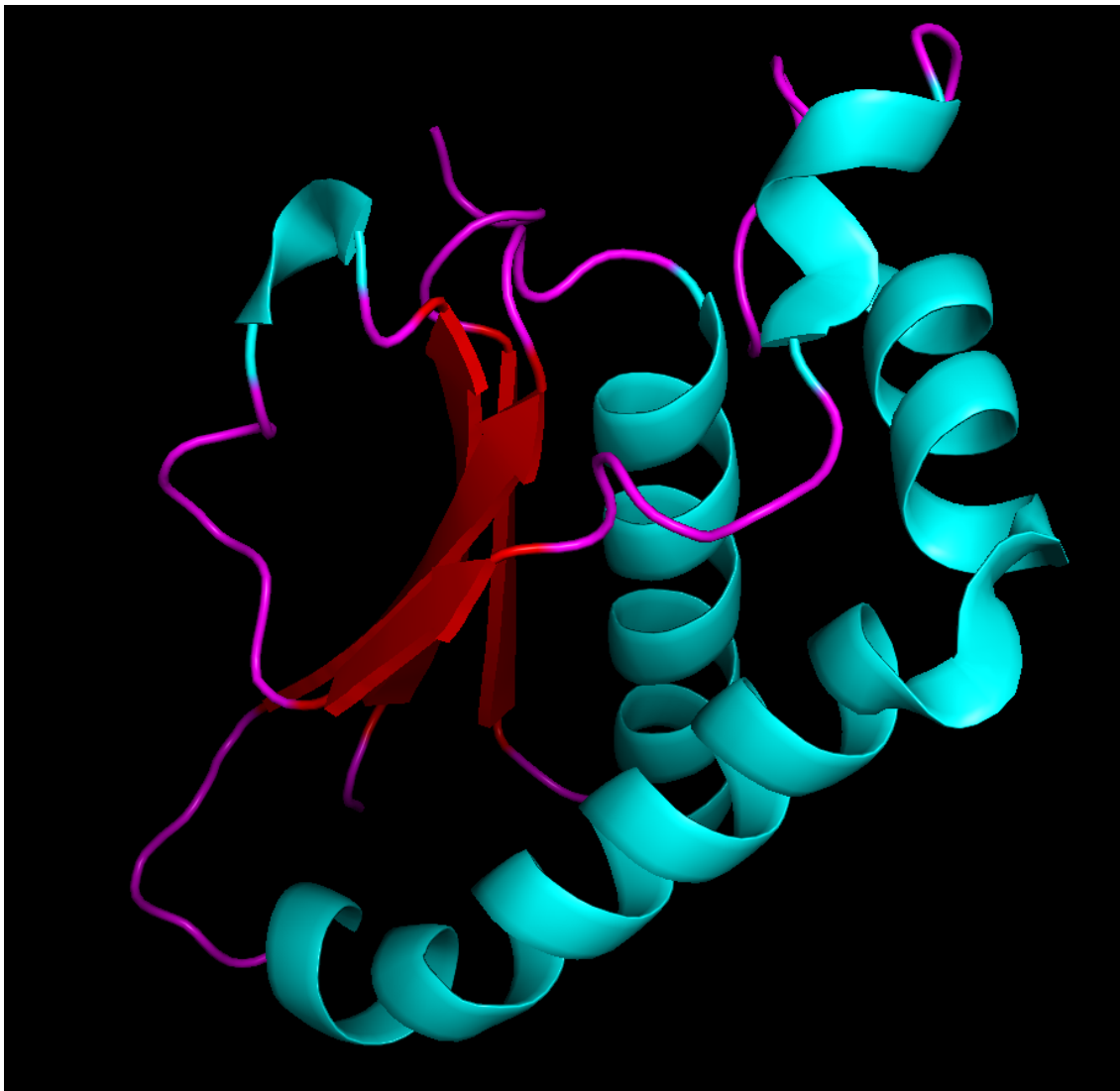


Определение вторичной структуры

В данном задании необходимо было определить вторичную структуру белка.

Необходимо было выбрать небольшую структуру, тем не менее содержащую интересные элементы вторичной структуры, желательно различные по длине, для того чтобы понять, как размер элемента структуры влияет на правильность предсказания.

Для определения была выбрана структура белка 5O5S. Это C терминальный домен RapZ из *E. coli*. Rap Z белок примечателен тем, что совсем недавно (2017-08-30) было обнаружено что он высококонсервативен среди бактерий поскольку он контролирует биогенез глюкоами фосфата. Данная структура была выбрана поскольку согласно аннотации rdb она имеет две 3/10 helix а также изгиб (coil) между двумя альфа-спиралями.



Белок 5O5S (раскраска по структуре).

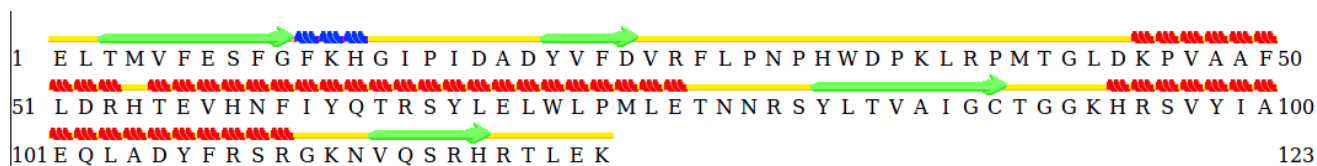
Далее структура была предсказана при помощи DSSP и Stride.

Для наглядности интересующие участки структуры были выделены цветом.



Белок 5O5S. 3/10 спираль 1 выделена синим, изгиб выделен желтым, 3/10 спираль 2 выделена оранжевым.

Сравним результаты, полученные при помощи STRIDE с аннотацией PDB.



Legend of secondary structure icons:

- | | |
|------------------------------------------|---------------|
| H Alpha-Helix | T Turn |
| E Extended Configuration (Beta-sheet) | C or " " Coil |
| B Isolated Beta Bridge | G 3-10 Helix |
| b Isolated Beta Bridge (Type 3 Fig 4,cd) | I Pi-Helix |

Результат выдачи STRIDE.



Аннотация PDB. Здесь указана полная структура, а не только 5о5s, поскольку C терминальный домен был расшифрован отдельно.

Примечательно то, что с помощью и DSSP и STRIDE удалось точно предсказать координаты 3/10 helix. STRIDE также предсказал положение изгиба (C, coil) в альфаспирали, DSSP также обнаружил изгиб, но выделил две позиции вместо одной, T (hydrogen bonded turn) и S (bend).

DSSP смог предсказать положение 3/10 спирали, указанной оранжевым (G). Однако STRIDE не смог предсказать положение 3/10 спирали, указанной оранжевым, и он указан как луп, что в данном случае классифицируется как T (Turn).

Интересно то, что именно в этом месте (там где 3/10 спираль указанная оранжевым) согласно аннотации pdb наблюдаются проблемы с расшифровкой структуры (poor fit to the electron density).

В pdb аннотации указано присутствие РНК – связывающего домена с координатами [266-281](#).

Внутри этого РНК-связывающего домена находится бета-лист с координатами 271-276.

Этот бета-лист был предсказан алгоритмом DSSP (как E - extended strand, participates in β ladder) но не алгоритмом STRIDE. Причем внутри этого бета-листа находятся участки, про которые в аннотации pdb validation указано что они имеют “проблему”.

Исходя из этого можно сделать вывод что STRIDE указывает более точные координаты при хорошем качестве расшифровки структуры, а DSSP способен предсказать тип вторичной структуры (альфа-спираль, бета-лист) более точно при худшем качестве расшифровки.

Различия в выдаче алгоритмов STRIDE и DSSP обусловлены принципиальными различиями в самих алгоритмах. Алгоритм DSSP основан на нахождении паттернов водородных связей, для их поиска используется электростатическая модель - водородные связи образуются только за счет электростатических взаимодействий, энергию водородной связи можно рассчитать по формуле). При нахождении водородных связей в алгоритме STRIDE используется не какая-то электростатическая модель, а статистика расстояний и углов, собранная по известным структурам (вторичные структуры которых были аннотированы экспертами).

Структура	STRIDE	DSSP	PDB	Качество расшифровки
Бета-лист	161-168	161-168	161-168	+
3/10 спираль	169-171	169-171	169-171	+
Бета-лист	179-182	179-182	179-182	+
3/10 спираль		289-291	289-291	-
Альфа-спираль	203-211	203-210	203-210	+
Альфа-спираль	213-234	213-233	213-233	+
Бета-лист	240-247	240-247	240-247	+
Альфа-спираль	252-268	253-267	253-267	+
Бета-лист	272-276	271-276	272-276	-

Сравнительная таблица координат элементов вторичной структуры