

Что может Альфафолд2

С помощью [ColabFold](#) можно получить предположительную структуру любого белка в течение часа. В этом практикуме я попробую сама запустить ColabFold для выполнения задания B - Helices.

Мне было выдано 3 последовательности - seqA, seqB, seqC:

```
PASEQETLVRPKPLLLKLLKSVGAQKDTYTMKEVLFYLGQYIMTKRLYDEKQQHIVYCSNDLLGDLFGVPSFSVKEHRKIYTMiYRNLVWN
```

```
PASEQETLVRPKPLLLKLLKSVGAQKDTYTMKEVLFPPPYIMTKRLYDEKQQHIVYCSNDLLGDLFGVPSFSVKEHRKIYTMiYRNLVWN
```

```
PASEQETLVRPKPLLLKLLKSVGAQKDTYTPPPPPPPPPPPPPPPLYDEKQQHIVYCSNDLLGDLFGVPSFSVKEHRKIYTMiYRNLVWN
```

SeqA отличается от двух других последовательностей заменами аминокислот на 4 или 16 подряд идущих пролинов.

Первым делом я сделала предсказание структуры для SeqA. В результате работы было получено пять структурных предсказаний (Рис. 1).



Рис. 1 Предсказание для структуры SeqA. (Структура с rank1 окрашена салатovým, с rank2 - голубым, rank3 - малиновым, rank4 - желтым, rank5 - светло-розовым)

Полученные структуры хорошо сходятся, они почти идентичны. Наблюдаются небольшие отличия лишь для концевых участков, которые, вероятно, являются подвижными.

Также, чтобы оценить качество выдачи программы, мне удалось найти расшифрованные структуры для seqA с помощью protein blast. Оказалось, что этой последовательности соответствует белок E3 убиквитин-лигаза. В результате наложения его структуры 4HBM на ту, что была собрана альфафолдом и имела ранг 1, выяснилось, что они почти не отличаются (Рис. 16). Структура была воссоздана очень достоверно.

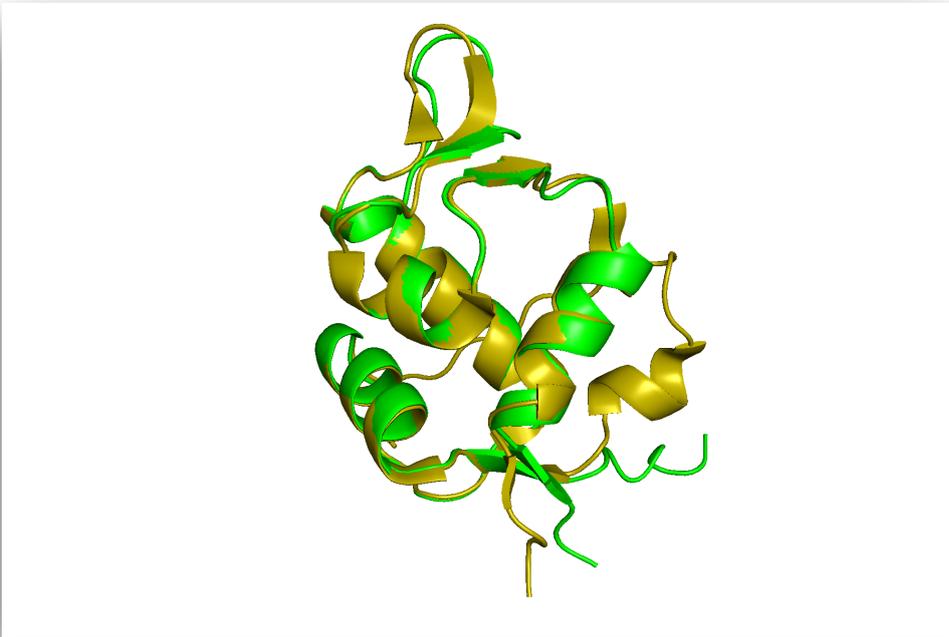


Рис. 16 Наложение структуры Seq1 rank1 на 4HBM

Я ожидаю, что при замене а.о. на пролины в соответствующей области будут торсионные углы, характерные для полипролинового участка. Проверю это предположение, изучив выдачу альфафолда для SeqB (Рис. 2) и SeqC (sРис. 4).

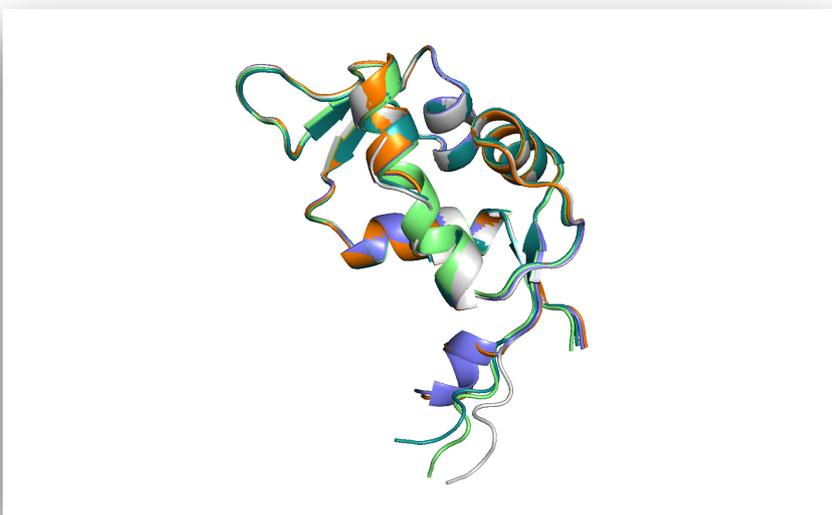


Рис. 2 Предсказание для структуры SeqB. (Структура с rank1 окрашена серым, с rank2 - фиолетовым, rank3 - оранжевым, rank4 -салатовым, rank5 - темно-зеленым)

На рисунке 2 видим «разночтения» для концевых элементов. Кроме того, структуры немного отличаются между собой как раз на том участке, где а.о. были заменены на пролины.

Рассмотрим выдачу подробнее на примере структуры с rank1 (Рис. 3).

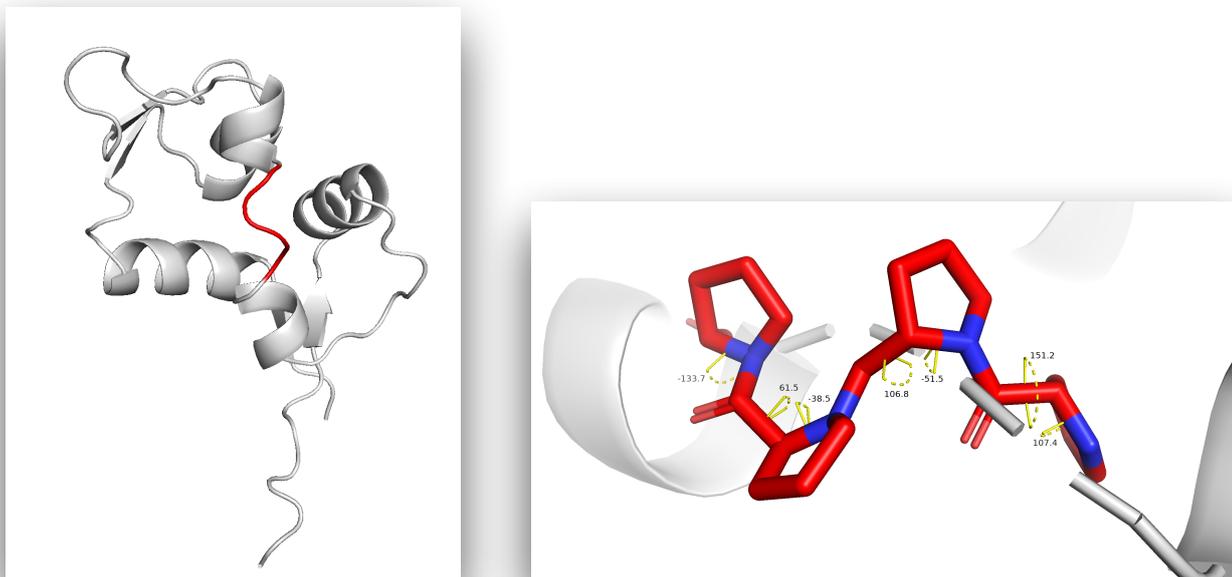


Рис. 3 Слева: Предсказание для структуры SeqV с rank1, справа - торсионные углы для участка из пролинов SeqV с rank1.

Торсионные углы для участка из пролинов лучшей выдачи плохо вписываются в значения, характерные для полипролинов (Рис.5, справа). Кроме того, если рассматривать выдачу с rank 4 (Рис. 3б), то для нее наблюдаются торсионные углы, принадлежащие альфа - спиральям, что также неверно.

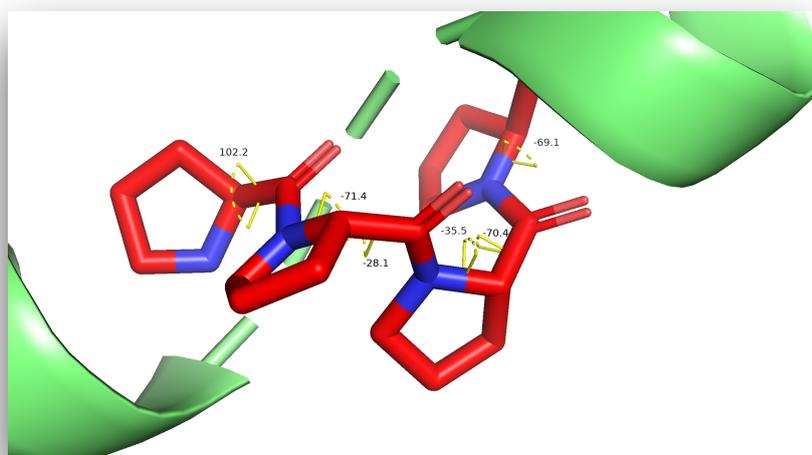


Рис. 3б Торсионные углы для участка из пролинов SeqV с rank4.

По данному примеру можно сделать вывод о том, что Альфафолд плохо справляется с обработкой небольшого количества замен в структуре.

Теперь изучим выдачу для SeqC (Рис. 4):

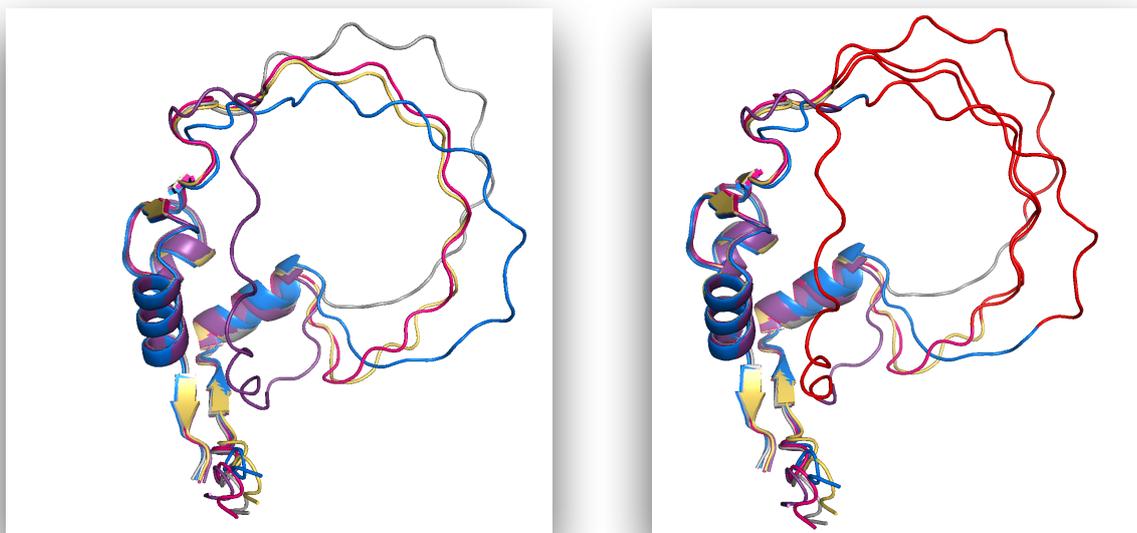


Рис. 4 Слева: Предсказание для структуры SeqC. (Структура с rank1 окрашена малиновым, с rank2 - желтым, rank3 - фиолетовым, rank4 -серым, rank5 - синим). Справа: структура SeqC с окрашенным красным участком, состоящим из 16 пролинов.

Если проверить торсионный углы для лучшей структуры, то они соответствуют области полипролина на карте Рамачандрана (Рис. 5). Это согласуется с моей гипотезой и говорит о хорошем качестве структуры, полученной альфафолдом при большой замене.

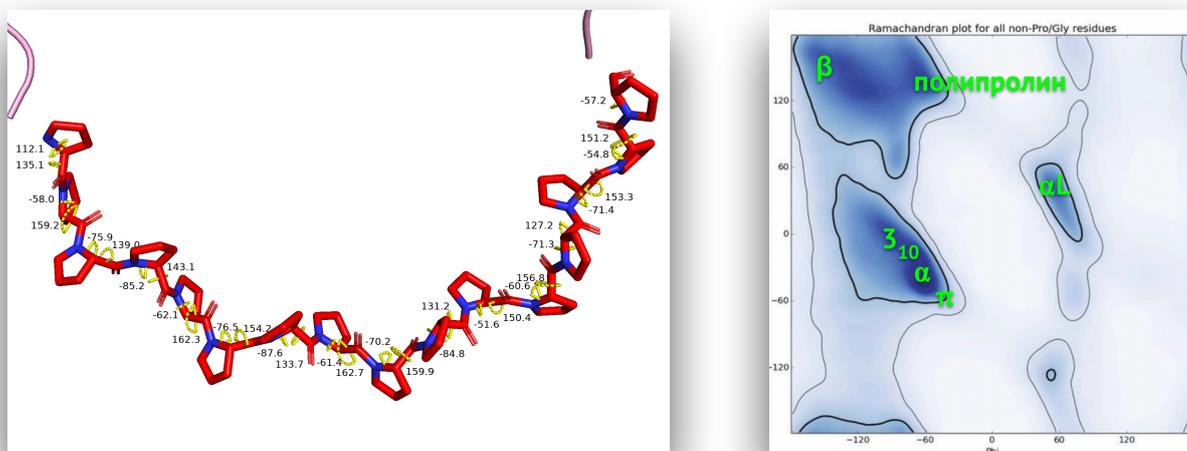


Рис. 5 Слева: Измерение торсионный углов для полипролинового участка SeqC структуры с rank1; Справа: карта Рамачандрана

[Выдача для SeqA](#)
[Выдача для SeqB](#)
[Выдача для SeqC](#)