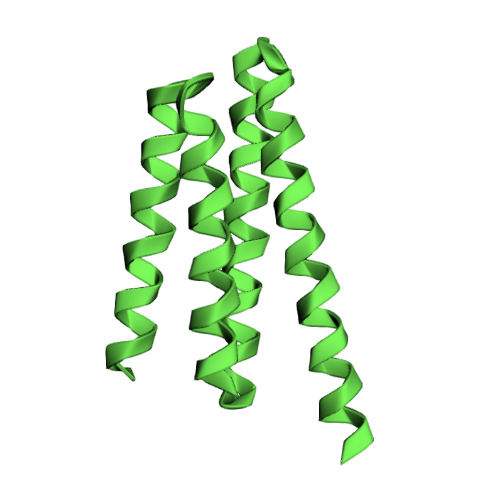
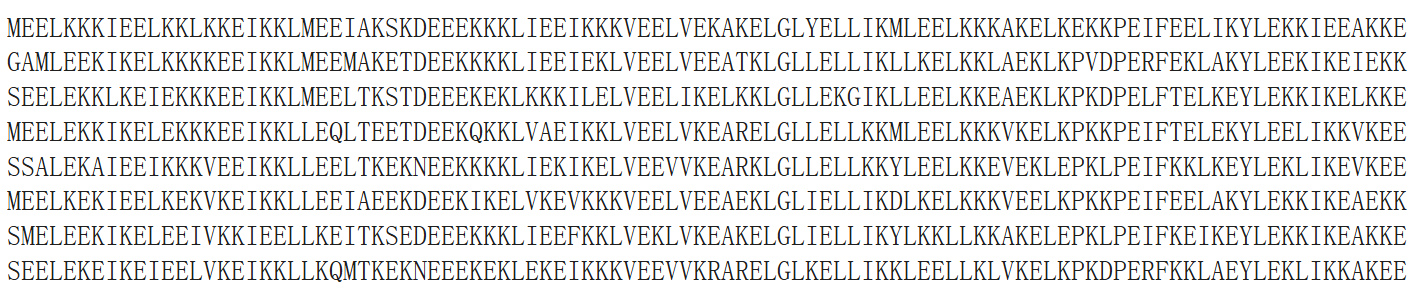
我们通过运行RFdiffusion，在contigs（重叠群）=100（即尝试生成100个不同的连续序列片段，从而找到最佳的蛋白质设计）和iterations（迭代次数）=50（即寻找最佳蛋白质序列的算法过程重复50次）的条件下生成了一个由四条α螺旋所组成的基本骨架，如下所示：



**图1 蛋白质骨架图**

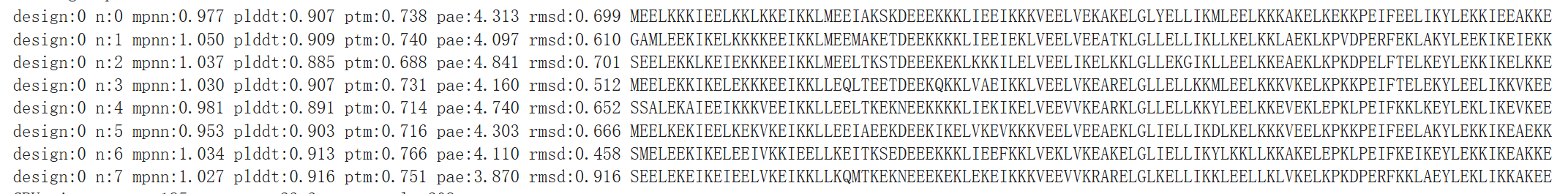
**1.2.2 序列设计**

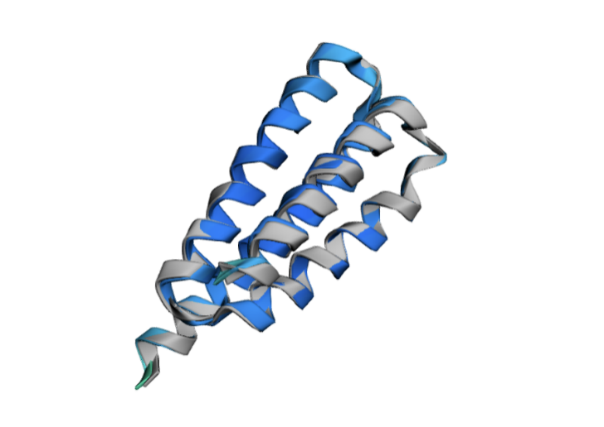
根据基本骨架，我们在运行一次循环和mpnn\_sampling\_temp（温度参数） = 0.1（即选择概率更高的氨基酸）的条件下运行ProteinMPNN生成八条序列：

**图2 氨基酸序列**

**1.2.3 序列筛选**

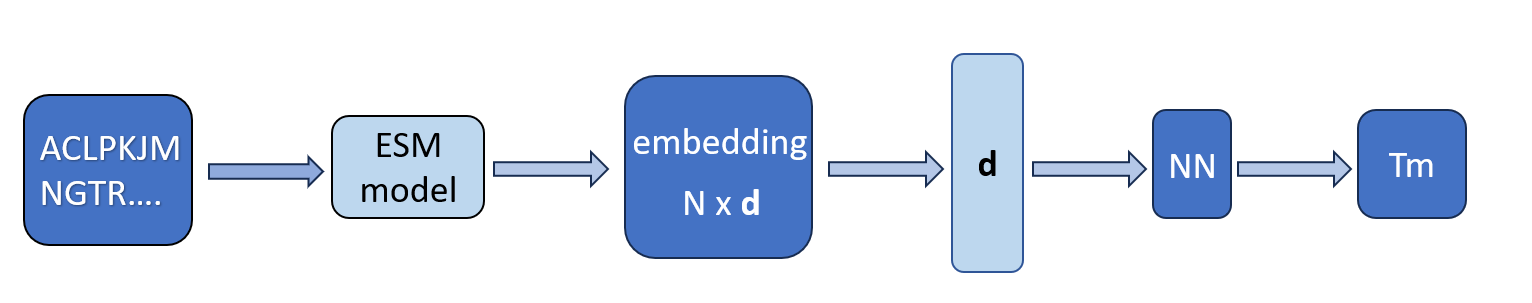
根据生成的序列，用AlphaFold2进行分析，取RMSD（预测结构与参考结构之间的均方根偏差）最小的序列进行三维结构展示，得到结果如下：

**图3 AlphaFold2分析结果**

**图4 预测结构与参考结构对比图**

从结构对比图中可见，所生成氨基酸序列与原蛋白骨架契合度较高，预测效果良好。

**1.2.4 利用ESM从氨基酸序列预测蛋白变性温度**

**图 5 神经网络结构示意图**

先向ESM传入蛋白的氨基酸序列，再将文本转化为数值向量（即embedding，用于捕捉文本的语义关系），并经MLP（Multilayer Perceptron，）处理，经过十轮epoch（十次完整训练数据集过程），输出蛋白的变性温度预测结果。

每轮epoch的预测效果如下图所示：

**图6 模型在训练集和测试集上的损失函数值**

由图6可知，本模型在测试集上的损失函数值始终维持在一个稳定水平，第二次循环时，已经能对训练集展示出较好的预测效果，接近测试集的水平，模型训练效果良好。