

Задание 1. Работа с разметкой вторичной структуры в ручном режиме

Для изучения разметки вторичной структуры я взяла структуру 3LDH из практикума 6.

Для нее была получена аннотации вторичной структуры с помощью [2Struc](#) и алгоритмов DSSP и STRIDE.

DSSP: <input checked="" type="checkbox"/>	STRIDE: <input checked="" type="checkbox"/>	PSEA: <input type="checkbox"/>	STICKS: <input type="checkbox"/>										
RESNUM	76	86	96	106	116	126	136						
SEQ	:TKQIDVYRSVVCPIILDEVIMGYNCTIFAYGQTGTGKFTMEGRSPNEEYTWEEDPLAGIIPRT												
CONSENSUS	①	00	HHHH	XXXXX	HHHHHHHH	X00	EEEEEEEE	00000	HHHHH	X0000	XXXX	0000000000	HHHHH
DSSP	①	00	HHHH	00000	HHHHHHHH	000	EEEEEEEE	00000	HHHHH	00000	HHH	0000000000	HHHHH
STRIDE	①	00	HHHHHHHHHHHHHHHHHH	HHHHHHHHHHHHHHHHHH	HHHHHHHHHHHHHHHHHH	HHHHHHHHHHHHHHHHHH	HHHHHHHHHHHHHHHHHH	HHHHHHHHHHHHHHHHHH	HHHHHHHHHHHHHHHHHH	HHHHHHHHHHHHHHHHHH	HHHHHHHHHHHHHHHHHH	HHHHHHHHHHHHHHHHHH	HHHHH

DSSP: <input checked="" type="checkbox"/>	STRIDE: <input checked="" type="checkbox"/>	PSEA: <input type="checkbox"/>				
RESNUM	196	200				
SEQ	KRGVIKGLEEITVHN					
CONSENSUS	①	000	EEEE	0000	EE	X00
DSSP	①	000	EEEE	0000	EE	000
STRIDE	①	000	EEEE	0000	EEEE	000

Рис.1 Аннотация тех участков вторичной структуры, которые будут рассмотрены ниже. (На рисунке представлена только разметка альфа - спиралей (H) и бета листов - (E))

Нельзя сказать, что разметки получились очень похожими даже по наиболее важным фрагментам структуры.

Рассмотрим участки, которые две программы аннотировали по-разному (Рис.2-4):

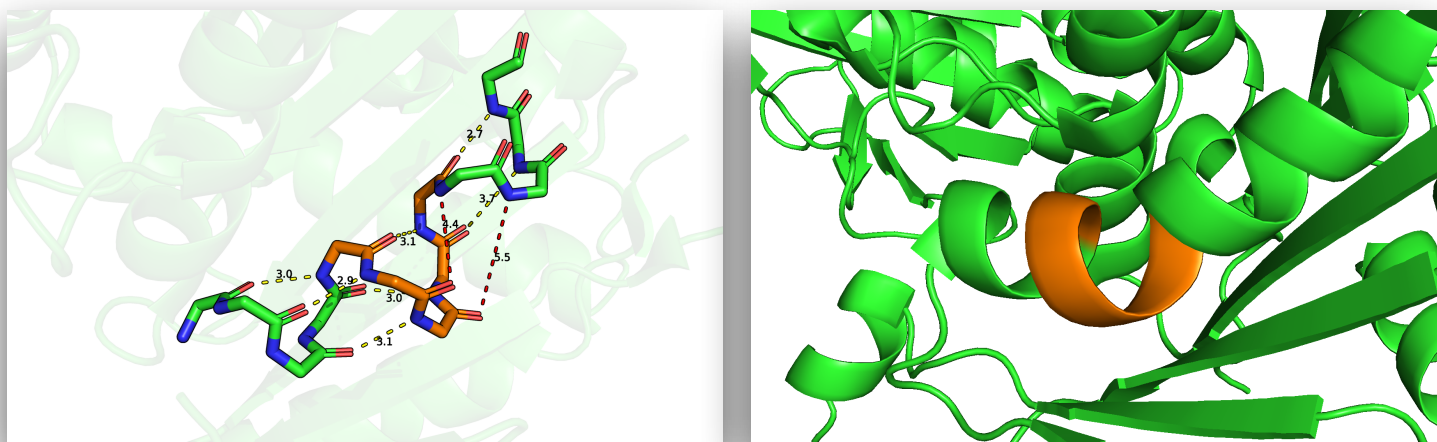


Рис.2 Слева: Оранжевым выделены остатки 82-86, для которых разметка различается. Справа: Разметка этого участка структуры ее авторами.

Для остатков 82-86 алгоритм DSSP показывает отсутствие какой-либо вторичной структуры. Однако STRIDE и авторы структуры размечают этот участок, как альфа-спираль. Водородных связей для остатков 83-85 действительно нет. Значит DSSP правильно аннотирует этот участок.

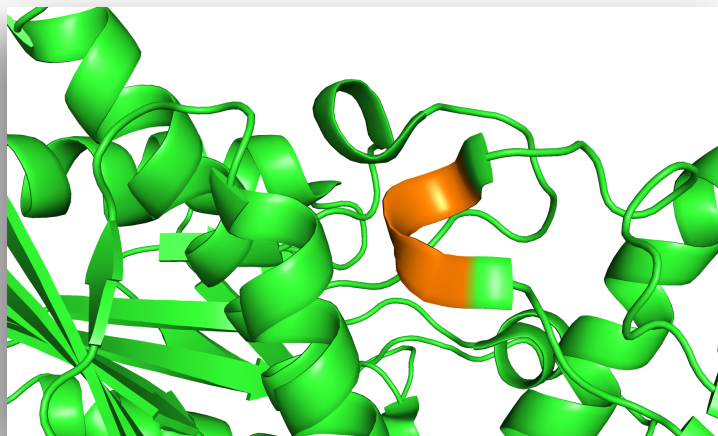
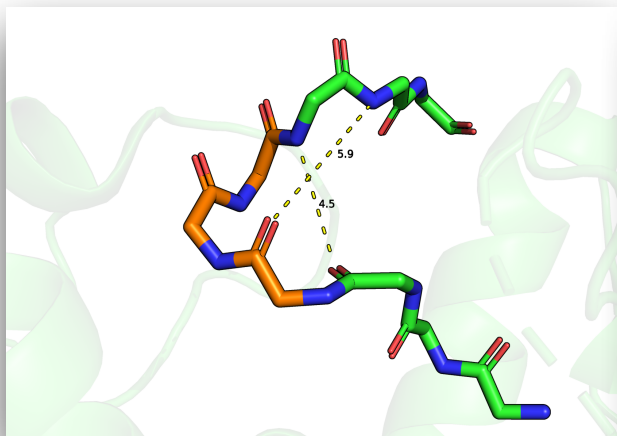


Рис.3 Слева: Оранжевым выделены остатки 121-123, для которых разметка различается. Справа: Разметка этого участка структуры ее авторами.

Наблюдается также различие в 121-123 позициях между разметками. DSSP предполагает наличие альфа спирали, а STRIDE ее отсутствие. На рисунке 3 нет водородных связей, а значит и альфа спирали. В данном случае STRIDE правильнее классифицировал вторичную структуру. Получается, чтобы точно определить, есть ли альфа спираль на данном участке в спорном случае, нельзя положиться ни на один из этих алгоритмов. Требуется самостоятельно внимательно изучить структуру.

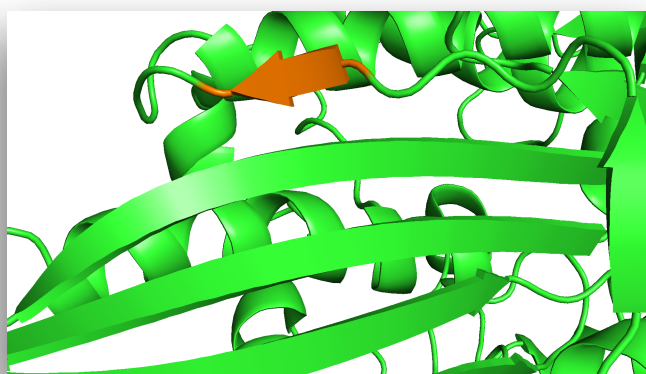
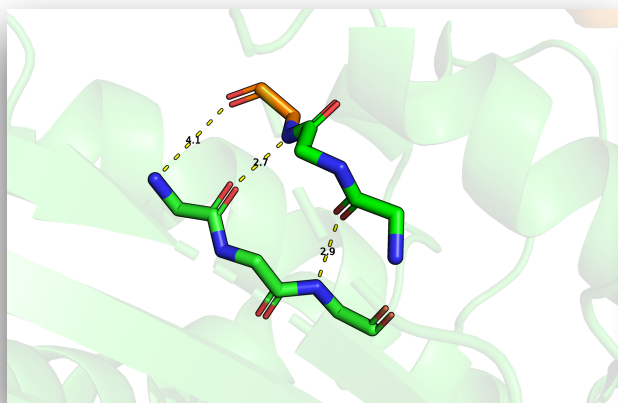


Рис.4 Слева: Оранжевым выделены остаток 204Val, для которого разметка различается. Справа: Разметка этого участка структуры ее авторами.

На рисунке 4 изображено последнее рассматриваемое различие. STRIDE и авторы статьи считают, что остаток валина 204 принадлежит бета-листу. А DSSP аннотирует его, как неструктурированный. Действительно, 204Val не удовлетворяют критерию бета листа так как не образует водородную связь с соседним бета-листом. Вновь STRIDE и авторы структуры неточны. По первому и по этому примеру можно подумать, что STRIDE более склонен ошибочно продлевать паттерны вторичных структур.

Задание 2. Работа с разметкой вторичной структуры в автоматическом режиме

В этом задании я буду оценивать склонность аминокислотных остатков образовывать ту или иную вторичную структуру по набору из 30 белков, используя следующую формулу с `codomo`:

$$P_{ik} = (n_{ik}/n_i) / (N_k/N)$$

Где P_{ik} это propensity аминокислотного остатка i образовывать тип вторичной структуры j

n_{ik} это количество остатков i в датасете, образующих тип вторичной структуры j

n_i это общее количество остатков i в датасете

N_k это общее количество остатков, образующих тип вторичной структуры j во всем датасете

N это общее количество остатков в датасете

Классификация вторичной структуры на альфа-спирали ('H'), бета-листы ('E') и петли ('C') при этом осуществляется с помощью `dssp`.

Таблица 1 - Склонности аминокислотных остатков образовывать определенные типы вторичной структуры:

	P	A	W	D	V	Y	E	G	H	F	L	K	M	I	R	S	T	N	Q	C
E	0.46	0.86	1.44	0.55	1.86	1.37	0.56	0.81	0.83	1.28	1.11	0.58	0.55	2.01	0.81	0.72	1.32	0.72	0.61	1.72
H	0.39	1.32	1.07	0.83	0.95	1.02	1.36	0.57	0.74	1.04	1.34	1.25	1.02	1.04	1.17	0.77	0.74	0.64	1.20	0.91
C	1.72	0.76	0.80	1.30	0.77	0.87	0.81	1.45	1.29	0.88	0.65	0.91	1.13	0.64	0.90	1.29	1.13	1.42	0.94	0.85

Заметно чаще образуют альфа спираль аланин, лейцин, глутамин, глутамат и лизин. То есть 2 гидрофобные аминокислоты и 3 полярные, 2 из которых несут заряд при стандартных pH.

В образовании бета-листов чаще всего участвуют гидрофобные валин, изолейцин и гидрофильный цистеин.

А для петель характерны гибкий глицин, аспарагин и пролин, который как раз удобно использовать для поворотов в неструктурированных участках (петлях).