

纳米孔测序技术简介及孔 蛋白MspA的结构分析

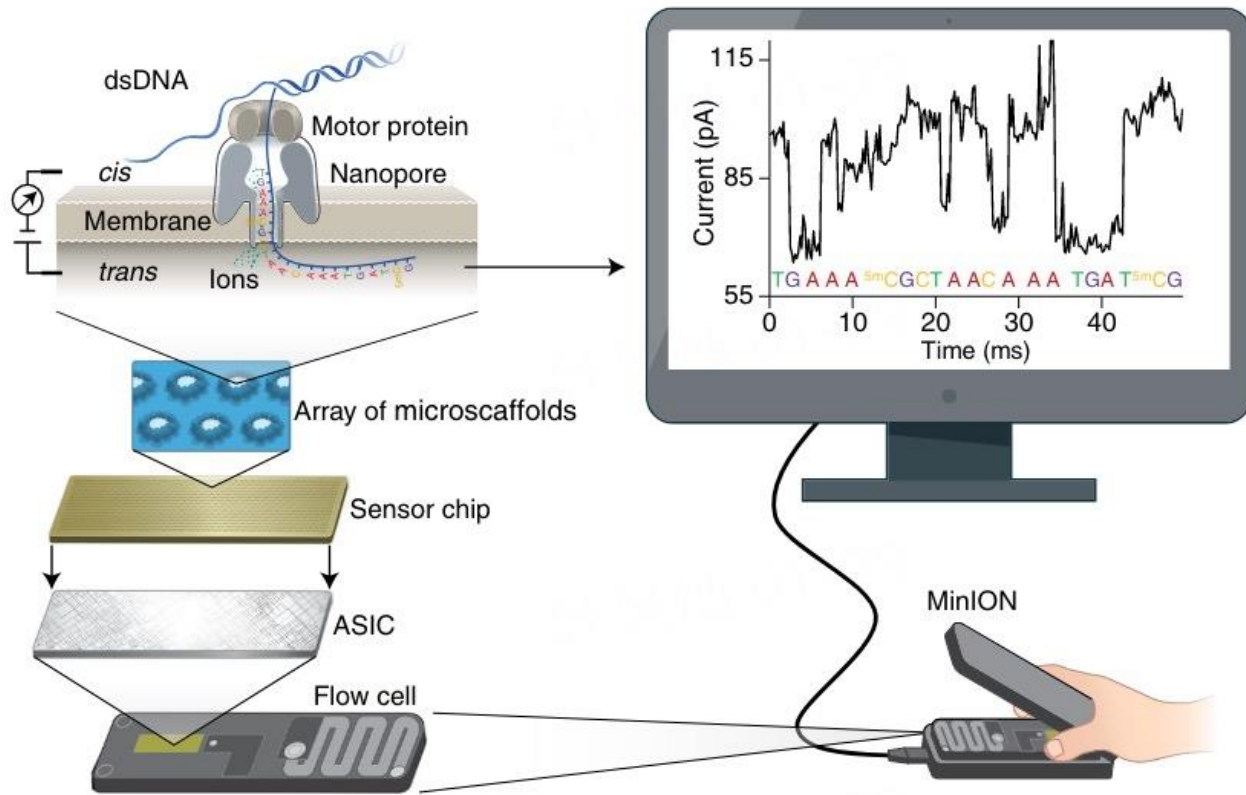
报告人: G2A 叶风灿

小组成员: G2B 刘雅文

G2C 陈详茂

G2D 阿克周力

纳米孔测序基本原理



纳米孔蛋白: α -Hemolysin, MspA, CsgG 等

马达蛋白: phi29 DNA聚合酶 等

电解质溶液: 提供阴阳离子

绝缘膜: 分隔电解质溶液

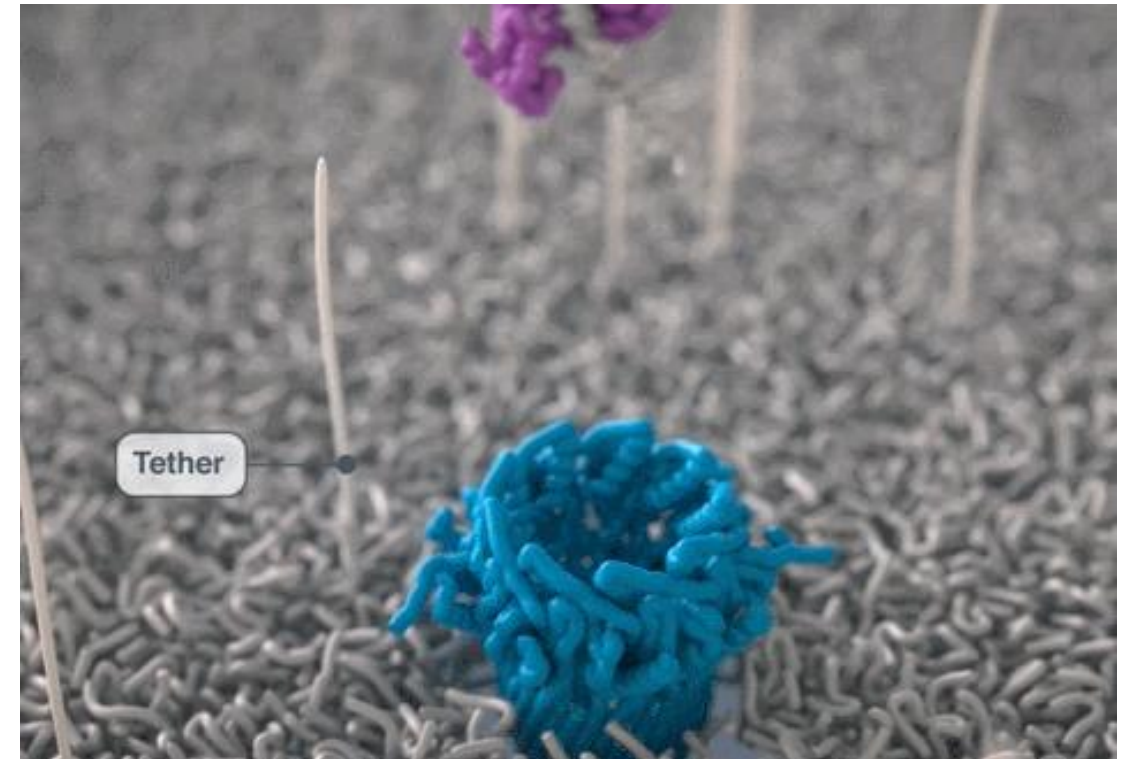
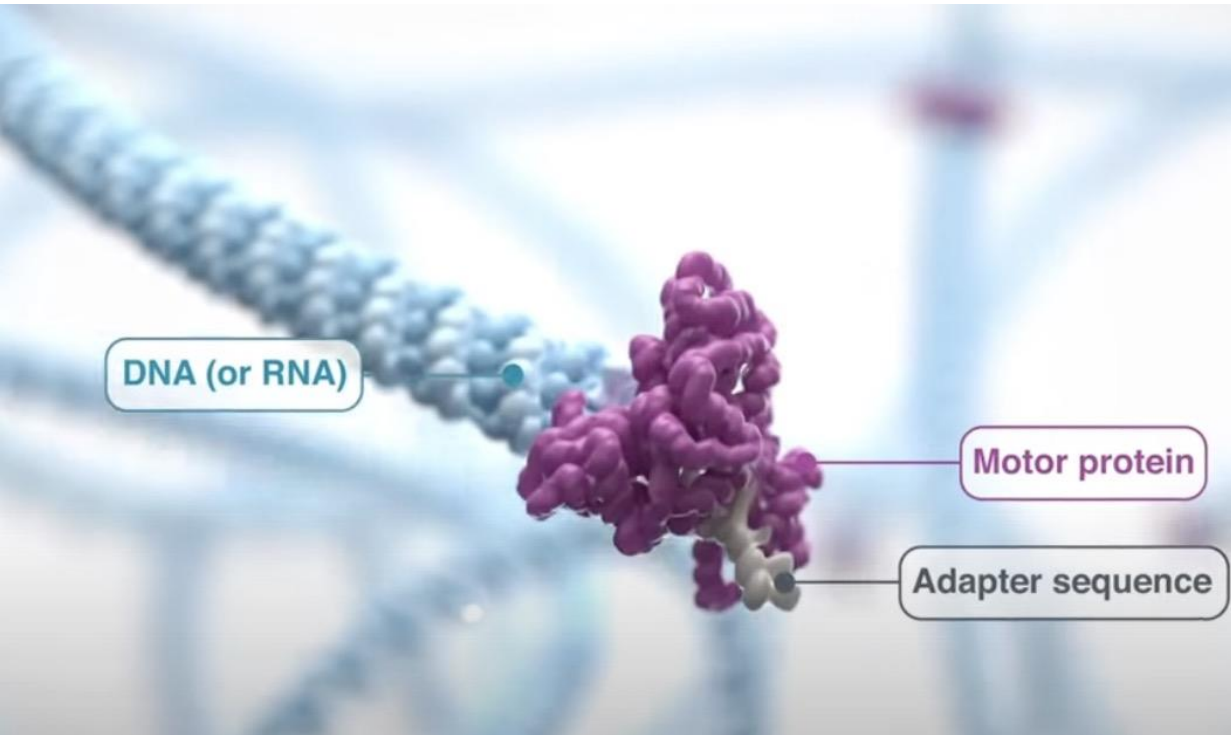
电极: 提供恒定电压

Application-**s**pecific **I**ntegrated **C**ircuit: 控制电极、测量电流

Wang, Y., Zhao, Y., Bollas, A., Wang, Y. & Au, K. F. Nanopore sequencing technology, bioinformatics and applications. *Nat Biotechnol* **39**, 1348–1365 (2021).

实时高通量单分子测序, 单个纳米孔测序速度可达450 bases/s

纳米孔测序基本原理



How nanopore sequencing works

<https://www.youtube.com/watch?v=RcP85JHLmnl>

纳米孔测序：不同测序方式

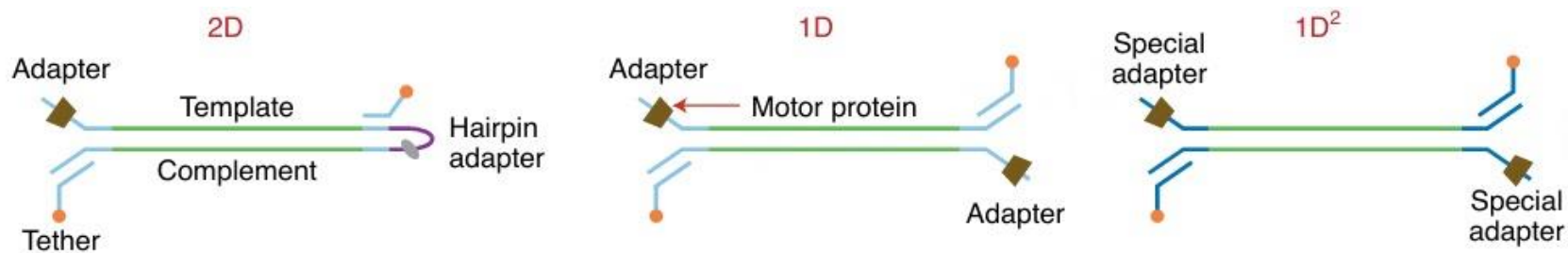
2D: 用发卡接头序列 (hairpin adapter) 连接待测序列和互补序列

1D: 直接对待测序列进行测序

1D²: 使用一种特殊的接头, 在待测链被纳米孔捕获 (capture) 后, 其互补链也有大概率 (>60%) 随即被同一个纳米孔捕获

RNA测序: 可直接对RNA进行测序, 无需反转录

Wang, Y., Zhao, Y., Bollas, A., Wang, Y. & Au, K. F. Nanopore sequencing technology, bioinformatics and applications. *Nat Biotechnol* **39**, 1348–1365 (2021).



Direct RNA sequencing

Full-length poly(A)⁺ RNA



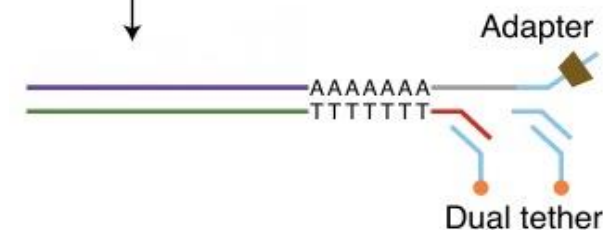
TTTTT
Primer annealing



Reverse transcription
(optional)

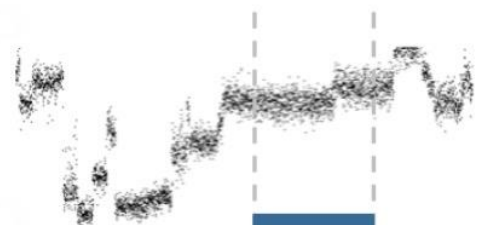


Adapter ligation



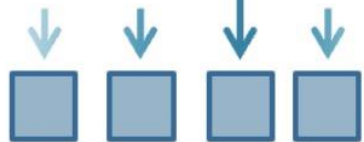
纳米孔测序： Base Calling

Base calling (RNN, raw)

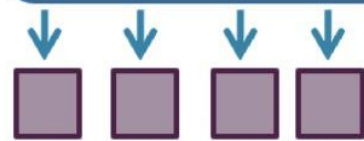
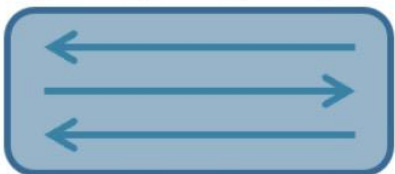


Parameters learned from training data

Extraction of blocks of features



Bidirectional information flow



Multi-base prediction



Decode to sequence

电流信号与孔道最窄处附近的5-6个碱基均有关

需要通过算法将电流信号转化为碱基序列

不仅可以测量碱基种类，还能够测量碱基上的修饰（如甲基化）

Oxford Nanopore Technologies Base Calling算法： Guppy

纳米孔测序的优缺点

优点:

便携

可进行单分子测序

测序速度快 (450 bases / s) , 可实时测序

高通量: 一个MinION设备可同时进行512条DNA的测序

可测序的DNA / RNA长度较长

缺点:

测序准确度低: R9.4 nanopore, 1D测序平均准确率86%, 2D测序平均准确率94%
1D²测序平均准确率95%

对比: 二代测序Illumina/Solexa的平均准确率为 99.2% 和 99.74%

纳米孔测序的应用

用于填充参考基因组（reference genome）中的空隙（gaps）

用纳米孔测序生成参考基因组：主要为非模式生物

鉴定基因组结构变异（structural variations）

检测RNA分子上的修饰

用于多肽序列的测定

.....

纳米孔测序的主要改进方法

对**纳米孔蛋白**以及**马达蛋白**的改进

对纳米孔蛋白的要求：能够让DNA / RNA链通过，通过的电流受邻近碱基的影响越小越好

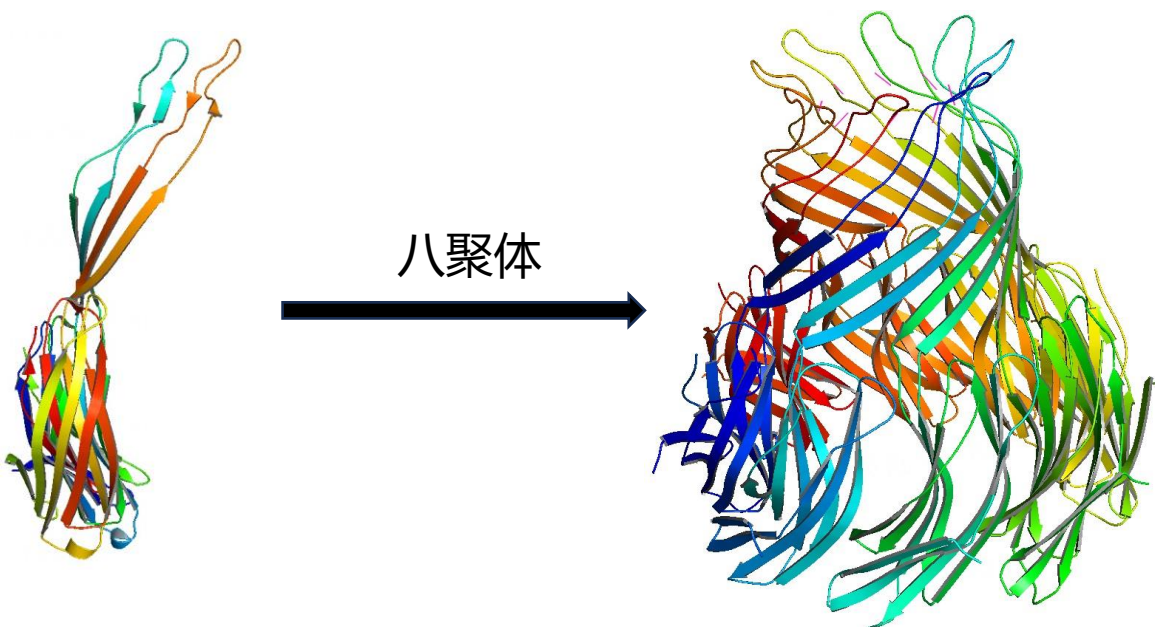
对马达蛋白的要求：控制DNA / RNA链通过孔蛋白保持在一个合适的速度 —— 提高信噪比

测序策略的改进：1D, 1D², 2D

Base calling 算法的改进

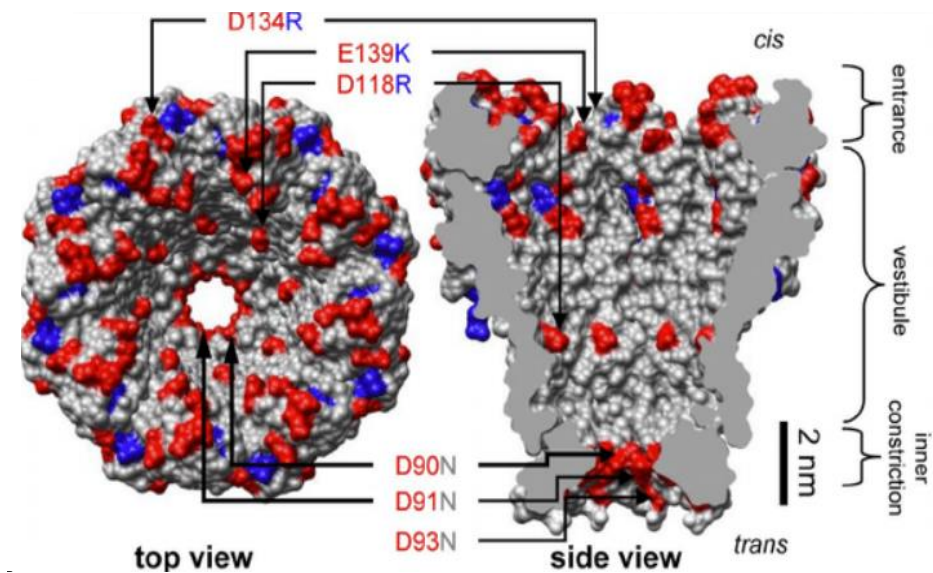
下面介绍实验采用的纳米孔蛋白突变体是如何从野生型蛋白改造而来的

MspA蛋白的结构与突变体



PDB entry: 1UUN

形成八聚体孔道



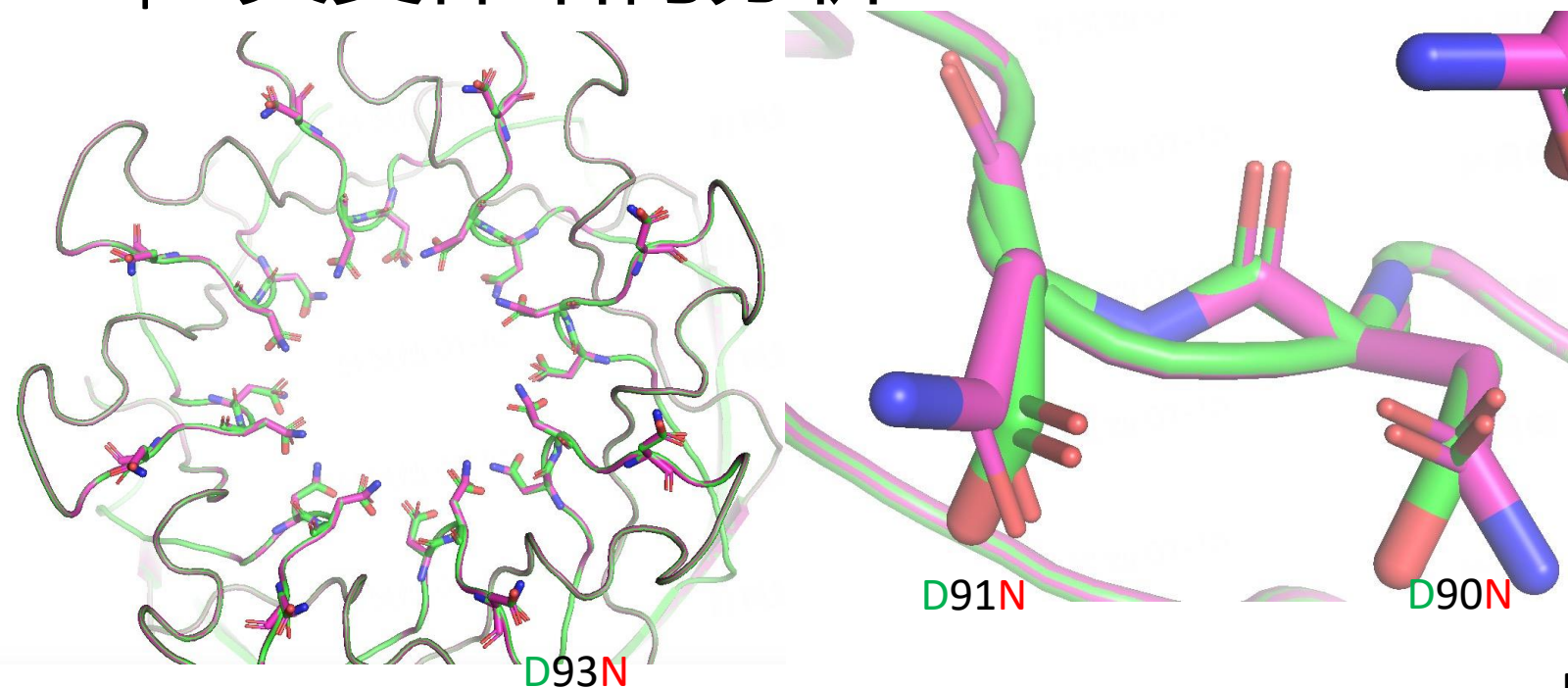
Butler, T. Z., Pavlenok, M., Derrington, I. M., Niederweis, M. & Gundlach, J. H. Single-molecule DNA detection with an engineered MspA protein nanopore. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 20647–20652 (2008).

实际测序用孔道采用突变体

主要突变孔道内带负电氨基酸

这样突变的原因？

MspA突变体结构分析



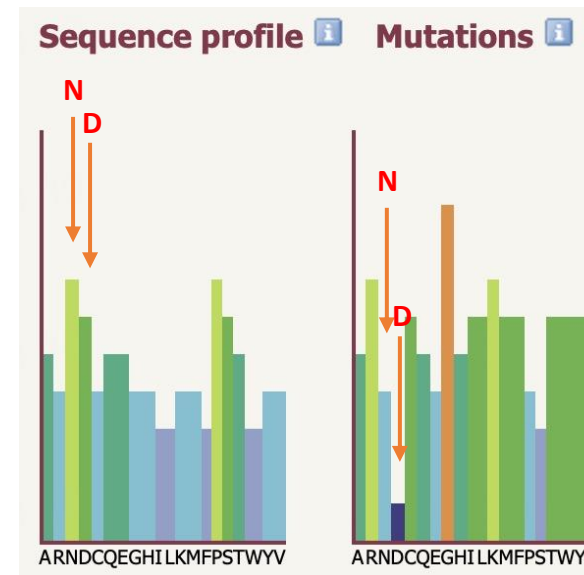
D90N, D91N, D93N: 位于孔道入口处

D带负电荷, 不利于DNA / RNA穿过

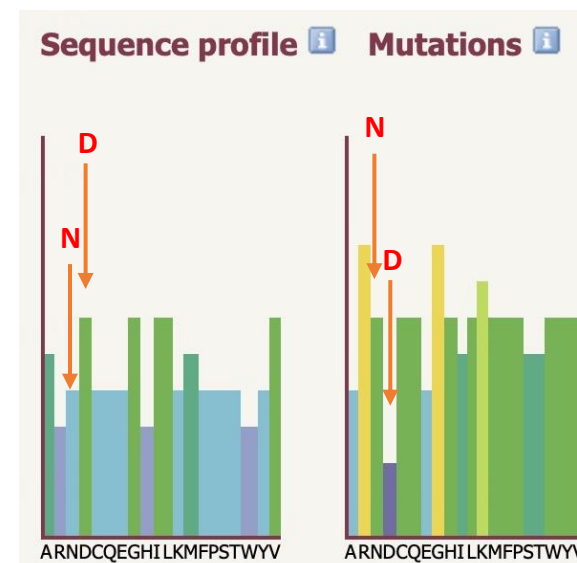
D的等电点为2.77, N的等电点为5.41

根据Phyre2 对D90, D91位点突变后是否影响功能的评估, 不影响功能的突变中 N 具有较高的等电点, 带有较少负电荷

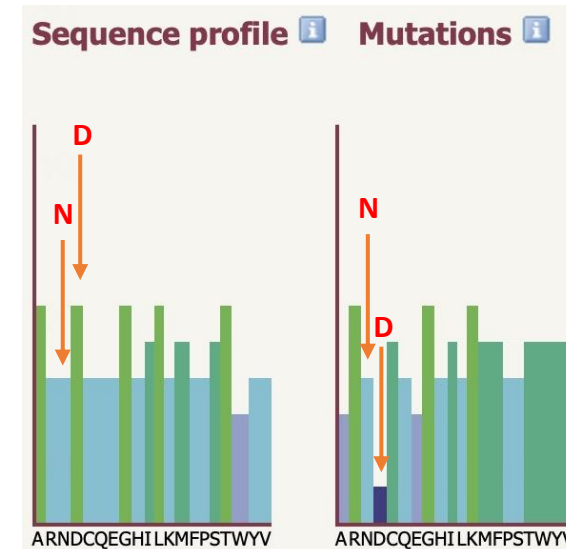
D90



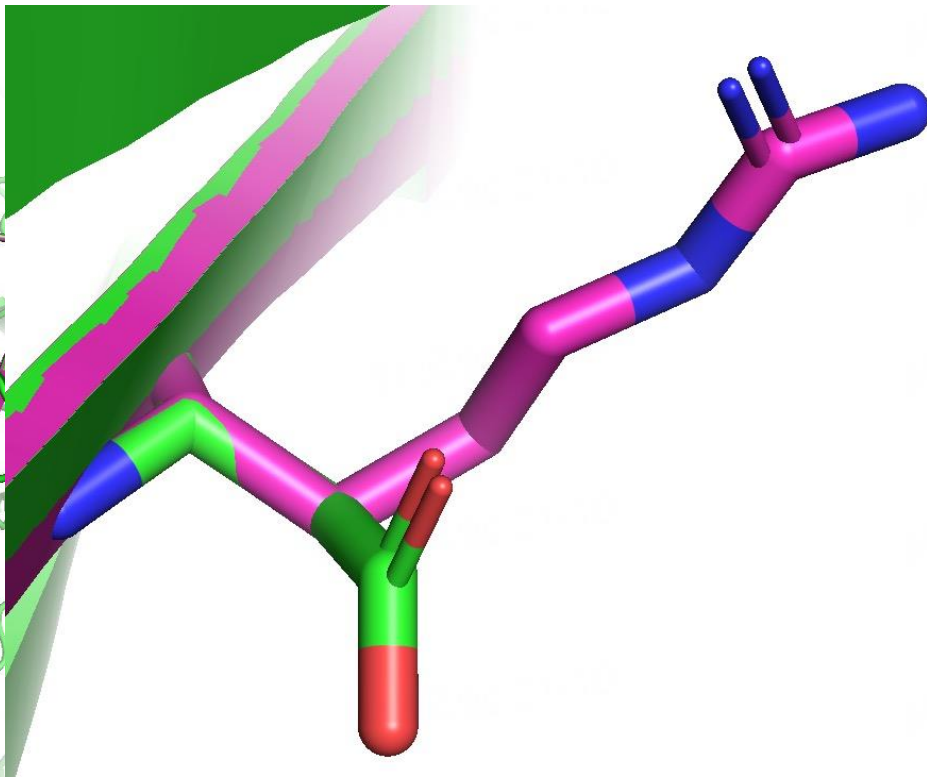
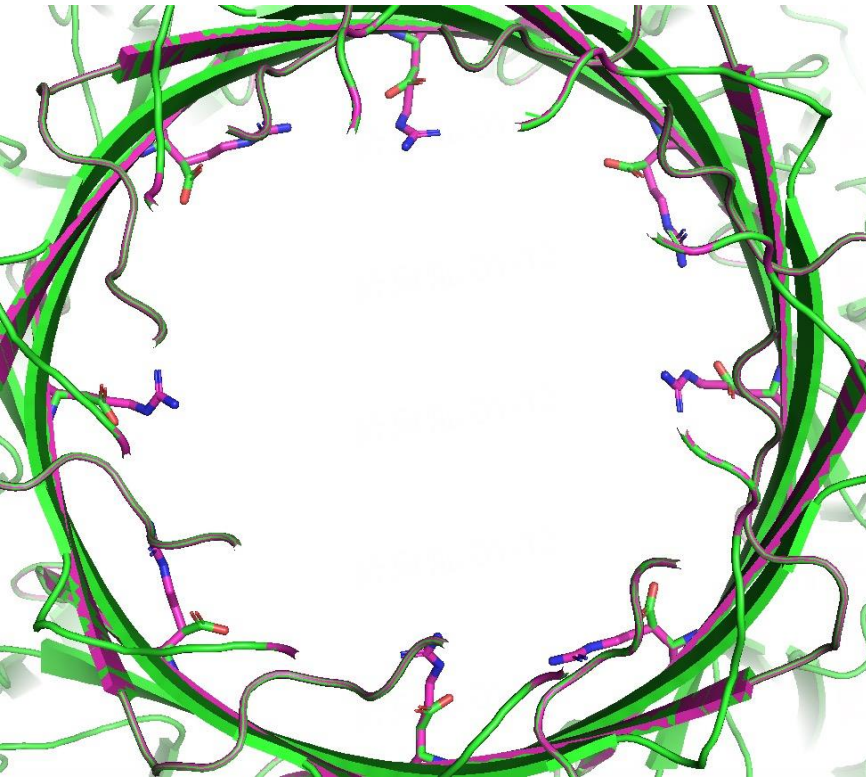
D91



D93



MspA突变体结构分析

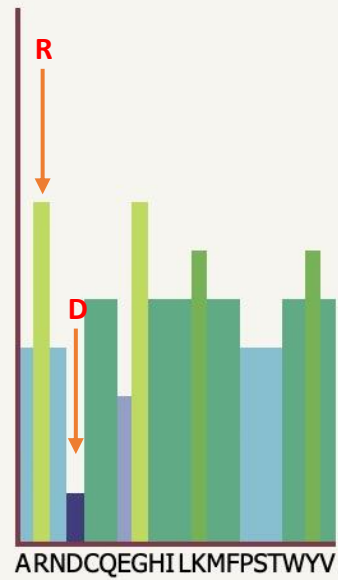
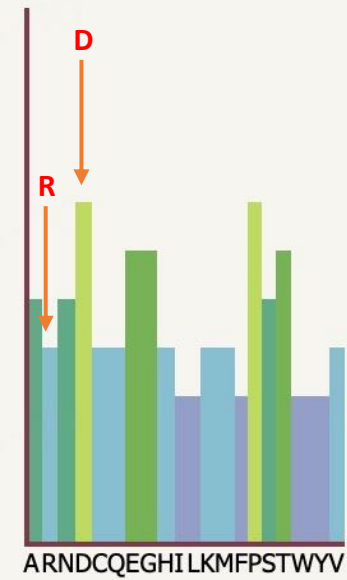


D118R

D118

Sequence profile [i](#)

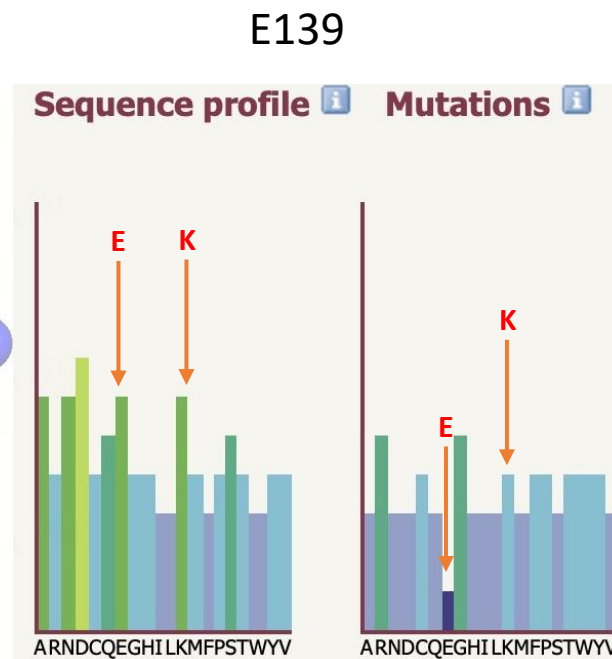
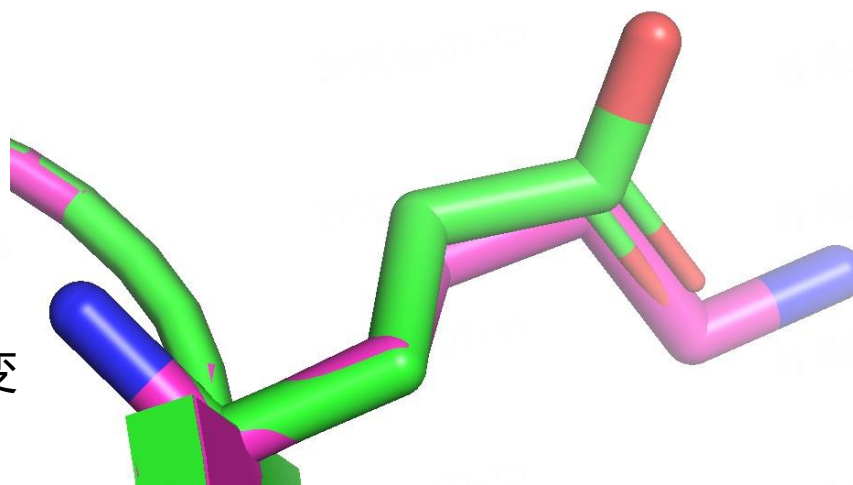
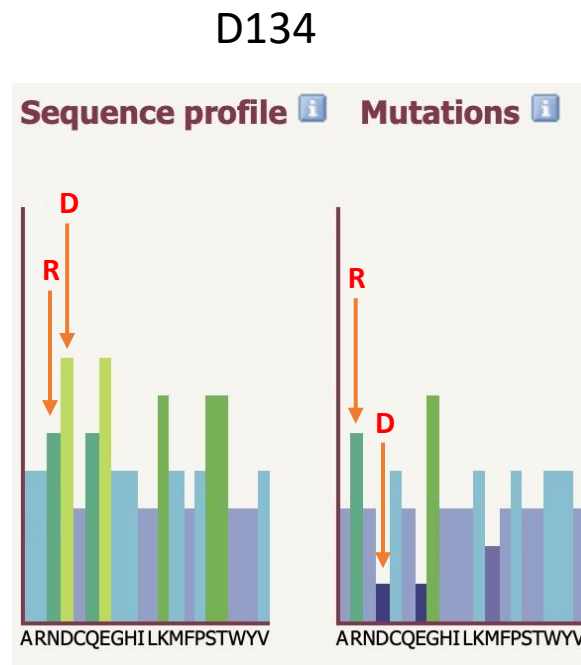
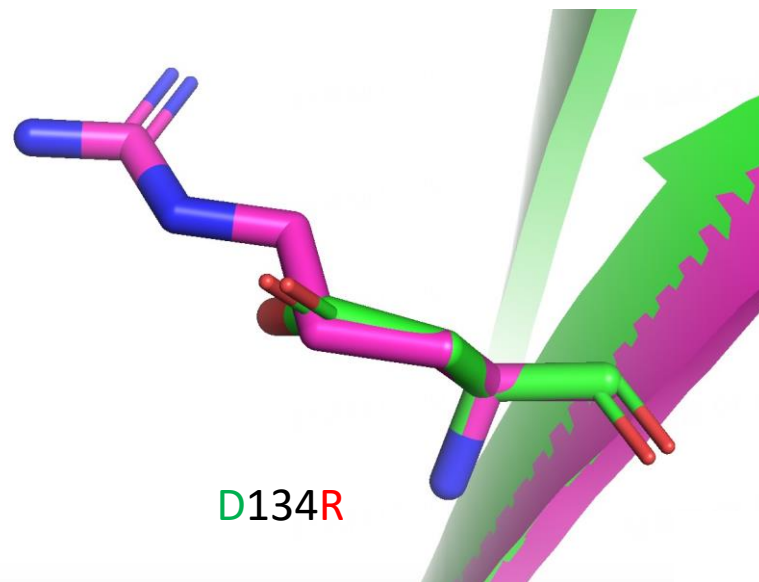
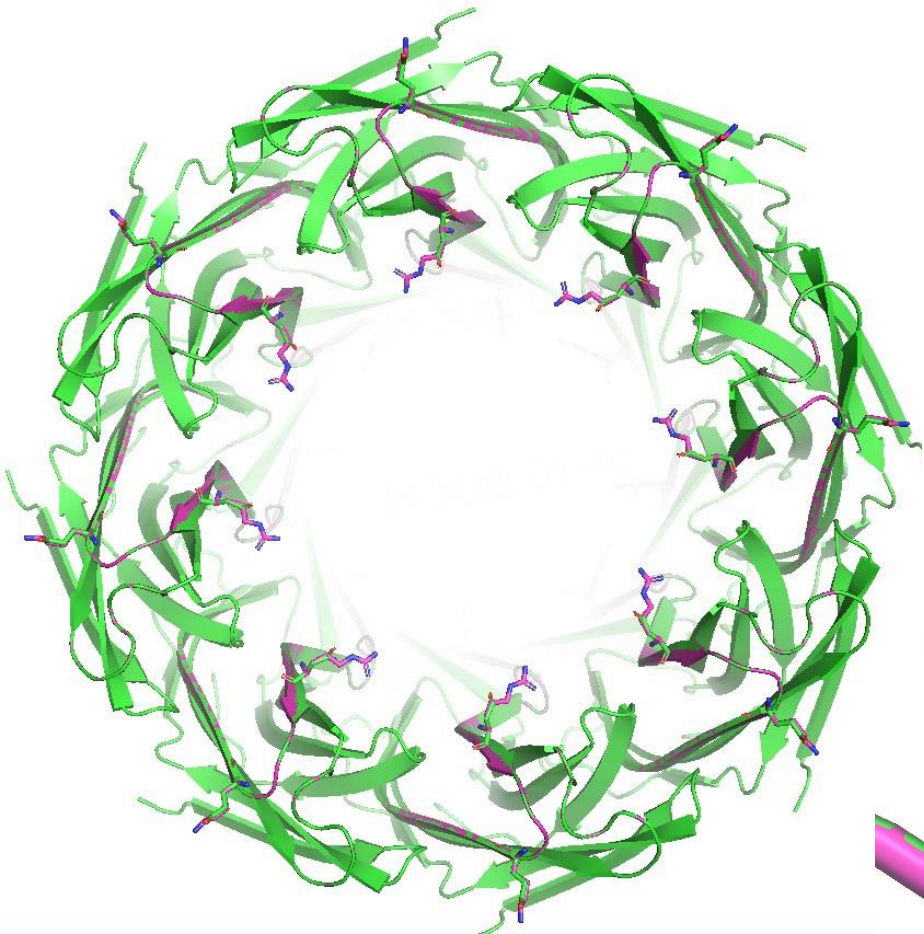
Mutations [i](#)



在通道的较宽处，将带负电的D突变为带正电的R

对结构影响较小，但能促进DNA / RNA通过

MspA突变体结构分析



在孔道开口处继续将带负电氨基酸突变为带正电氨基酸

突变体总结: 将对结构影响不大的负电氨基酸突变成带负电更少/带正电的氨基酸, 促进DNA / RNA链的通过

课程感悟

互联网上资源十分丰富，各种数据库数据量极大，可以为实验工作者提供很多便利，对于主要进行计算或是模拟研究的人员，提供了非常宝贵的数据和素材，需要好好挖掘利用

个人认为许多生信工具的作用在于寻找亲缘关系或是演化关系，要用计算的方式解决好生物问题，利用亲缘或演化信息十分重要。例如结构预测中的Phyre2，以及最近的深度学习蛋白质结构预测模型AlphaFold，都利用了亲缘或演化信息。

即使是在分子动力学这类和物理较为相关的领域，事实上也可以利用演化信息。

Prody Tutorial: <http://prody.csb.pitt.edu/tutorials/> 其中有将序列关系和大分子的动力学性质联系起来的教程，以及blast等工具在研究大分子结构中的应用

谢谢！