

Что может AlphaFold2

В этом практикуме изучались возможности AlphaFold2. Это программа, предсказывающая структуру белка по его аминокислотной последовательности.

Я выбрал кейс B. Helices (B6)

Мне было дано 3 последовательности:

SeqA:

```
MSRSQRAGLQFPVGRIHRHLKSRTTSHGRVGATAAVYSAAILEYLTAEVLELAGNASKDLKVKRITPRHL  
QLAIRGDEELDSLK
```

SeqB:

```
MSRSQRAGLQFPVGRIHRHLKSRTTSHGRVGATAAVYSAAILEYLTAPPPLAGNASKDLKVKRITPRHL  
QLAIRGDEELDSLK
```

SeqC:

```
MSRSQRAGLQFPVGRIHRHLKSRTTSHGRVPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPKVKRITPRHL  
QLAIRGDEELDSLK
```

Последовательность seqB представляет из себя SeqA, в которой позиции 48-51 заменены на пролины, а seqC – в которой на пролины заменены позиции 31-60.

Сначала при помощи ColabFold были получены предсказания для seqA. В выдаче содержалось 5 предсказаний, ранжированных по качеству. Они были визуализированы при помощи PyMol (и выровнены с его помощью):

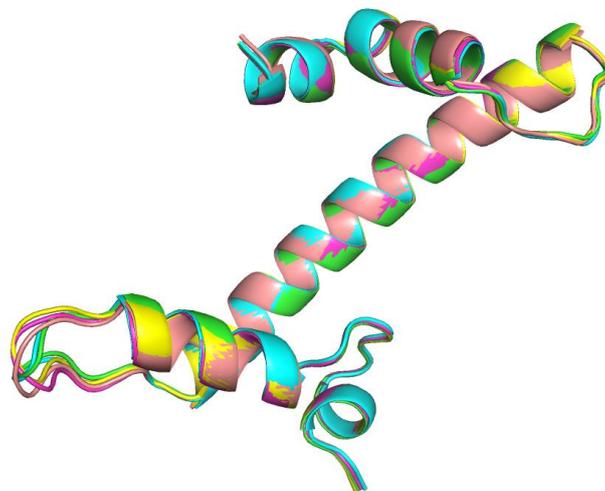
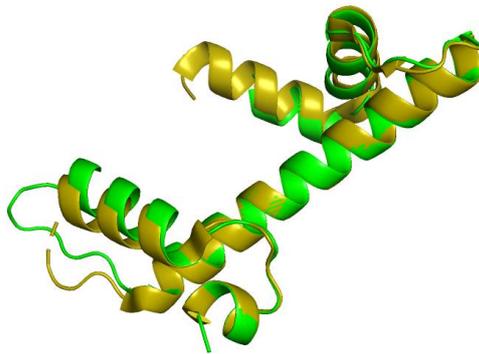


Рисунок 1. Предсказанные модели

Видно, что предсказания практически совпадают, есть лишь небольшие отличия в положении остатков одной из петель, для всех остальных петель и альфаспиралей модели очень похожи.

Также для того, чтобы оценить, насколько предсказанная структура близка к реальной, был проведен поиск при помощи protein BLAST с указанием PDB в качестве базы данных. Это оказалась последовательность гистона H2A.Z человека (PDBid 4CAY). Сравним предсказанную структуру с реальной, полученной при помощи PCA.



Предсказанная структура очень похожа на реальную, отличается лишь немного одной петлей, что вполне может объясняться её подвижностью.

Рассмотрим, где расположены участки, которыми SeqB и SeqC отличаются от SeqA.

SeqB:

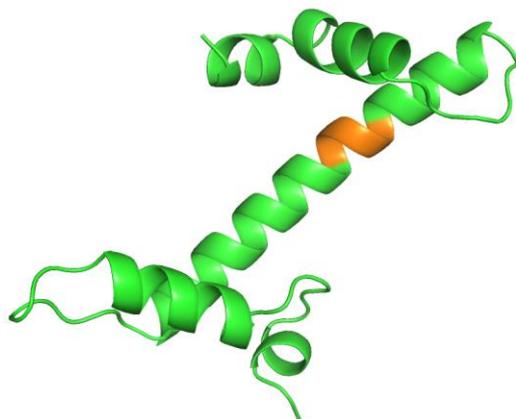


Рисунок 2. Лучшая модель для seqA. Оранжевым выделен участок, которым отличается SeqB

Этот участок расположен в альфа-спирали. Так как пролин обладает неудобными для неё структурами, можно ожидать, что в этом месте произойдет «надлом» и останутся две небольшие альфа-спирали, связанные полипролиновым участком.

SeqC:

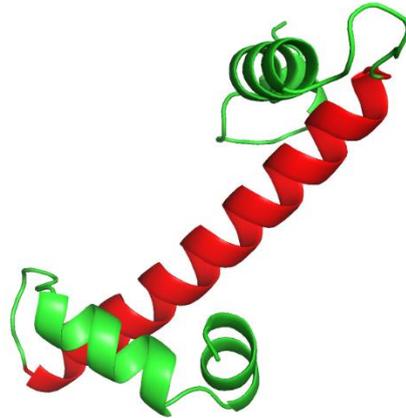


Рисунок 3. Лучшая модель для seqA. Красным выделен участок, которым отличается SeqC

Тут на пролины заменяется вся длинная альфа-спираль, можно ожидать, что произойдет её замена на полипролиновую спираль.

Результаты:

SeqB:

Все полученные модели соответствовали ожиданиям, в них большая альфа-спираль разделилась на две.

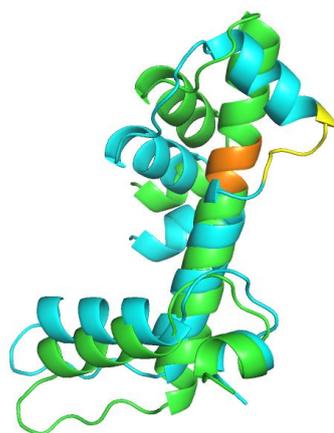


Рисунок 4. Лучшие предсказания для seqA (зеленым) и seqB (цианом). Отличающийся участок выделен оранжевым и желтым

Более внимательно изучим лучшую модель

Рассмотрим её пролиновый участок, построим торсионные углы. В предсказанной структуре 3 из 4 пролинов соответствуют разрешенным состояниям, 3 из 4 пролинов образуют участок PolyPro спирали, не характерные углы для последнего пролина могут объясняться тем, что он находится на границе между альфа-спиралью и полипролиновым участком. На мой взгляд, это хорошие результаты.

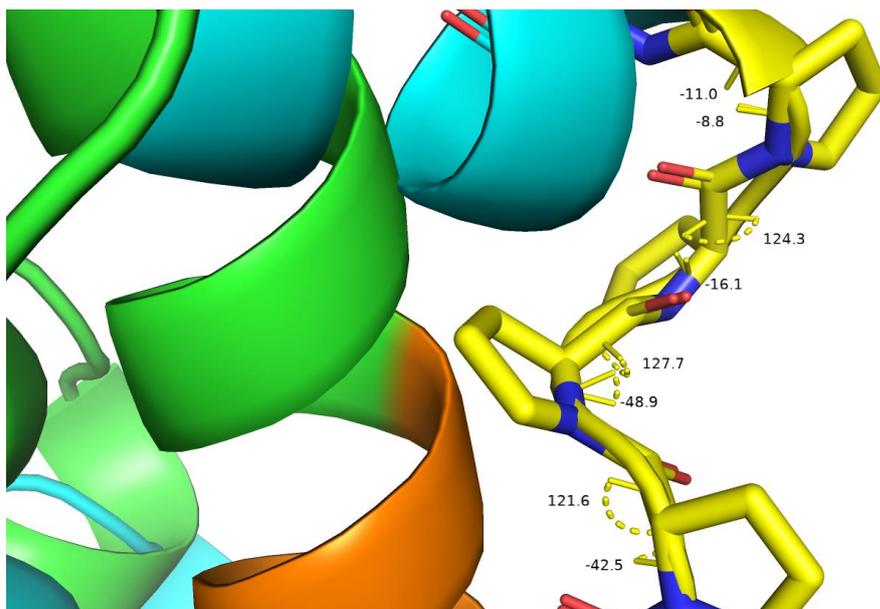


Рисунок 5. Полипролиновый участок лучшей из предсказанных для seqB структур.

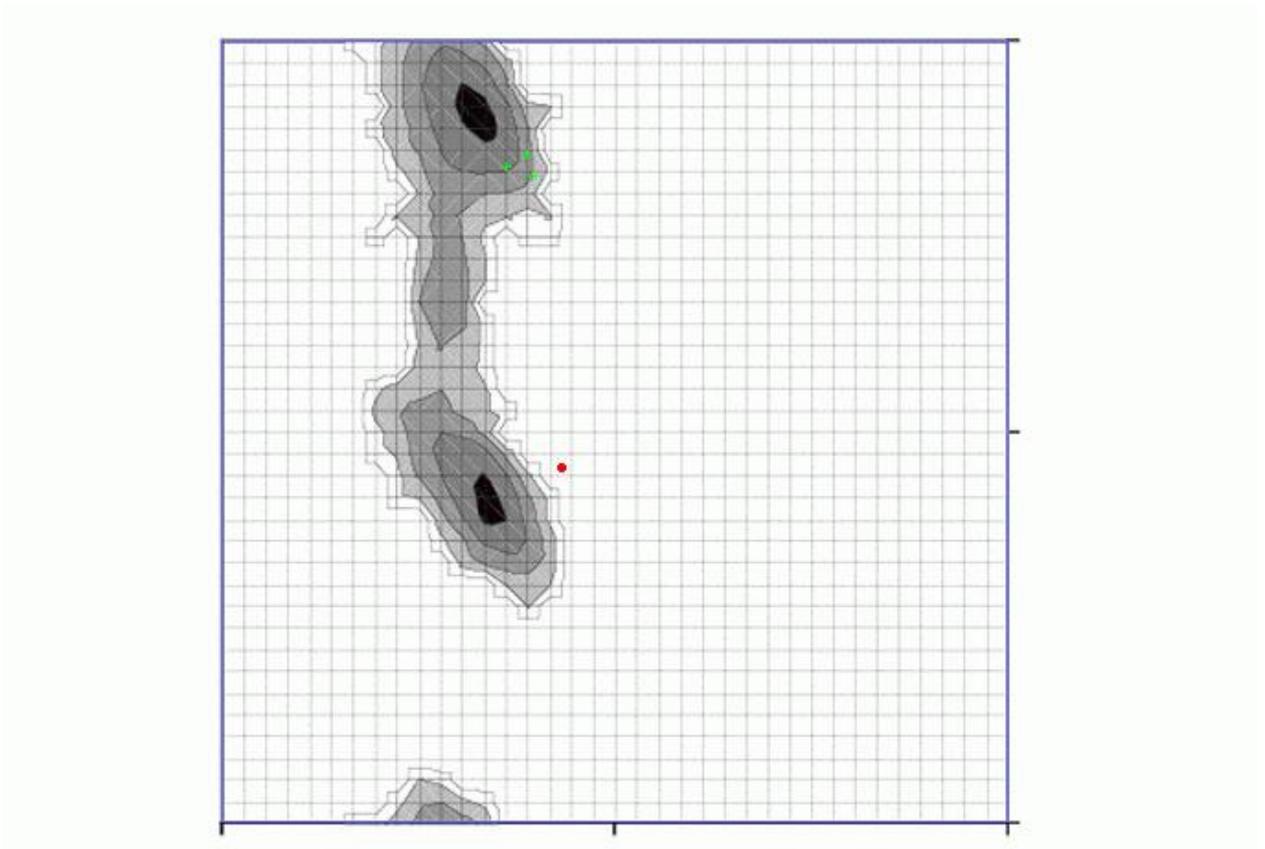


Рисунок 6. Карта раманчадрана для пролинов полипролинового участка

SeqC:

Для лучшей модели seqC Alpha фолд также оправдал ожидания: при замене всех остатков альфа-спирали на пролины на её месте была предсказана PolyPro спираль, соединяющая оставшиеся неизменными концевые участки белка. Все пролины попали в разрешенную область, более того, для всех кроме одного углы близки к идеальным.

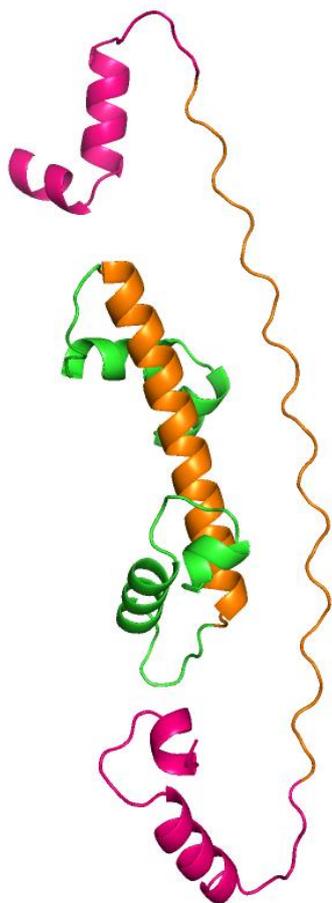


Рисунок 7. Лучшие предсказания для seqA (зеленым) и seqC (розовым). Отличающиеся участки покрашены в оранжевый

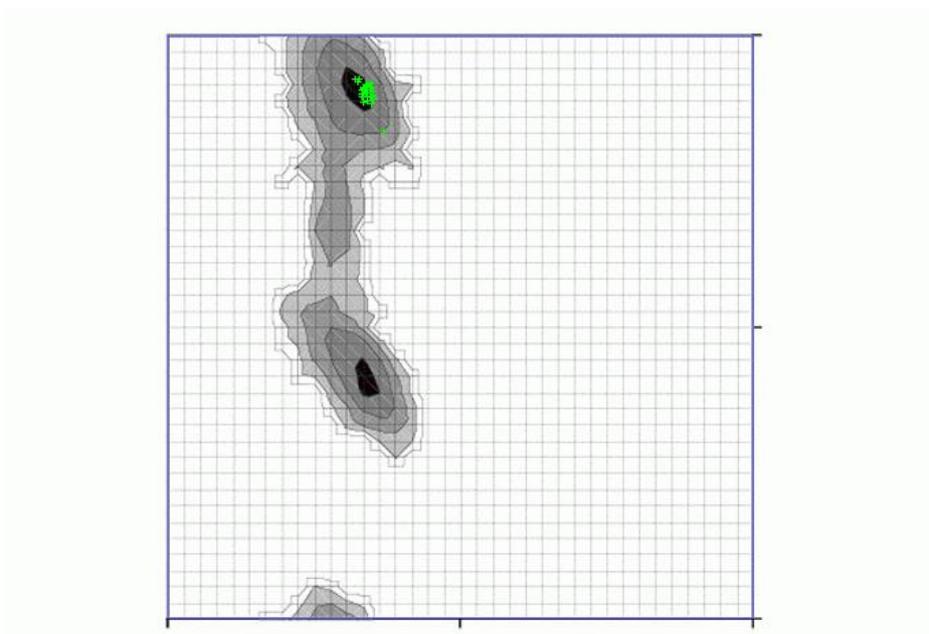


Рисунок 8. Карта Раманчадрана для пролинов отличающегося участка

Таким образом, на мой взгляд AlphaFold2 хорошо справился с предсказанием структуры с большим полипролиновым участком.

Таким образом, AlphaFold2 хорошо справился с задачей предсказания реальной структуры и на мой взгляд построил достаточно реалистичные предсказания для последовательностей как с небольшим числом замен, так и с заменой целой альфа-спирали.

[Выдача для seqA](#)

[Выдача для SeqB](#)

[Выдача для seqC](#)