKELLKELKELGGLKELKELLKELKELGGGLKELKELLKELKELGGGLKELKELLKELKELG（对应的原始序列为GLKELKELLKELKELG）



(iv) GLEKLEKLLKLEKLLGGLEKLEKLLKLEKLLGGLEKLEKLLKLEKLLGGLEKLEKLLKLEKLLG（对应的原始序列为GLEKLEKLLKLEKLLG）



顺便给出GLKKLKKLLKKLLKG、GKKLKKLLKKLLKLLG、GLKELKELLKELKELG、GLEKLEKLLKLEKLEKG的螺旋轮图



3.第一问：在蛋白质中，原子间的共价键键长和键角变化不大，而引起蛋白质折叠过程中能量变化的主要因素是共价键的二面角，以及残基间的相互作用，因此对于蛋白质结构稳定性贡献大的主要为二面角扭转能和非键相互作用能。

第二问：二硫键不会影响蛋白质的三维结构（更准确说，二硫键对肽链的正确折叠并不是必要的，但它对稳定折叠态结构作出贡献，参见朱圣庚《生物化学》第四版P136-137），证据为将多肽链变性并还原其二硫键后，多肽链仍能折叠出正确的结构，并形成正确位置的二硫键，说明二硫键不是随机形成的，是在多肽链正确折叠后形成的。

第三问：目前由于（超级）计算机算力有限，因此分子动力学模拟尚无法做到模拟速度很慢（如秒数量级）的变化，最多只能模拟毫秒数量级的变化。

4.第一个问题的解答：首先，直接解决能量最小化问题为NP问题；其次，蛋白质包含的原子数目太大

(几千至数万个)，导致优化的参数过多，直接使用能量极小化算法，可能会因为参数过多而使时间复杂度大大增加，甚至难以计算；最后，即使计算能力允许使用能量极小化算法进行计算，也会因为蛋白质存在亚稳态构象（局域极小值，但不是全局最小值），导致得到的能量极小值点未必对应于蛋白质的能量最低构象。

第二个问题的解答：分子动力学或蒙特卡洛模拟退火等方法主要解决的便是搜索的状态空间过大的问题，利用随机微扰结合Metropolis采样（蒙特卡洛法），或通过给定坐标和初速度模拟分子运动轨迹（分子动力学），可以在尽可能不丢失最优结构的情况下，使待优化分子的搜索空间大大缩小。

5.第一问：蛋白质与配体结合过程的主要驱动力包括：范德华力、氢键、静电相互作用、偶极相互作用、 π - π 堆积相互作用、疏水相互作用（可能还有配位作用？）。

第二问：在已经形成的复合物中，蛋白质与配体间的主要作用力有范德华力、氢键，对于带电荷配体和蛋白质结合位点还会有静电相互作用，对于含有金属的蛋白质结合位点还会有配位作用。

第三问：某个蛋白质P在实验条件下可以与配体L结合，说明该反应的 $\Delta G < 0$ ，但实验测出来的 $\Delta H > 0$ ，说明焓效应不是驱动结合的主要因素。根据 $\Delta G = \Delta H - T\Delta S$ 可知，当 $\Delta H > 0$ 时，由于正常温度为正值

($T > 0$)，为使 $\Delta G < 0$ ，只能使 $\Delta S > 0$ ，从而熵效应（具体而言，主要为疏水效应）是驱动结合的主要因素。