

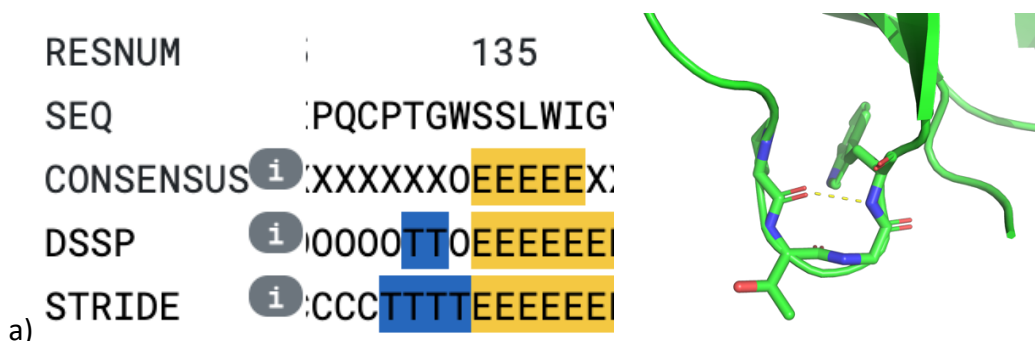
Задание 1. Работа с разметкой вторичной структуры в ручном режиме

Мы произвели разметку вторичной структуры используя сервис 2Struc. Программа предлагает несколько вариантов разметки, основанных на разных алгоритмах. Мы сравнили разметки, полученные с помощью DSSP и STRIDE. В целом они оказались очень похожими. Две разметки хорошо выравниваются по самым важным фрагментам структуры: бета листам и альфа цепям. Рассмотрим же несколько различий между двумя разметками.

Две разметки отличаются по 130 и 134 позициям (Рисунок 1, а). DSSP предполагает отсутствие структуры на этом участке, а STRIDE рассматривает это как регион поворота. Признаком поворота [<https://www.ncbi.nlm.nih.gov/labs/pmc/articles/PMC3904766/>] является наличие водородной связи между атомами остова остатков, судя по визуализации, между 130 и 134 остатками действительно есть водородная связь. Значит stride правильнее классифицировал этот участок. В этом случае проявляется способность лучше предсказывать вторичную структуру крайних аминокислот блока.

Существует также различие на 166 и 167 позициях между разметками. DSSP предполагает, что остатки образуют складку (bend), а stride, что они образуют поворот. Мы здесь не находим водородных связей между атомами остова – значит это не поворот. Однако при этом между C альфа атомами аминокислот высокие значения торсионных углов, что свидетельствует о том, что они согнуты (curved) – это свидетельствует в пользу того, что это складка. Значит DSSP делает лучшую разметку в этом случае. Это объясняется тем, что STRIDE не различает между поворотом и складкой.

И наконец последнее различие, которое мы рассмотрим, находится на 122 и 123 позициях. DSSP размечает их, как неструктурированный участок и фрагмент складки, а STRIDE, как конец бета-листа. Анализ структуры показывает, что эти остатки, только частично удовлетворяют критерию бета листа. Во-первых, только одна из аминокислот образует водородную связь с другим бета листом. Во-вторых, три из четырех торсионных углов совсем не соответствуют типичным для бета-листов углам. Значит DSSP в этом случае лучше. Этот случай является контрастом к первому, здесь склонность STRIDE удлинять блок остатков с одинаковой вторичной структурой ведет к ложным результатам.



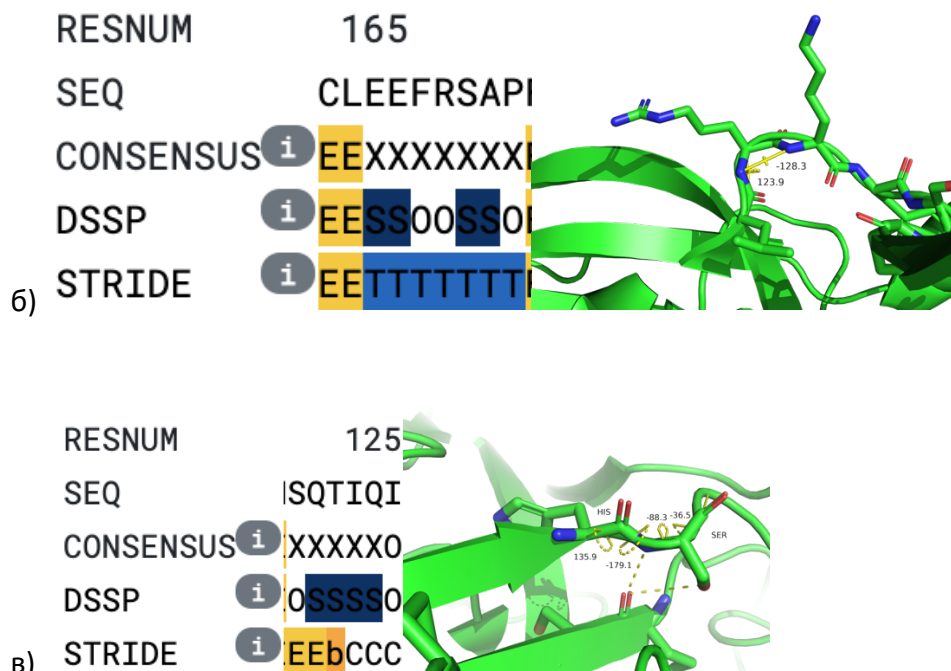


Рисунок 1. Различие в разметке вторичной структуры разными алгоритмами

Задание 2. Работа с разметкой вторичной структуры в автоматическом режиме

В этом задании нам предлагалось оценить склонность каждой из аминокислот к образованию разных типов вторичной структуры. Для этого мы сгенерировали dssp разметку для 30 белков. Склонность к образованию каждого типа вторичной структуры была рассчитана по следующей формуле: $P_{ik} = (n_{ik}/n_i) / (N_k/N)$. 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. Результаты представлены в таблице 1 и на рисунке 2.

Среди аминокислот, которые чаще всего образуют альфа-спирали оказались М, А, L, E, К, Q – они обладают достаточно прочной структурой, но при этом их радикалы не мешают образованию альфа-спирали. Реже всего в этом типе структур встречаются пролин и глицин. Первому мешает то, что он не может быть донором водородной связи с амидного атома, а также его конформация стерически ограничивает образование альфа петель. Глицин просто очень подвижная аминокислота и он будет дестабилизировать альфа-спираль за счет энтропийного фактора.

Среди аминокислот, которые более склонны образовывать бета-листы, как и ожидалось, нашлись ароматические аминокислоты F, W, Y и разветвленные аминокислоты I, V, T. Ароматические аминокислоты часто находятся в бета листах за счет того, что они могут образовывать стекинговые взаимодействия. Интересно, что на первом месте оказался цистеин. Хуже всего бета-листы образуют пролин, а также дикарбоновые и амиды дикарбоновых аминокислот. Конформация пролина ограничивает образование бета листов.

В петлях чаще всего находятся пролин и глицин. Это происходит из-за того, что они оказываются исключены из других типов вторичной структуры по вышеназванным причинам. Реже всего в петлях находятся те аминокислоты, которые склонны образовывать альфа-спирали и бета-листы.

АК	Альфа-спирали	Бета-листы	Петли
Q	1.3886825798489013	0.5526989609891164	0.9393193931939319
A	1.3552203490091688	0.9307076584097913	0.7934284243174659
E	1.3564546116221097	0.557847051162264	0.9585677414151191
N	0.6842145621703211	0.4720419355816138	1.4460574868906584
I	1.0230584987618236	1.7456568119475773	0.6526888430649013
K	1.232714554748453	0.7242996399402829	0.9671805616361249
V	0.9449120184906139	1.7559861421957876	0.7003209677067186
T	0.7297340340818602	1.326416479620734	1.035342934401004
L	1.376395666962437	1.1234548482605478	0.6934818957564576
M	1.380670496408402	0.8025103991826608	0.83348058832701
R	1.2040611562350692	0.6391789557592666	1.0242194152710757
Y	1.033392422991741	1.3034237529208577	0.8426247497769096
S	0.7224850204982656	0.8490134852280137	1.2526850516849541
D	0.9068895920413558	0.5016951803695376	1.2840318874886862
G	0.44468167701863354	0.5935233160621762	1.552078528597786
F	0.8550795059224519	1.8895844348101936	0.7008952104827171
H	0.7905452035886819	0.8902849740932643	1.1888261070110702
P	0.2542456936164605	0.47720467623089546	1.7311326002707264
C	0.5929089026915114	1.9586269430051815	0.8453874538745388
W	0.7257464164092817	1.8681389620317677	0.7968816163571472

Таблица 1. Склонность разных аминокислот к образованию некоторых типов вторичной структуры

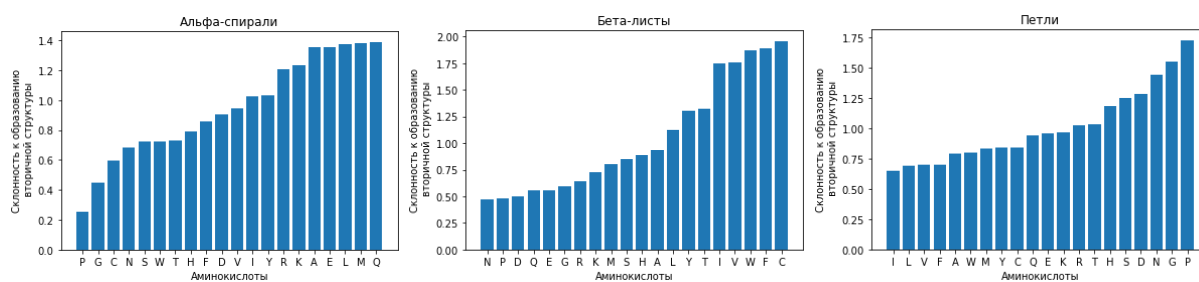


Рисунок 2. Склонность разных аминокислот к образованию некоторых типов вторичной структуры