

Практикум № 8. Вторичная структура

Иосиф Финкельберг

1 Введение

В первом задании этого практикума мы работали с PDB-структурой **1sbr**. Эта структура была получена с помощью рентгено-структурного анализа (РСА) в работе Jing J. He и Florante A. Quijano 1993-го года [1]. 1sbr является структурой периплазматического сульфат-связывающего ABC-транспортера (**Sulphate Binding Protein, SBP**) гамма-протеобактерии *Salmonella typhimurium*. SBP прочно и специфично связывает сульфат-ионы [2]. Именно эта его особенность и заинтересовала авторов работы [1]: они использовали SBP в связанном с заряженным субстратом (ионами сульфата) состоянии, как хорошую модель для изучения электростатических взаимодействий. Авторы предложили схему, объясняющую взаимодействия субстрата с белком (см. рис 1 и 2). В непосредственном связывании субстрата (иона сульфата) задействованы 7 водородных связей. Находящиеся недалеко от связанного иона SO_4^{2-} положительно заряженные аминокислотные остатки (Arg, His) участвуют в электростатическом взаимодействии с ним. Кроме того, авторы замечают, что в сторону иона SO_4^{2-} направлены своими N-концами три альфа-спирали, что, возможно, тоже частично компенсирует отрицательный заряд субстрата (такая модель альфа-спиралей называется “macro-dipole model”).

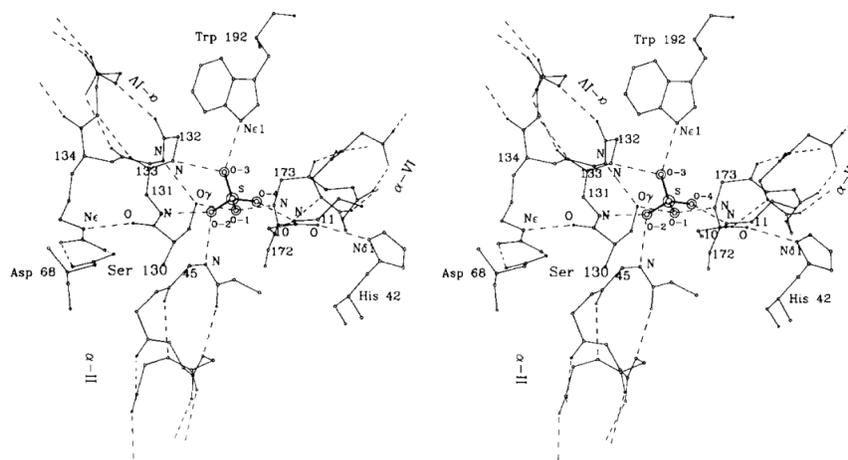


Рис. 1: Предложенная в статье [1] схема связывания SBP с субстратом (ионом SO_4^{2-}). Показаны три альфа-спирали, NH-группы остовов которых образуют водородные связи с кислородами субстрата. Спираль α -VI образует две водородные связи с атомами кислорода SO_4^{2-} . Остатки Trp-192 и Ser-130 образуют по одной водородной связи с субстратом своими боковыми цепями. Еще одну водородную связь с субстратом образует азот остова остатка Gly-131, который находится у границы альфа-спирали IV. Итого, в связывании субстрата непосредственно задействованы 7 водородных связей.

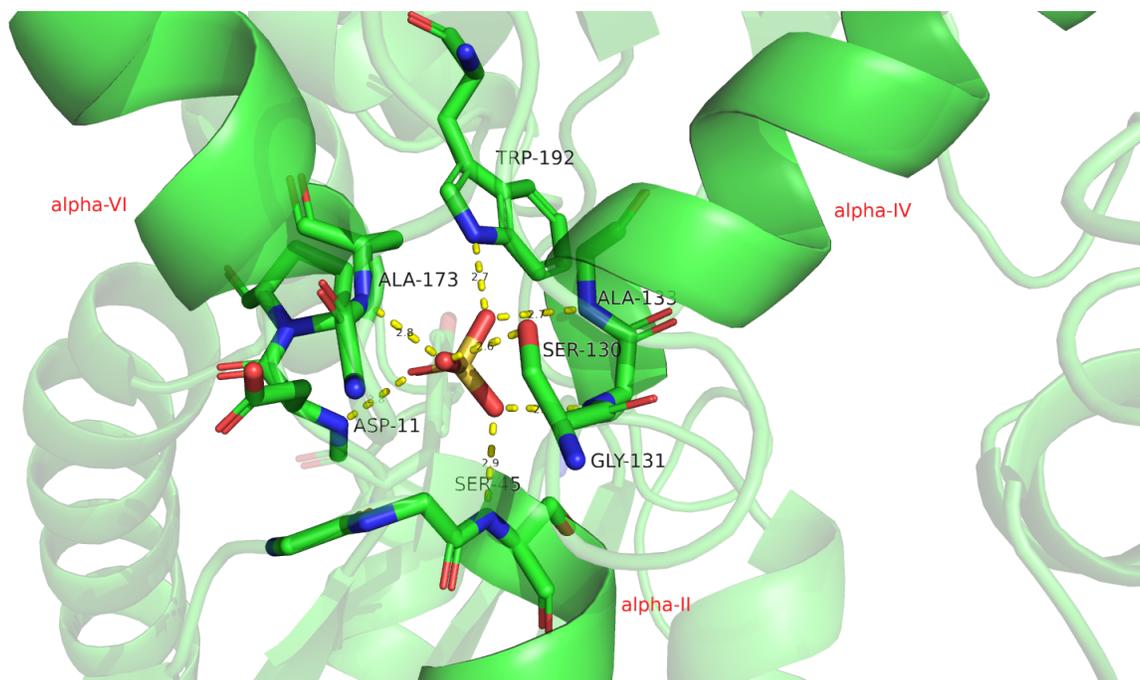


Рис. 2: Предложенная в статье [1] схема связывания SBP с субстратом (ионом SO_4^{2-}): изображение в rупол. Видим, что ключевыми для связывания с субстратом остатками можно в первом приближении считать: Asp-11, Ser-45, Ser-130, Gly-131, Ala-133, Ala-173, Trp-192.

2 Задание 1. Работа с разметкой вторичной структуры в ручном режиме

В этом задании мы воспользовались сервером [2StrucCompare](#) (и реализованном на нем методом 2Struc), чтобы разметить вторичную структуру в нашей PDB-модели. Сперва мы рассмотрели лишь разметку на α -спирали, β -листы. Результаты работы алгоритмов DSSP, STRIDE приведены на рисунке 3.

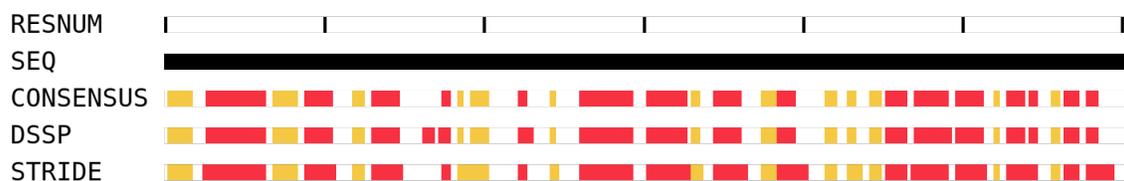


Рис. 3: Разметка вторичной структуры с помощью алгоритмов DSSP, STRIDE. Из элементов вторичной структуры на схеме приведены только α -спирали (красн.) и β -листы (желт.).

Сравним выдачу программ DSSP, STRIDE. Мы видим, что в общих чертах разметки совпадают, однако часто встречаются отличия на 1-4 аминокислотных остатков в длинах размеченных участков вторичной структуры. Отобразим теперь на схеме все элементы вторичной структуры, которые алгоритмы умеют определять (см. рис 4). Теперь мы видим больше различий, причем они в большинстве своем касаются различий между менее типичными элементами вторичной структуры (т.е. всеми, кроме α -спиралей и β -листов).

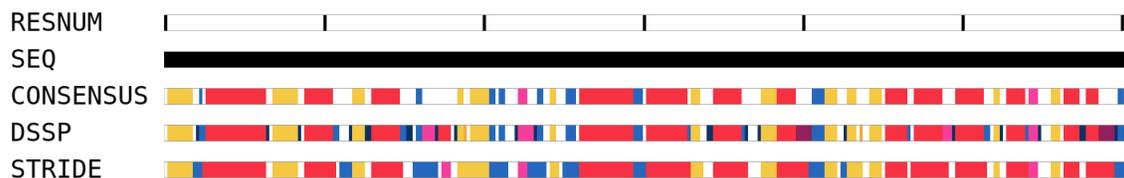


Рис. 4: Разметка вторичной структуры с помощью алгоритмов DSSP, STRIDE. Элементы вторичной структуры обозначены цветами: α -спираль (красный), 3_{10} -спираль (розовый), β -лист (желтый), β -мост (оранжевый), π -спираль (фиолетовый), поворот (turn, голубой), излом (bend, темно-синий).

Рассмотрим более детально некоторые из различий. Визуально кажется, что одно из самых значительных разногласий между алгоритмами заключается в разметке С-концевого участка белка (см. рис 5).



Рис. 5: Разметка С-концевого участка.

Сразу отметим, что хотя формально DSSP отмечает остатки Ala-299, Gly-302 как S (т.е. Bend, излом = резкое изменение направления хода цепи), а STRIDE размечает эти же остатки как T (т.е. Turn, поворот), реального разногласия здесь нет. Действительно, с точки зрения STRIDE-классификации Bend не существует. Фактически оба алгоритма как раз выдают одно и то же предсказание: между двумя спиралями (левая из спиралей α или π – это сейчас обсудим) имеется изгиб, петля.

При этом модели отличаются в своих предсказаниях типа спирали, предшествующей изгибу. Обе модели сходятся на том, что остатки 290-293 входят в состав альфа-спирали. Остатки 294-298 согласно STRIDE являются продолжением этой альфа-спирали, но согласно DSSP представляют собой π -спираль. Казалось бы, различие значительное – целых 5 остатков. Тем не менее, мы сейчас покажем, что противоречие, к счастью, становится не таким уж разительным (и состоит по большей части в формальных правилах разметки).

Мы видим, что по индексам в DSSP-альфа-спираль входит отрезок [290 : 293] (включительно). Однако по определению α -спирали кислород остатка Trp-290 должен образовывать водородную связь с азотом остатка Gln-294, который, заметим, по DSSP разметке уже не входит в альфа-спираль. Таким образом, можно сделать вывод, что в данном случае DSSP алгоритм отметил входящими в альфа-спираль только те остатки, атом кислорода в остовах которых участвует в образовании водородной связи этой альфа-спирали. Вместе с тем по определению π -спирали водородные связи, которые ее образуют, соединяют атомы остатков с индексами $i, (i + 5)$. Мы видим, что из 5 остатков (Gln-294, ..., Phe-298) только Gln-294 может образовывать кислородом своего остова водородную связь (с азотом остова Ala-299). Кстати, мы тут сталкиваемся с отсутствием общего правила для разметки спиралей в DSSP: действительно, если бы данная π -спираль размечалась так же, как предшествующая ей α -спираль, то размечен был бы только остаток Gln-294. Про STRIDE здесь сложно угадать, по какому правилу он красит остатки, входящие в альфа-спираль, однако разумно предположить,

что STRIDE помечает альфа-спиральными все те остатки, у которых в водородной связи, образующей спираль, участвует либо N, либо H атом. И так, судя по всему, все отличие состоит в предсказании одной водородной связи: DSSP предсказывает связь между кислородом Gln-294 и азотом Ala-299, STRIDE – между кислородом Gln-294 и азотом Phe-298. Давайте же проверим нашу гипотезу в PyMol (рис. 6).

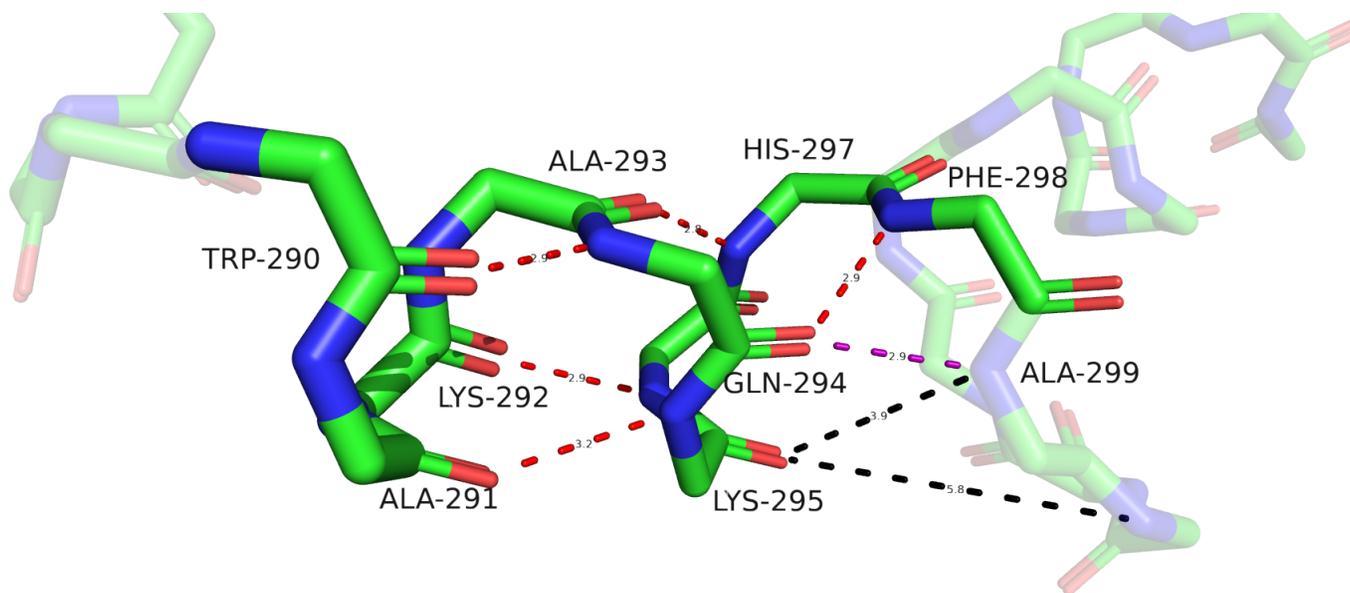


Рис. 6: Суровая реальность: вся разница между DSSP и STRIDE состоит в предсказании одной водородной связи: Gln-294(O) : Phe-298(N) (STRIDE) или Gln-294(O) : Ala-299(N) (DSSP). STRIDE предсказывает ее по альфа-спиральному типу (покрашена красным), DSSP – по π -спиральному типу (покрашена фиолетовым). Черным показаны расстояния между теми атомами, которые в теории могли бы продолжать α/π -спираль.

Да, действительно, наша гипотеза подтверждается. Интересно, что по расстояниям ни одна из связей (Gln-294(O) : Phe-298(N) или Gln-294(O) : Ala-299(N)) не является предпочтительной: обе имеют длину 2.9 Å. Про угол мы ничего не можем сказать, за неимением водородов в PCA-структуре. Мы отдаем предпочтение STRIDE (т.е. α -спирали) по той причине, что считаем глупым из-за наличия одной альтернативной водородной связи 5 атомов, входящих в состав спирали, переименовывать из α в π . Кроме того, настоящая π -спираль – явление редкое, при наличии двух равнозначных альтернатив стоит из эмпирических соображений выбирать более часто встречающуюся.

Нас заинтересовало также, как эту часть разметили авторы структуры

(кстати, структура была получена в 1993 году [1], т.е. за 2 года до того, как STRIDE-алгоритм был опубликован [4], но уже после опубликования DSSP-алгоритма [3]): см. рис. 7.

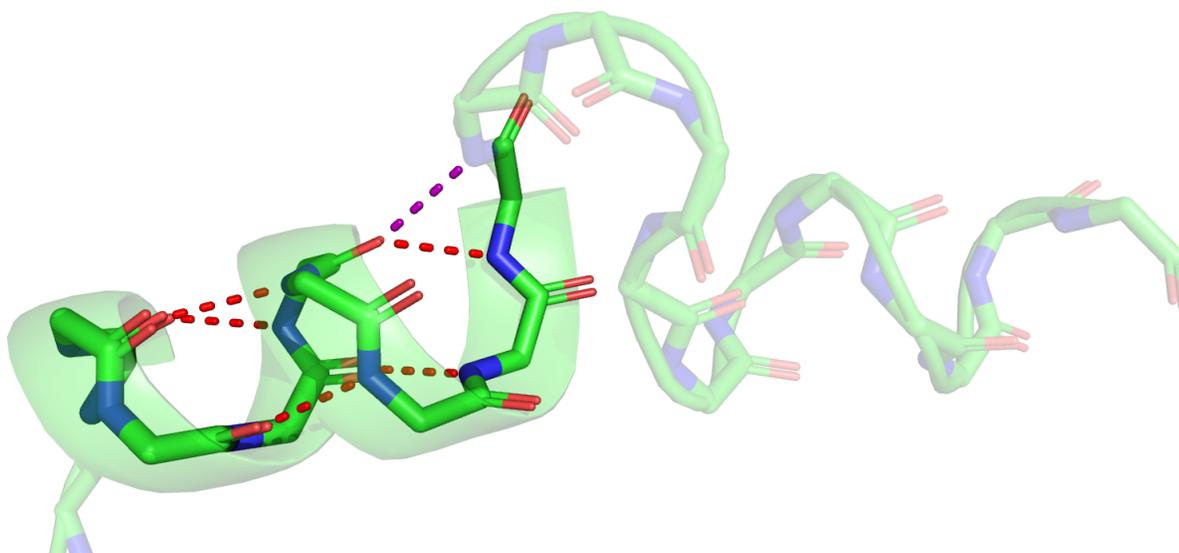


Рис. 7: Авторская разметка С-концевого участка структуры.

Похоже, что авторы разметили данный участок, как α -спираль (т.е. так же, как предлагает алгоритм STRIDE, который к тому времени еще не был написан). Любопытно, что С-концевой участок (на рис. 7 справа) оба алгоритма разметили, как α -спираль, и глазами мы видим то же самое. Однако авторы решили этот участок не размечать (скорее всего, потому что он уже никакого интереса не представляет для изучения связывания SO_4^{2-}). Тут же хорошо заметен предсказываемый обоими алгоритмами перегиб между двумя короткими спиральями.

Посмотрим теперь еще на пару отличий в предсказаниях. Разберем сначала похожий на предыдущий случай (см. рис. 8)

RESNUM	231	241	251	261
SEQ	DKVVEKKDTKAVAEAYLKYLSPYEGQEIAAKNFYRPR			
CONSENSUS	i HHHHHHHX	HHHHHHHHHXXX	HHHHHHHHHHX	HOEEHC
DSSP	i HHHHHHHT	HHHHHHHHHGGG	HHHHHHHHHTT	HOEEHC
STRIDE	i HHHHHHHC	HHHHHHHHHHHC	HHHHHHHHHHHC	HOEEHC

Рис. 8: Различия в DSSP- и STRIDE-разметках.

На этот раз опять DSSP в конце α -спирали (остатки 236-244) предсказывает еще другую спираль: 3_{10} -спираль (остатки 245-247). STRIDE предсказывает только одну длинную α -спираль (которая, к тому же, на один остаток раньше начинается: остатки 235-246). Посмотрим, как предлагаемые разметки выглядят в Rmol (см. рис. 9)

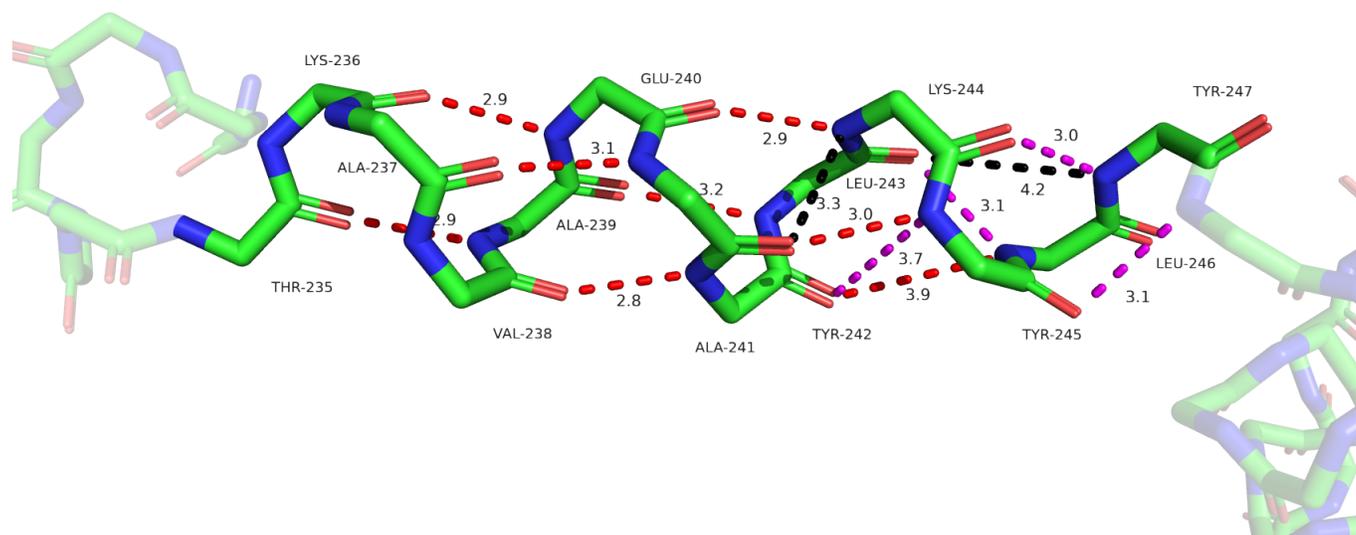


Рис. 9: Различия в DSSP- и STRIDE-разметках. Красным отмечены характерные для α -спирали водородные связи, розовым – характерные для 3_{10} -спирали водородные связи.левой черной пунктирной линией обозначена гипотетическая водородная связь (Ala-241(O) : Lys-244(N)), которая должна была быть, если бы 3_{10} -спираль продлевалась налево, правая черная пунктирная линия показывает ту связь (Leu-243(O) : Tyr-247(N)), которая имела бы, если α -спираль продлевалась направо.

Тут все интереснее. Мы видим, что ближе к N-концу водородные связи, характерные для α -спирали (т.е. $i, (i+4)$) являются вполне валидными. Однако ближе к C-концу спирали по расстоянию более выгодными становятся связи, характерные для 3_{10} -спирали (т.е. $i, (i+3)$): 3 правые розовые связи, показанные на рис. 9). В то же время, при попытке продлить связи 3_{10} -типа в направлении N-конца, они оказываются менее выгодными по расстоянию, чем связи α -типа (левая черная и самая левая розовая линии на рис. 9 и черные пунктирные линии на рис. 10).

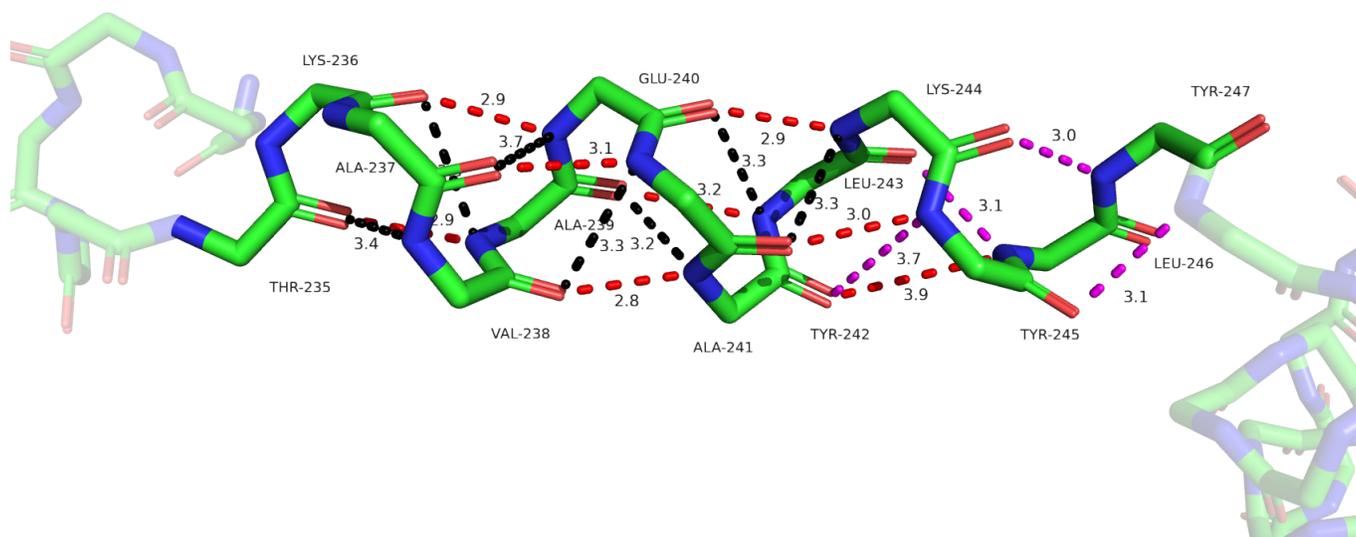


Рис. 10: Различия в DSSP- и STRIDE-разметках. Черными пунктирными линиями отмечены водородные связи 3_{10} -спирального типа.

Моментом перехода с α на 3_{10} спираль мы предлагаем считать бы остаток Leu-243. Нам кажется, что кислород остова этого остатка связывается водородной связью 3_{10} -типа с азотом остова Leu-246; при этом последняя (если считать с N-конца) водородная связь α -спирального типа представлена на азотом остова остатка Lys-244 и кислородом остова остатка Glu-240 (см. рис. 9, 10). Таким образом, на наш взгляд предсказание STRIDE в данном случае не вполне корректно, в отличие от предсказаний DSSP. Мы предполагаем, что так как 3_{10} -спираль является более редкой, чем α -спираль, то последняя имеет больший приоритет для STRIDE. Таким образом, STRIDE может отдавать ей предпочтение (заметим, что все водородные связи, соединяющие атомы остовов остатков 235-246, имеют длину меньше 4 Å, и, вероятно, с точки зрения STRIDE вполне допустимы).

Теперь рассмотрим несоответствие в предсказаниях β -тяжей (см. рис. 11)

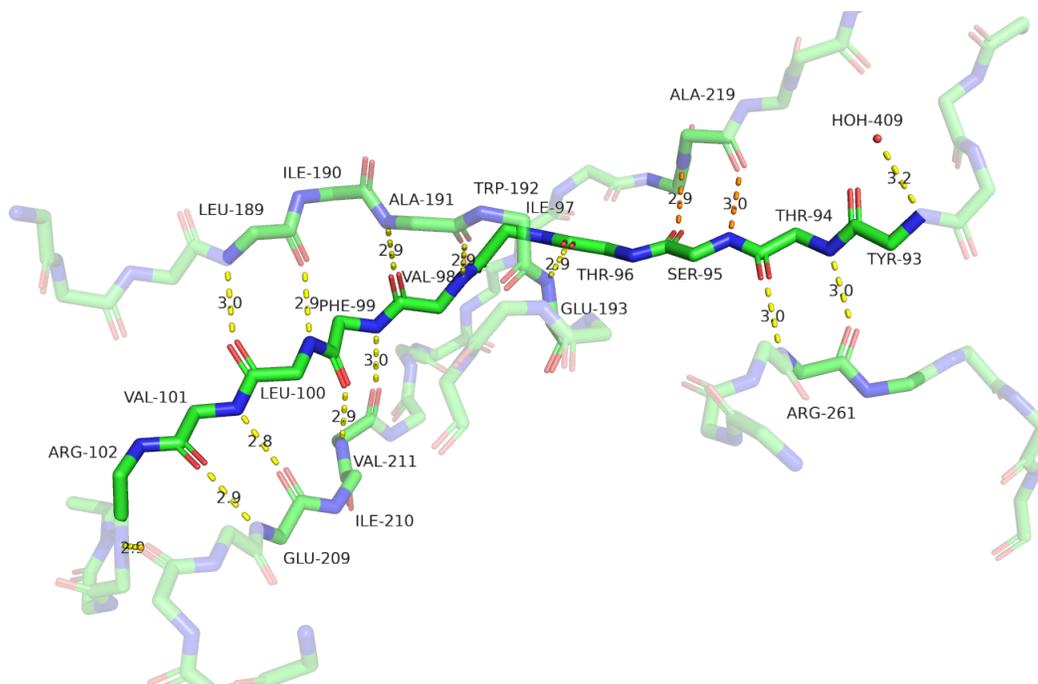


Рис. 12: Окружение бета-тяжа, представленного остатками 93-102. Желтым цветом показаны водородные связи, характерные для антипараллельного бета-листа; оранжевым – β -мостиковые связи Ser-95 : Ala-219. Заметим, что для остатка Туг-93 нам так и не удалось найти водородной связи с другими атомами белка. При поиске ближайших (4 Å) атомов кислорода/азота к основным кислороду/азоту Туг-93 нам удалось найти только кислород молекулы воды. Для остовных атомов Arg-102 нам не удалось найти водородных связей.

Картина оказалась достаточно хитрой: в общей сложности в образовании β -листа, не считая нашу β -нить, оказались задействованы 4 β -нити: остатки 189-193, 209-211, 219, 261. Предсказания для этих участков структуры представлены на рис. 13, 14.



Рис. 13: Разметка вторичной структуры для участков 189-193, 209-211, 219, которые, как мы предполагаем, образуют вместе с нашим β -тяжем бета-лист.



Рис. 14: Разметка вторичной структуры для 261 остатка, который, как мы предполагаем, образуют вместе с нашим β -тяжем бета-лист.

Теперь, наконец, давайте обсудим различие в предсказаниях DSSP и STRIDE для тяжа 93-102 (рис. 11). STRIDE предсказывает здесь один целый тяж из 10 остатков. DSSP размечает остатки Ser-95 как β -мостик (B-bridge), Thr-96 – как не имеющий вторичной структуры участок; остальные остатки – так же, как и STRIDE. Остаток Ser-95 образует две водородные связи своим остовом с остовом остатка Ala-219. Если посмотреть на предсказания для остатка Ala-219 (см. рис. 13), то мы видим, что DSSP последовательно предсказывает остаток Ala-219 тоже как β -мостик, а STRIDE, соответственно, как β -лист. В данном случае, на наш взгляд прав оказался DSSP, а STRIDE, как и раньше, сильно загрубил реальность. Действительно, на остатке Ala-219 заканчивается небольшой бета-лист (215-219) (см. рис 12 и 13), β -нити 93-102 и 215-219 не параллельны друг другу. Таким образом, Ser-95 подходит под определение β -мостика.

В случае Thr-96 ситуация чуть-чуть сложнее. Этот остаток образует водородную связь с Glu-193 кислородом своего остова (см. рис 12). Эта связь напоминает связь между антипараллельными бета-нитеями, однако Glu-193

согласно обеим разметкам входит в состав α -спирали (см. рис 13) и, соответственно, формально остаток Thr-93 не похож на остаток β -листа. Т.е. нам кажется, что здесь снова оказался прав DSSP.

Хочется подвести итог: DSSP, видимо, любит всякие нетипичные вторичные структуры, но зато, как мы видим, часто оказывается прав, когда их предсказывает. STRIDE, судя по нашей структуре, любит загроублять описание: предсказывать длинную α -спираль / β -лист. Таким образом, STRIDE, вероятно, больше подходит для какого-то крупного статистического анализа, а в деталях на него полагаться не следует.

3 Задание 2. Работа с разметкой вторичной структуры в автоматическом режиме

Здесь мы воспользовались DSSP алгоритмом, чтобы разметить структуры 30 PCA-моделей из PDB. Мы поделили элементы вторичной структуры на α -спирали (H), β -листы (E) и все остальное (C) (см. рис 15).

aa	V	L	D	T	F	G	P	R	E	N	W	Y	I	C	S	K	A	M	H	Q
H	113	235	101	74	76	62	23	92	152	70	24	55	104	19	85	111	216	46	37	84
E	140	94	35	70	54	50	12	40	37	15	15	45	103	20	57	36	59	19	22	24
C	106	138	158	140	66	269	174	95	135	130	27	64	71	40	183	116	137	30	50	69

Рис. 15: Разметка аминокислот 30 PCA-моделей.

Далее мы посчитали для нашего датасета склонности аминокислот присутствовать в размеченных DSSP элементах вторичной структуры. Мы определяли эту склонность (P_{ij}) так:

$P_{ij} = \frac{n_{ij} \cdot N}{n_i \cdot N_j}$, где n_{ij} – это количество остатков i в датасете, образующих тип вторичной структуры j ; n_i – это общее количество остатков i в датасете; N_j – это общее количество остатков, образующих тип вторичной структуры j во всем датасете; N – это общее количество остатков в датасете.

Для каждого типа вторичной структуры мы отсортировали аминокислоты по значениям P_{ij} (см. рис 16).

Amino acid propensities for secondary structures

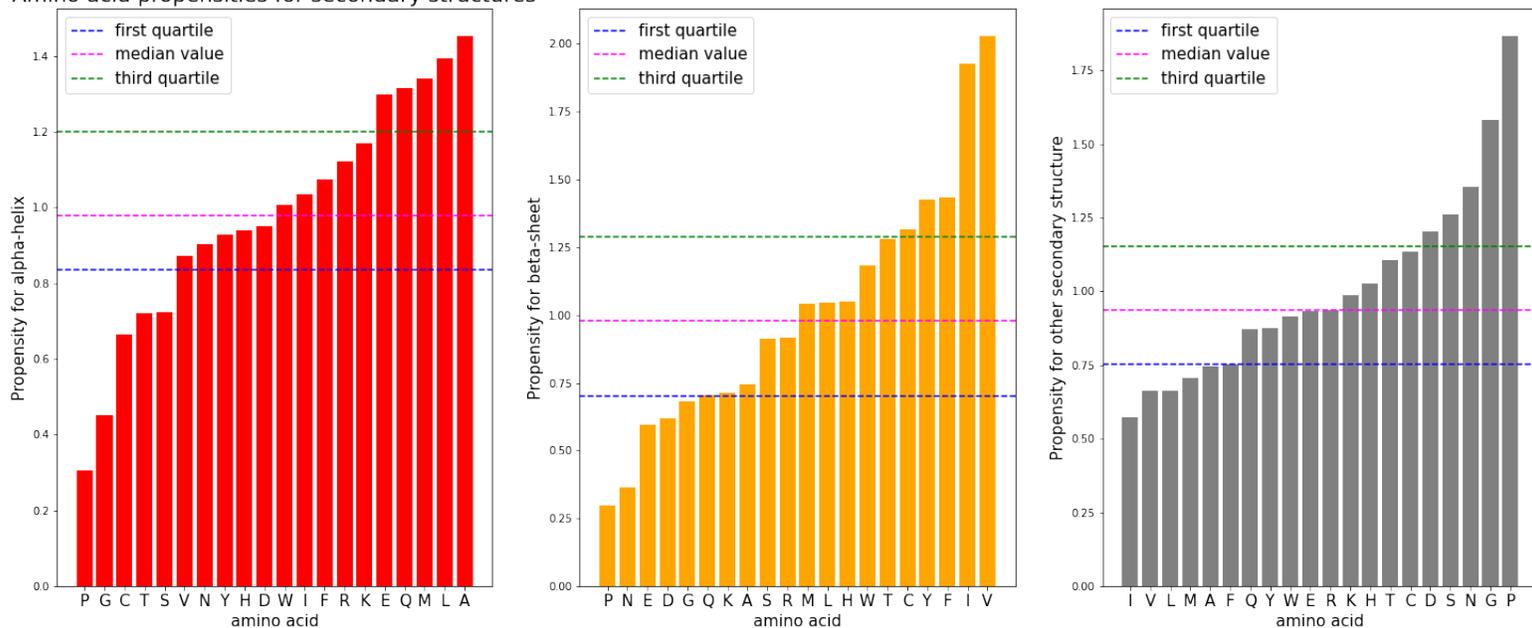


Рис. 16: Склонности образования аминокислотами α -спиралей, β -листов и других типов вторичной структуры (в первом приближении – петель).

Мы можем заключить, что для нашего датасета выполняются закономерности: для α -спиралей наиболее характерны остатки аланина, лейцина, метионина, глутамина, глутамата (эти остатки часто встречаются в α -спирали и достаточно редко в других структурах), для β -листов характерны остатки валина, изолейцина, фенилаланина, тирозина; наконец, для петель (в нашем приближении) характерны остатки пролина, глицина, аспарагина.

4 Сессии Pymol + jupyter-notebook

<https://kodomofbb.msu.ru/~osyafinkelberg/term7/prac8/session0.pse>
<https://kodomofbb.msu.ru/~osyafinkelberg/term7/prac8/session1.pse>
https://kodomofbb.msu.ru/~osyafinkelberg/term7/prac8/analyze_propensity.ipynb

5 ССЫЛКИ

[1] He and Quioco, 'Dominant role of local dipoles in stabilizing uncompensated charges on a sulfate sequestered in a periplasmic active transport protein' 1993 Protein Science

[2] Langridge et al., 'Sulfate-Binding Protein from Salmonella typhimurium: Physical Properties' 1970 Science

[3] Kabsch, Sander, 'Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features', 1983 Biopolymers (DSSP algorithm)

[4] Frishman, Argos, 'Knowledge-based protein secondary structure assignment' 1995, Proteins (STRIDE algorithm)