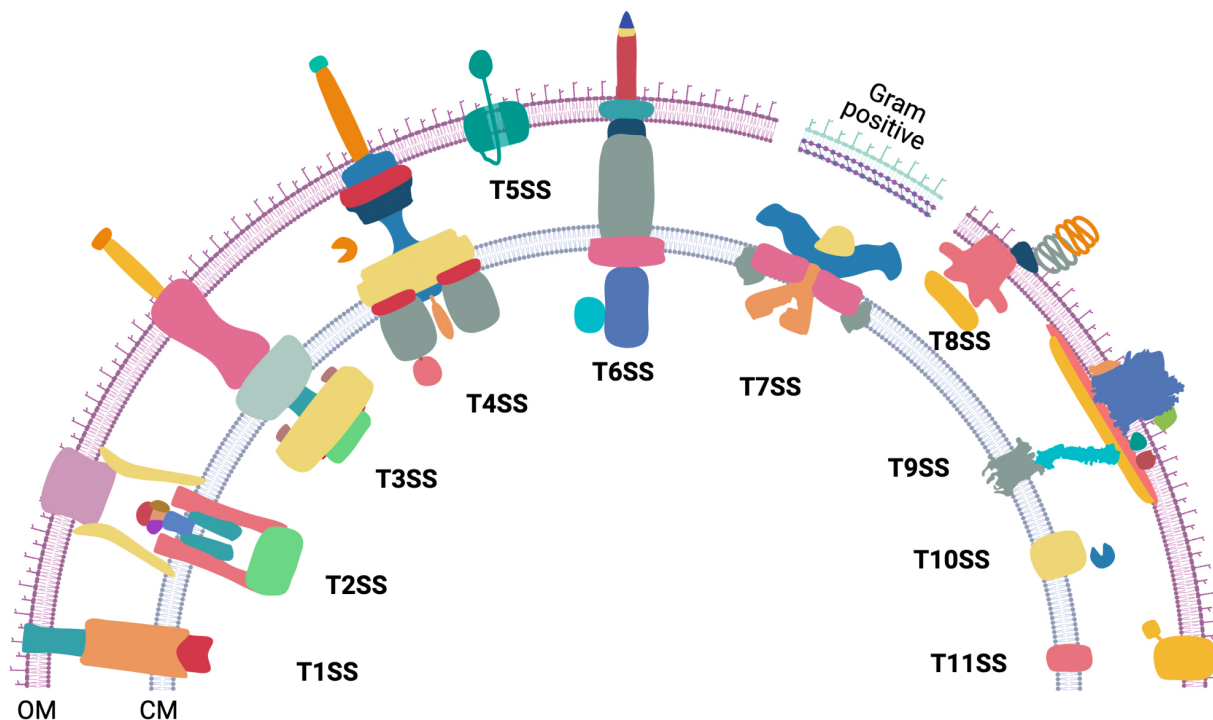


Pr6

β -спиральный белок

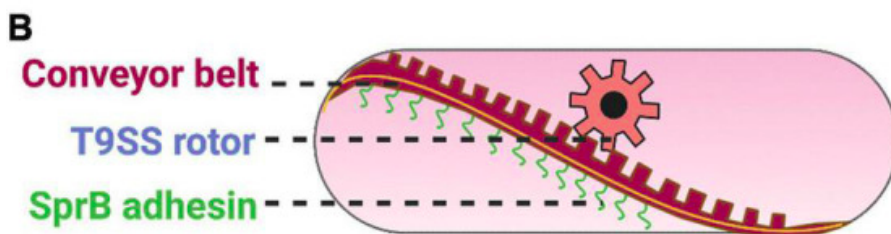
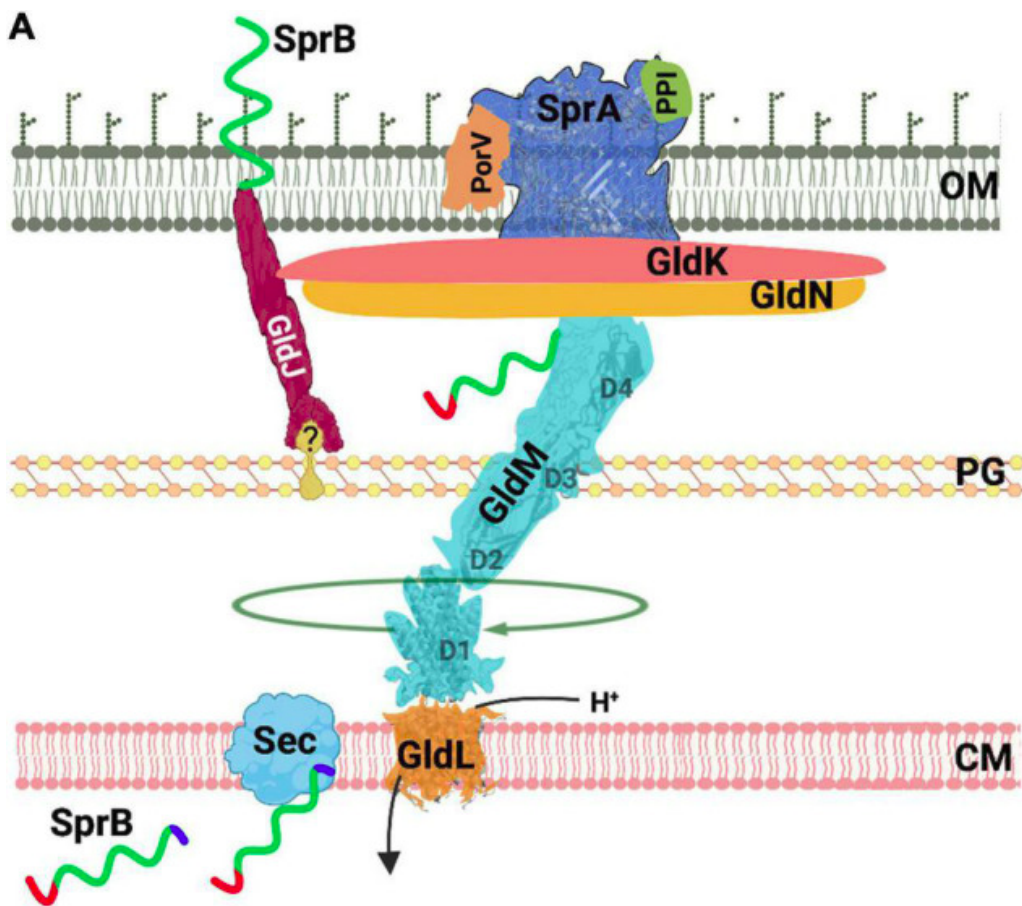
Мотивация

Где же искать трансмембранные β -листные белки, как не среди белков бактериальных систем секреции?



Бактериальные системы секреции, иллюстрация из [статьи](#).

Очень интересна система IX типа (T9SS, Type 9 secretion system) грамотрицательных бактерий, уникальная тем, что секреция в ней осуществляется за счёт работы... протонного мотора! Это одна из всего-навсего трёх моторных систем во всей биологии (ещё две - АТФ-синтаза и бактериальный жгутик).



В качестве блока переноса из цитоплазмы в периплазму T9SS использует белки системы Sec. Белком, обеспечивающим протон-зависимое вращение, на данный момент считается GldL. Функцию белка-транслокона в этой системе исполняет гигантский белок SprA, в его состав входит более 2000 аминокислот. Вот им мы сегодня и займёмся.

Что-то описательное про белок

Параметр	Значение
Name	T9SS-translocon
Название	T9SS-транслокон
UniProt ID	UP000006694
PDBID	6H3I

Координаты трансмембранных участков с OPM

Субъединиц 2, но приведу координаты для большей, поскольку предсказание будем проводить для неё.

A- Tilt: 2

TM segments:

1(140- 150),
 2(168- 179),
 3(190- 195),
 4(202- 208),
 5(219-223),
 6(242- 249),

```
7( 252- 260) ,
8( 756- 766) ,
9( 770- 779) ,
10( 799- 808) ,
11( 829- 839) ,
12(1520-1528) ,
13(1533-1540) ,
14(1564-1569) ,
15(1582-1591) ,
16(1636-1645) ,
17(1660-1667) ,
18(1685-1694) ,
19(1728-1735) ,
20(1764-1772) ,
21(1780-1785) ,
22(1821-1827) ,
23(1842-1849) ,
24(1881-1887) ,
25(1945-1954) ,
26(2014-2022) ,
27(2029-2034) ,
28(2175-2183) ,
29(2199-2207) ,
30(2253-2258) ,
31(2263-2271) ,
32(2291-2298) ,
33(2322-2329) ,
34(2353-2361) ,
35(2369-2376) ,
36(2391-2399)
```

Разумеется, нам теперь интересно сравнить эту информацию с выдачей DeepTMHMM.

Запуск

Ставить локально - не очень приятное занятие, поэтому воспользуемся версией для терминала, которая отправляет `fasta`-файл для работы на сервер:

```
pip install pybiolib
biolib run DTU/DeepTMHMM --fasta Q5I6C7.fasta
```

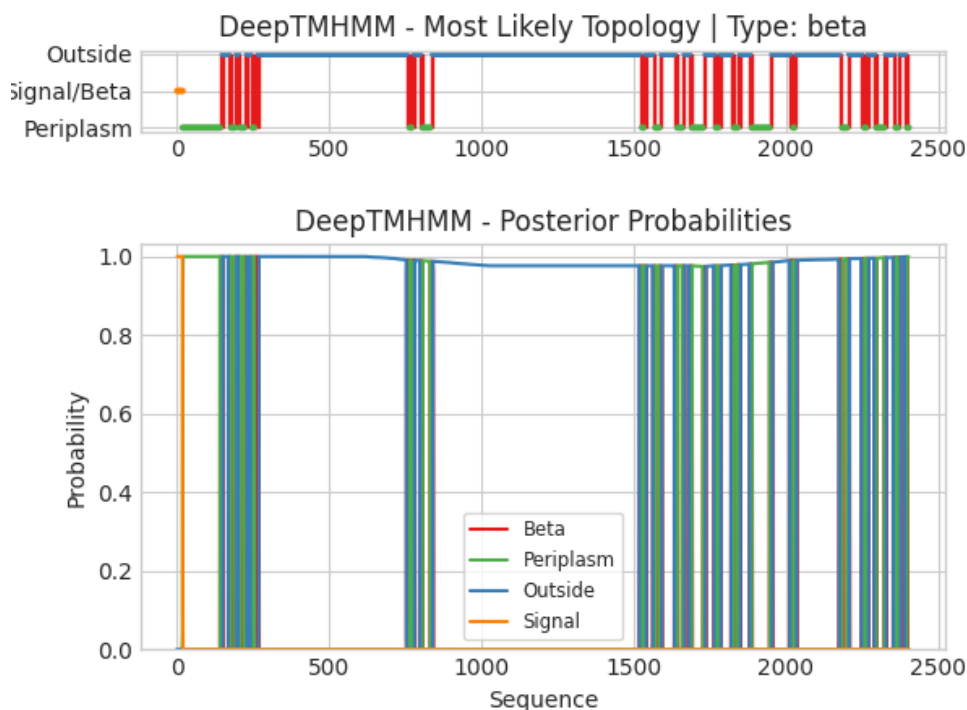
Результаты

```
signal 1      18
periplasm    19      141
Beta sheet  142     150
outside 151   168
Beta sheet  169     177
periplasm   178     190
Beta sheet  191     196
outside 197   200
Beta sheet  201     207
periplasm   208     220
Beta sheet  221     230
```

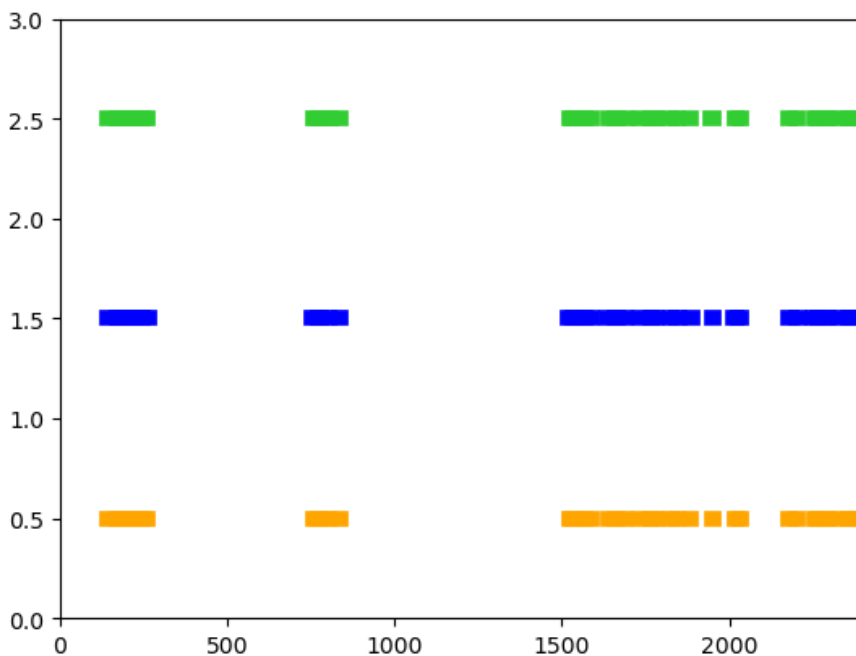
outside 231	240	
Beta sheet	241	247
periplasm	248	253
Beta sheet	254	268
outside 269	753	
Beta sheet	754	764
periplasm	765	770
Beta sheet	771	780
outside 781	797	
Beta sheet	798	806
periplasm	807	831
Beta sheet	832	840
outside 841	1518	
Beta sheet	1519	1527
periplasm	1528	1532
Beta sheet	1533	1541
outside 1542	1562	
Beta sheet	1563	1570
periplasm	1571	1583
Beta sheet	1584	1592
outside 1593	1633	
Beta sheet	1634	1643
periplasm	1644	1659
Beta sheet	1660	1667
outside 1668	1682	
Beta sheet	1683	1691
periplasm	1692	1727
Beta sheet	1728	1736
outside 1737	1763	
Beta sheet	1764	1770
periplasm	1771	1777
Beta sheet	1778	1786
outside 1787	1819	
Beta sheet	1820	1829
periplasm	1830	1841
Beta sheet	1842	1850
outside 1851	1880	
Beta sheet	1881	1888
periplasm	1889	1946
Beta sheet	1947	1955
outside 1956	2012	
Beta sheet	2013	2023
periplasm	2024	2027
Beta sheet	2028	2035
outside 2036	2174	
Beta sheet	2175	2184
periplasm	2185	2200
Beta sheet	2201	2208
outside 2209	2247	
Beta sheet	2248	2257
periplasm	2258	2263
Beta sheet	2264	2271
outside 2272	2289	
Beta sheet	2290	2297
periplasm	2298	2322
Beta sheet	2323	2329

outside	2330	2352
Beta sheet	2353	2360
periplasm	2361	2368
Beta sheet	2369	2376
outside	2377	2390
Beta sheet	2391	2398
periplasm	2399	2403

Выглядит внушающе, но давайте лучше посмотрим на графическую версию:



Верхняя часть графика отображает, какую топологическую аннотацию DeepTMHMM присваивает каждому из участков белка. Нижняя картинка отображает степень уверенности выданных предсказаний. Видно, что модель классифицирует все участки белка очень уверенно. Если считать трансмембранными все интервалы, которые DeepTMHMM классифицирует как Beta sheet, можно посмотреть на пересечения информации из БД и того, что предсказала модель. Следующая картинка показывает нам, что предсказания НММ, размещённые на OPM, очень похожи на предсказания DeepTMHMM:



По оси x - вся длина белка, над которой нарисованы трансмембранные интервалы. Зелёный цвет отвечает НММ, синий - нейронной сети, оранжевый - их пересечению.

α -спиральный белок

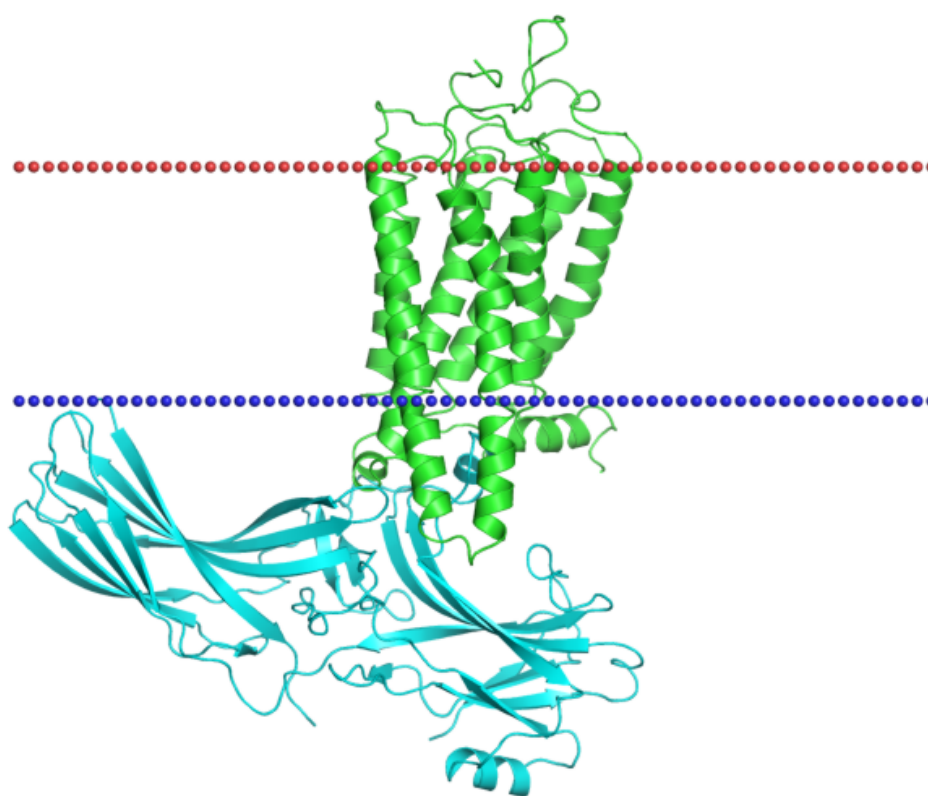
Описательное

Параметр	Значение
Name	Human rhodopsin
Название	Родопсин человека
UniProt ID	OPSD_HUMAN

Координаты с OPM

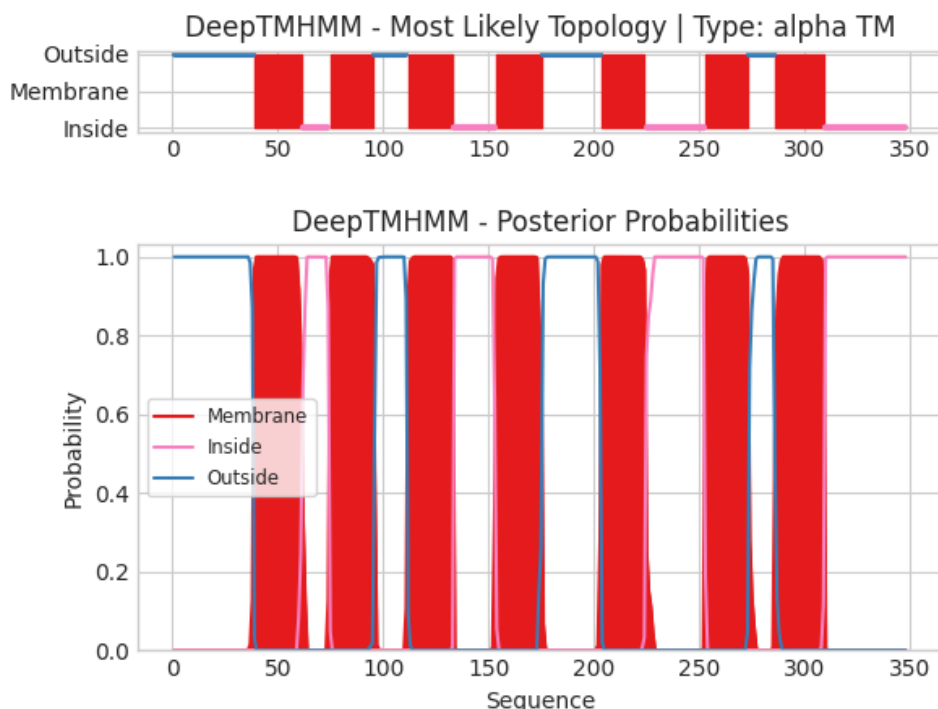
До настоящего времени мне так и не пришли результаты предсказания PPM (по всей видимости, проблемы на сервере), поэтому, по договорённости с Дарьей Владимировной, я взял уже готовые TM-координаты для этого белка с OPM:

```
A - Tilt: 9
TM segments:
1(35-63),
2(73-99),
3(109-134),
4(151-173),
5(201-224),
6(253-277),
7(286-309)
```



DeepTMHMM

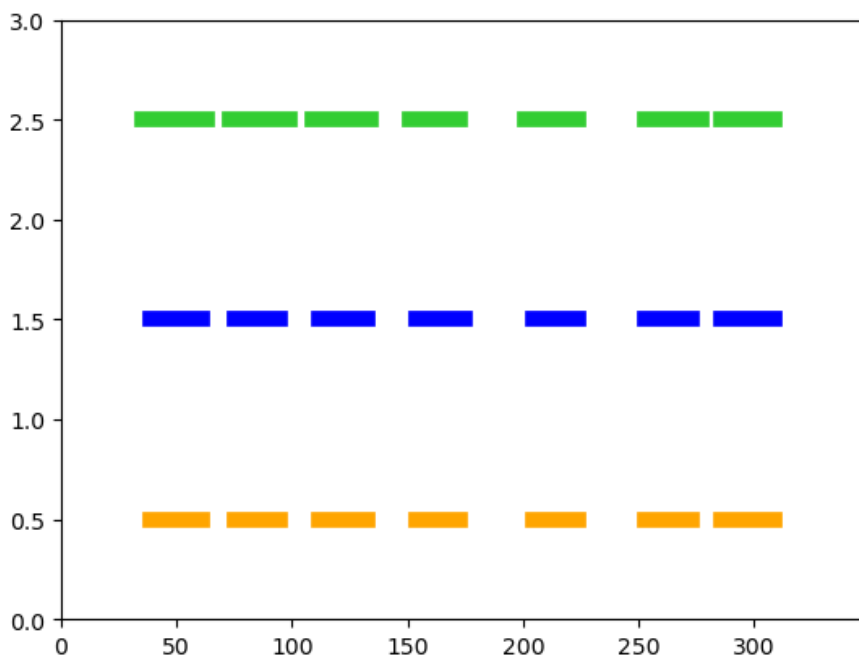
Координаты предсказанных трансмембранных участков:



Участки, погружённые в мембрану, отмечены красным, вне мембраны - розовым, внутри клетки - синим цветом.

Сравнение результатов

Как и в прошлый раз, рисуем картинку с пересечениями:



По оси x - вся длина белка, над которой нарисованы трансмембранные интервалы. Зелёный цвет отвечает НММ, синий - нейронной сети, оранжевый - их пересечению.

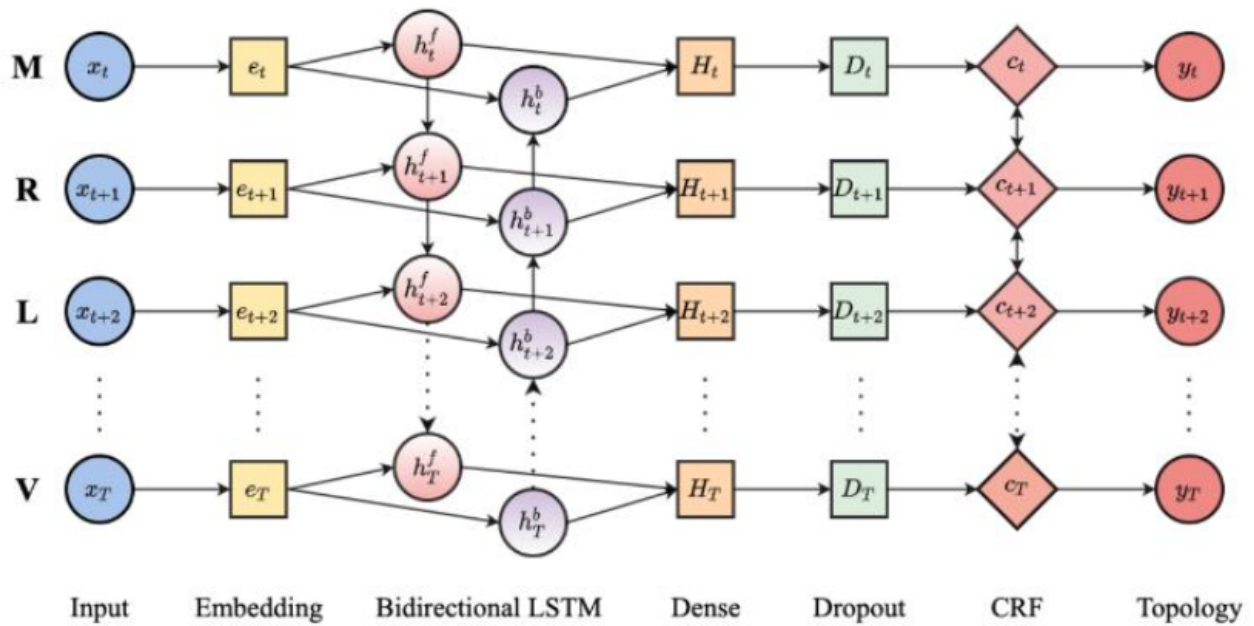
Видно, что обе модели предсказали примерно одно и то же, но нейронная сеть для каждого интервальчика "нагадывает" немного меньше. При этом мы вновь видим, что ситуаций, когда одна модель заметила ТМ-участок, а другая - нет - не возникает.

Вероятно, хотя бы для родопсина хорошую сходимость результатов можно объяснить тем, что этот белок широко известен \implies изучен \implies скорее всего, был в обучающей выборке DeepTMHMM.

Мысли DeepTMHMM

- удалось найти только [препринт](#) на bioRxiv'e, он не то что бы очень подробный
- [train](#), [выдача](#) с кросс-валидации

- в трейне около 3576 белков
- архитектура: аминокислотные эмбединги из ESM1b -> двунаправленная LSTM - > полносвязный слой с дропаутом-> CRF (conditonal random field, подобие марковской цепи):



- пока не очень понимаю, как тут текут градиенты при обратном распространении ошибки, но непосредственное соединение рекуррентной сети и подобия НММ выглядит очень интересно, никогда не видел такого раньше
- видимо, сейчас в архитектуре DeepTMHMM уже не RNN, а трансформер, во всяком случае так утверждают информационные сообщения при запуске через `biolib`, однако в препринте про трансформер ни слова (но про это можно найти аналогичные [статьи](#))