

Структурная биоинформатика

Практикум 8. Вторичная структура.

Поддьяков Иван

Задание 1. Разметка вторичной структуры в ручном режиме	2
Пример 1	3
Пример 2	4
Пример 3	6
Задание 2. Разметка вторичной структуры в автоматическом режиме	10
Источники	10

Задание 1. Разметка вторичной структуры в ручном режиме

В данном задании я буду работать со структурой 1JZ7 [1]. Это структура фермента β -галактозидазы E.coli со связанным лигандом - β -D-галактопиранозой, полученная в рамках исследования по установлению каталитического механизма [2]. Ассиметрическая единица представляет собой гетеротетрамер. Я буду изучать разметку вторичной структуры на примере цепи А с помощью сервиса [2Struc](#). Разметка всей цепи очень длинная, поэтому я поместил ее на [страницу практикума](#). На Рисунке 1 представлена часть всей размеченной последовательности, красными прямоугольниками отмечены участки, которые я буду рассматривать.

Я сравню аннотацию вторичной структуры цепи, построенную тремя алгоритмами: DSSP [3], STRIDE [4], P-SEA [5]. DSSP сначала находит водородные связи между кислородами и азотами остова с достаточной энергией, которые размечаются как повороты или мостики. Затем идущие подряд повороты объединяются в спирали, а мостики - в лестницы и листы.

В основу STRIDE также положено определение энергии водородных связей остова, но данный алгоритм учитывает аннотации набора структур, для которого разметка проведена вручную. В соответствие с этими аннотациями в алгоритме подобраны веса для отнесения остатков к той или иной вторичной структуре и рассчитаны вероятности, что остаток с некоторыми значениями торсионных углов ϕ и ψ может принадлежать тому или иному типу вторичной структуры.

Алгоритм P-SEA, в отличие от двух описанных выше, может относить остатки только к спиралям или тяжам. Используемый подход в данном случае несколько проще и главным образом основывается на взаимном расположении СА атомов. Для i -го остатка вычисляются расстояния $(i-1, i+1)$, $(i-1, i+2)$, $(i-1, i+3)$ и углы $(i-1, i, i+1)$, $(i-1, i, i+1, i+2)$ на вершинах в СА. Затем остатки, для которых расстояния или углы удовлетворяют установленным интервалам значений, аннотируются как спирали или тяжи, а остальные относятся к петлям.

RESNUM	13	23	33	43	53	63	73	594	604	614															
SEQ	RRD	WENPGVTQLN	RLAAHP	PFASWRN	SEEARTRDRPS	SQLRSLNGE	WRFAPPAEAVF	ESWLECDLP	AYGGDFG	DTPNDRQFC	MGLVFADRTPHPALTE														
CONSENSUS	00GGGO	TTTOXEX	00000X0000	00000XX	HHHHHH0000	TTTT00	EEEEEEEE00	GGG0CH	HHH000T	EXTTT	X00X0CXXXX000XX0TTX0B0HHHHH														
DSSP	00GGGO	TTTOXEX	000000	00000B	SHHHHHH	T000	TTTT00	EEEEEEEE	SSGGG	0CT	HHH000T	E0TTTT	0000	00GGG	0000	BO	TTT	0B0	HHHHH						
STRIDE	CCGGG	TTTTTE	EE	BCCCC	CCCCB	T	HHHHH	CCC	TTTT	EEEE	EEEEEEEE	CCGGG	CC	H	HHH	CC	T	E	TTTTTT	TTTT	TTTT	EEEE	TTTT	CB	HHHHH
PSEA	cccccccccccc	bbbb	cccccccc	aaaaaa	cccccccccccc	bbbbbbb	cccccccc	aaaa	cccc	bb	cccccccc	cccc	cccccccccccc	aaaa											

Рисунок 1. Фрагменты разметки вторичной структуры 1JZ7, цепь А.

Пример 1

Рассмотрим аннотацию вторичной структуры разными алгоритмами для аминокислот с 27 по 31. STRIDE не относит их ни к какому типу вторичной структуры, DSSP размечает только пролин-31 как излом (bend), а вот P-SEA распознает данный участок как β -тяж. На Рисунке 2 представлены измерения, на основании которых данный алгоритм относит, в частности, ALA-28 к β -тяжу. В допустимые значения, указанные авторами метода [5], укладываются все 5 измерений, кроме двугранного угла, построенного по СА аминокислот 27-30 (-111.3° при допустимых $-170 \pm 45^\circ$). Таким образом, данный остаток полностью подходит под критерий по расстояниям и определяется как тяж.

Два других метода изначально ищут водородные связи между атомами остова, образования которых на данном участке не происходит ни внутри него самого, ни с соседними участками. Поэтому они аннотируют данные аминокислоты как петли, а не тяжи. На Рисунке 3 показан угол между СА аланина-29 и фенилаланина-33 с вершиной в СА пролина-31. Последний определяется DSSP как излом, так как происходит изменение направления цепи более чем на 70° ($180 - 103.2 = 76.8^\circ$). В других рассмотренных алгоритмах такой тип вторичной структуры не распознается.

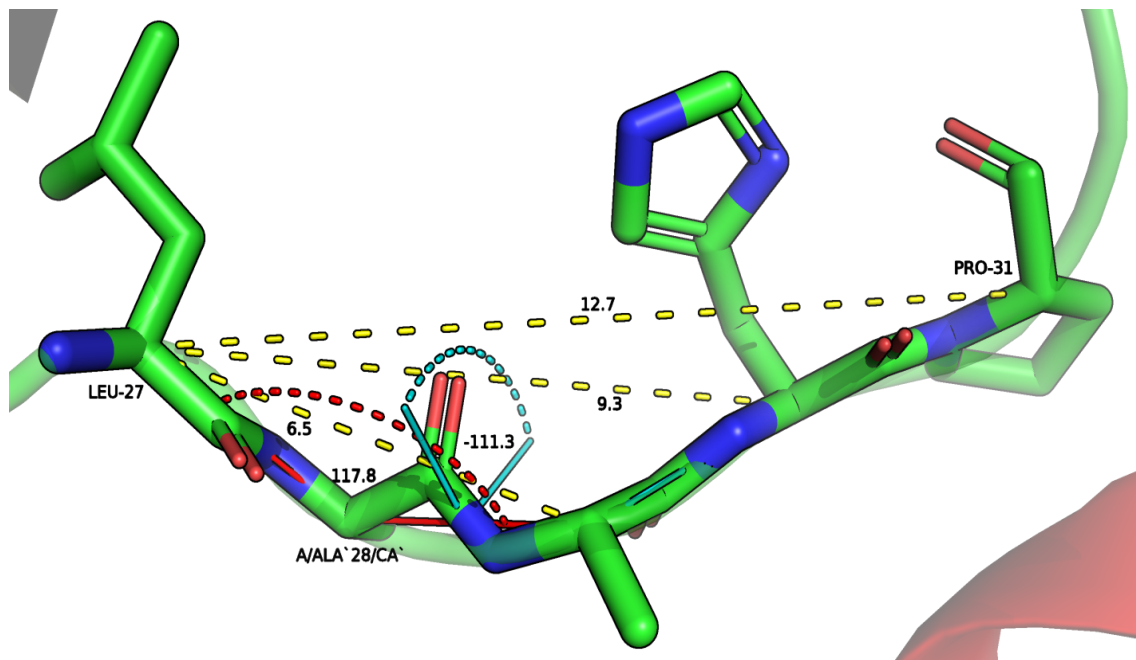


Рисунок 2. Расстояния и углы относительно СА атома аланина-28.

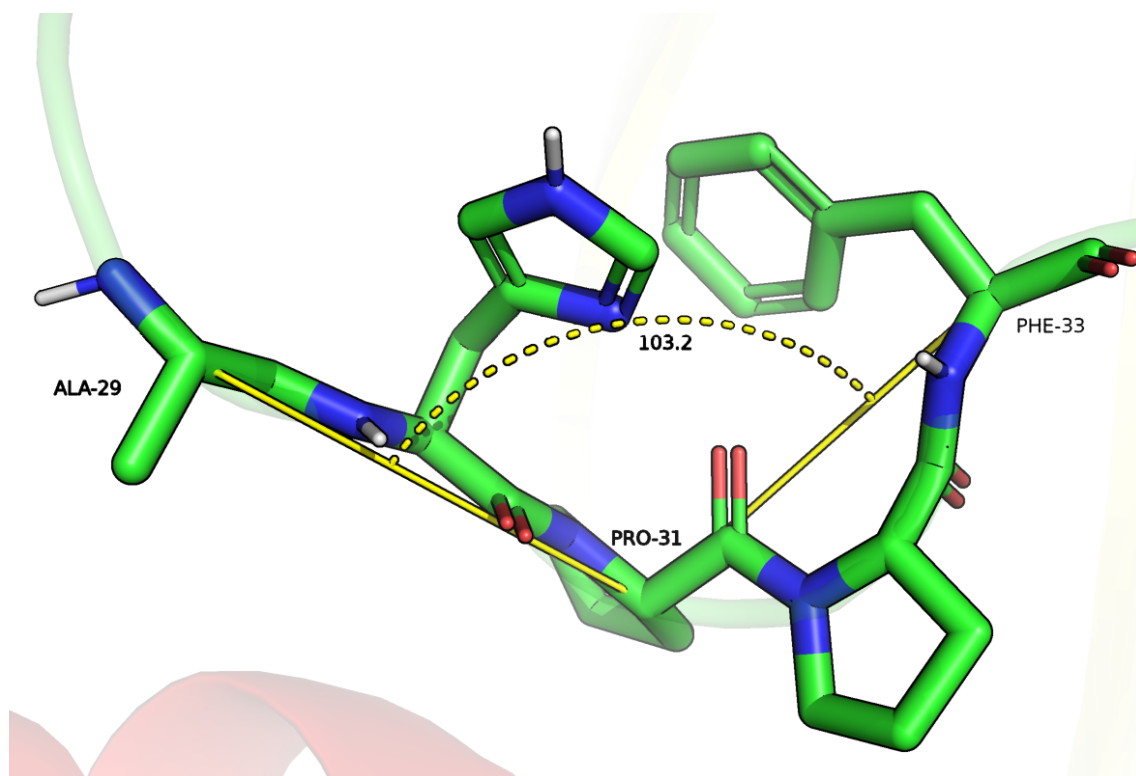


Рисунок 3. Угол ALA29/CA - PRO31/CA - PHE33/CA.

В данном случае я согласен с разметкой DSSP - нет никаких оснований аннотировать этот участок как тяж или спираль. В PDB файле этот участок тоже никак не аннотирован. На мой взгляд, отметить излом также важно, так как происходит сильное изменение направления цепи.

Пример 2

Рассмотрим тип вторичной структуры, который аннотирован для глутамата-71 тремя алгоритмами. STRIDE и P-SEA определяют его как α -спираль (спираль 4_{13} в случае STRIDE), а DSSP - как поворот. На Рисунке 4 представлен участок, содержащий данную аминокислоту, ее торсионные углы ϕ и ψ , возможные водородные связи кислорода остова с азотами групп $i + 3$ и $i + 4$. Водородная связь с азотом глутамата-75 кажется мне достаточно слабой из-за расстояния между донором и акцептором, а связь с азотом лейцина-74 - затрудненной геометрически. Если рассчитать энергию этих связей, как в алгоритме DSSP, то у первой такое значение получится значительно меньше второй (-0.71 ккал/моль против 1.61 ккал/моль), но все равно выше порога -0.5 ккал/моль. Мне кажется, в этом случае алгоритм принимает во внимание связь с высокой энергией и считает эту аминокислоту не последней в α -спирали, а 3-поворотом.

STRIDE определяет эту аминокислоту как часть 4_{13} -спирали, так как ее торсионные углы лежат в допустимой области значений, согласно данным из вручную аннотированных структур, а эмпирические веса и энергия водородной связи позволяют преодолеть порог значимости для такой аннотации.

На Рисунке показаны расстояния и углы с вершиной в СА глутамата-71, на которые опирается алгоритм P-SEA. Все значения удовлетворяют критериям алгоритма, кроме длины связи между $i - 1$ и $i + 3$ СА (8.0 \AA при допустимых $6.4 \pm 0.6 \text{ \AA}$). Так как достаточно соответствия только критериям по углу или расстоянию, данный остаток включается в α -спираль.

Глутамат-71 аннотирован в PDB-файле также как спираль. Я склонен считать, что водородная связь O GLU-71 с N LEU-74 маловероятна из-за геометрических затруднений, и отнес бы данный остаток к 4_{13} -спирали, в пользу чего также говорят значения торсионных углов, часто встречающиеся у типа правой спирали.

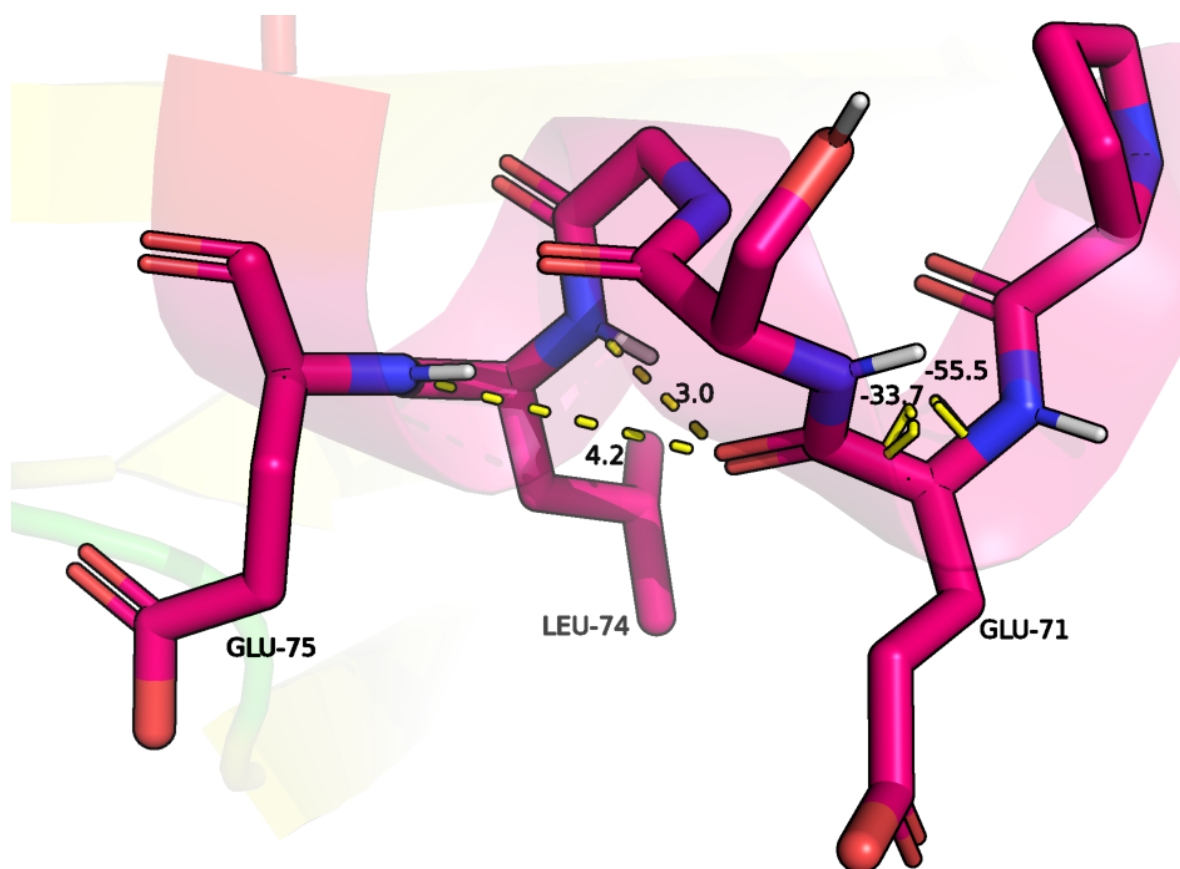


Рисунок 4. Глутамат-71. Водородные связи и торсионные углы .

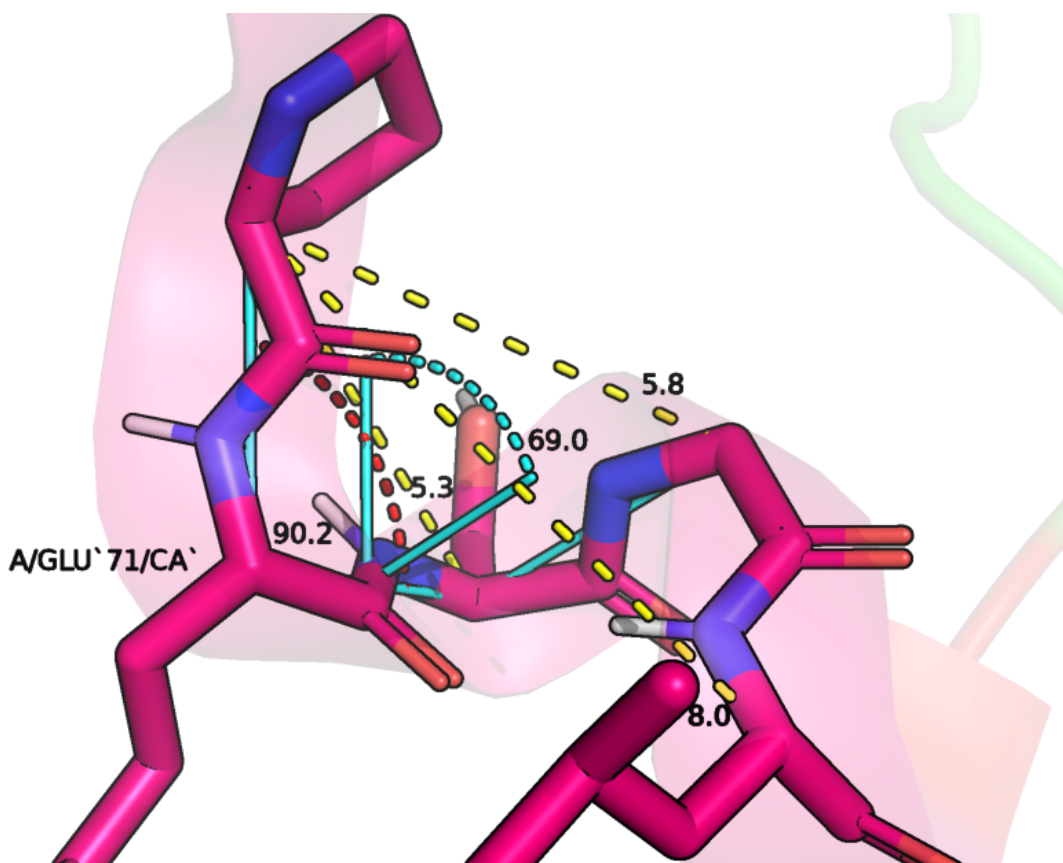


Рисунок 5. Глутамат-71. Дистанции и углы в алгоритме P-SEA.

Пример 3

Для следующего примера я выбрал участок с 599 по 602 остаток, который аннотируется как 3_{10} -спираль алгоритмом DSSP, как повороты - STRIDE и как неструктурированный - P-SEA. В PDB-файле участок также аннотирован как спираль.

На Рисунке 6 показан данный участок и водородные связи между кислородами и азотами остова. Шесть остатков с 598 по 603 формируют 3 идущих подряд 3-поворота, согласно изначальной разметке DSSP. Энергии данных связей, рассчитанные по аналогии с алгоритмом, составляют -1.95, -0.70, -2.70 ккал/моль для пар 598-601, 599-602, 600-603 соответственно, все значения выше порогового. Это позволяет данному алгоритму объединить этот участок в 3_{10} -спираль, аннотировав так 4 центральных остатка.

Для идентификации поворотов алгоритм STRIDE использует классификацию по торсионным углам ϕ и ψ , приведенную в соответствующем исследовании [6]. На Рисунке 7 приведены значения торсионных углов для рассматриваемого участка. Их значения соответствуют типичным значениями поворотов типа I (α) для остатков $i + 1$

и $i + 2$ относительно остатка, кислород которого акцептирует водородную связь. Алгоритм мог бы разметить данный участок как 3_{10} -спираль, так как есть минимум 3 идущих подряд 3-поворота, однако он так не делает. Ввиду редкости 3_{10} -спиралей, в STRIDE для их аннотации используется алгоритм, как в DSSP, но определение водородных связей происходит по-другому [7], так как для стандартного “knowledge-based” подхода недостаточно данных. В данном случае накладываются ограничения на углы с вершинами в атомах донора и акцептора (2 ограничения) и двугранные углы между плоскостями донора или акцептора и плоскостью, в которой лежит связь и атом C, ковалентно связанный с O или N (2 ограничения). Также длина водородной связи должна быть меньше суммы радиусов образования водородной связи у атомов в соответствующей конфигурации.

На Рисунке 8 представлены три из четырех измеряемых углов водородной связи 600/O и 602/N. Величины углов с вершинами в атомах азота и кислорода подходят под установленные критерии, но, так как атом O и N находятся в sp_2 -гибридизации, необходимо проверить ограничения на планарность. Двугранный угол CA-C-O-N составляет -108.3° , что не попадает в допустимый интервал $0 \pm 90^\circ$. При этом условие на длину связи выполняется - $2.9 \text{ \AA} < 1.6 \text{ \AA} (sp_2 \text{ O}) + 1.9 \text{ \AA} (sp_2 \text{ N})$. Алгоритм не считает такую связь достоверной и не включает остаток в 3_{10} -спираль, но все равно обозначает его как поворот с водородной связью по другим критериям. Такая аннотация кажется мне противоречивой.

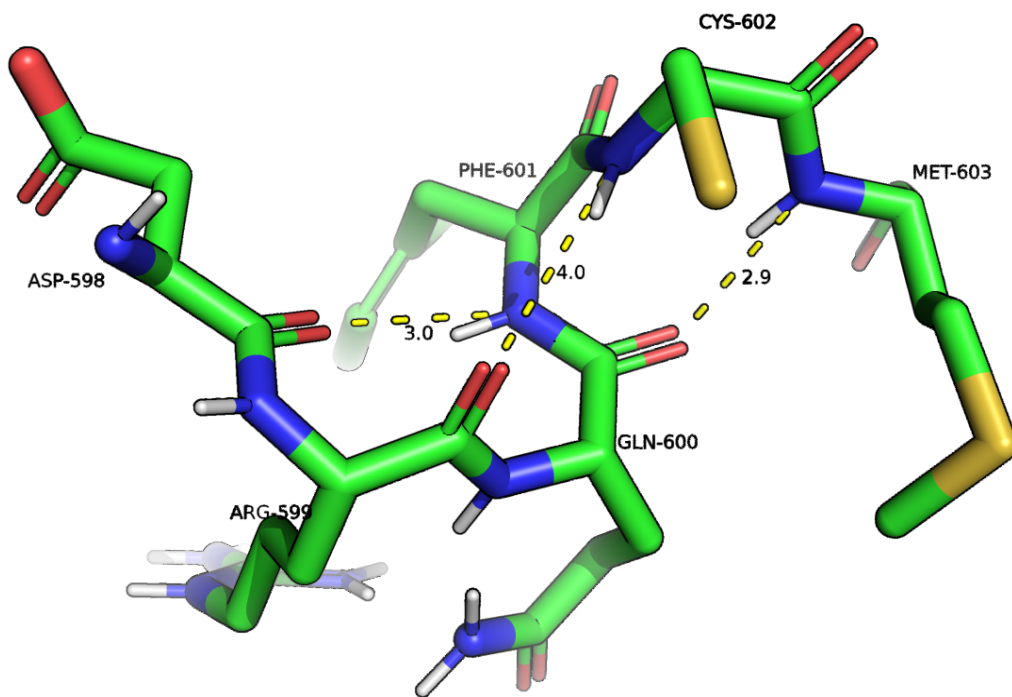


Рисунок 6. Водородные связи остова на участке 598-602.

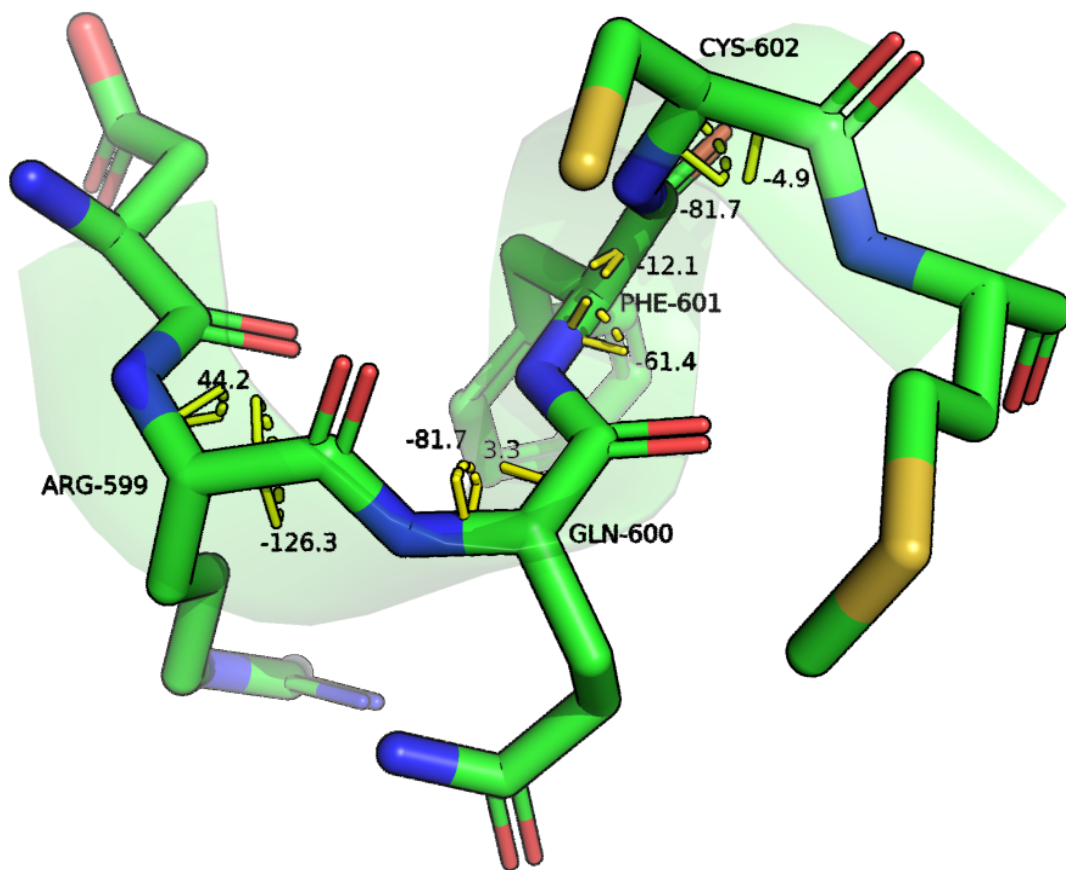


Рисунок 7. Торсионные углы ϕ и ψ на участке 598-602.

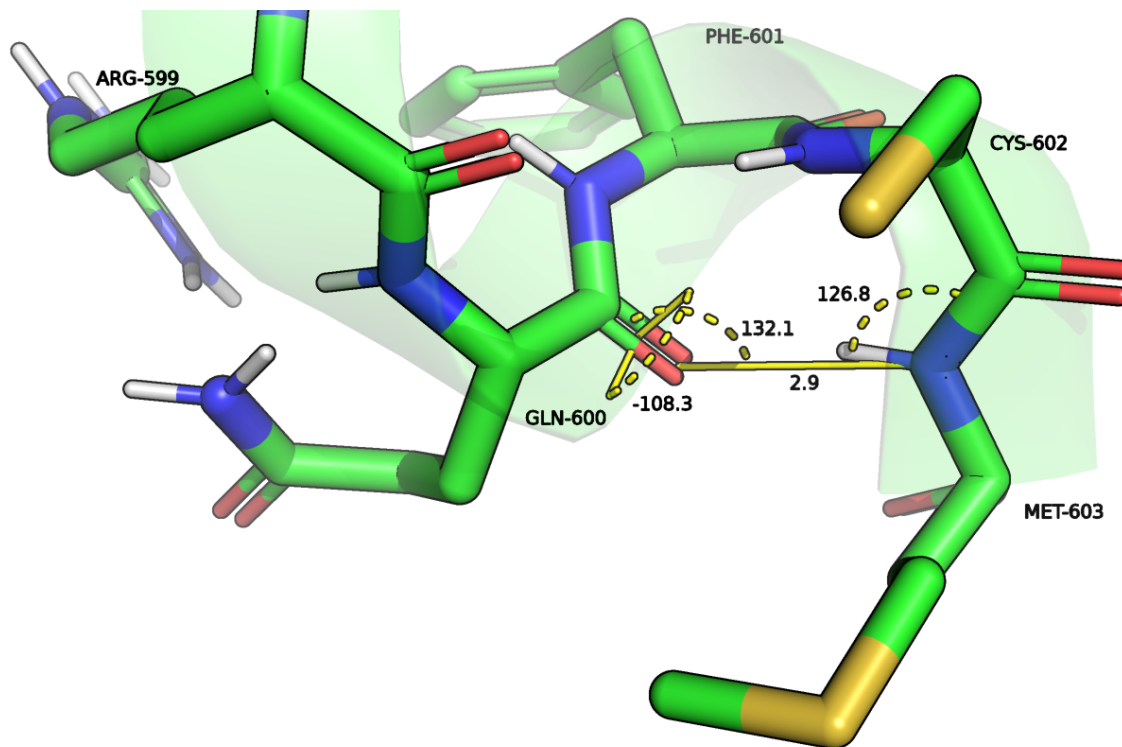


Рисунок 8. Углы вокруг возможной водородной связи 600/O и 602/N.

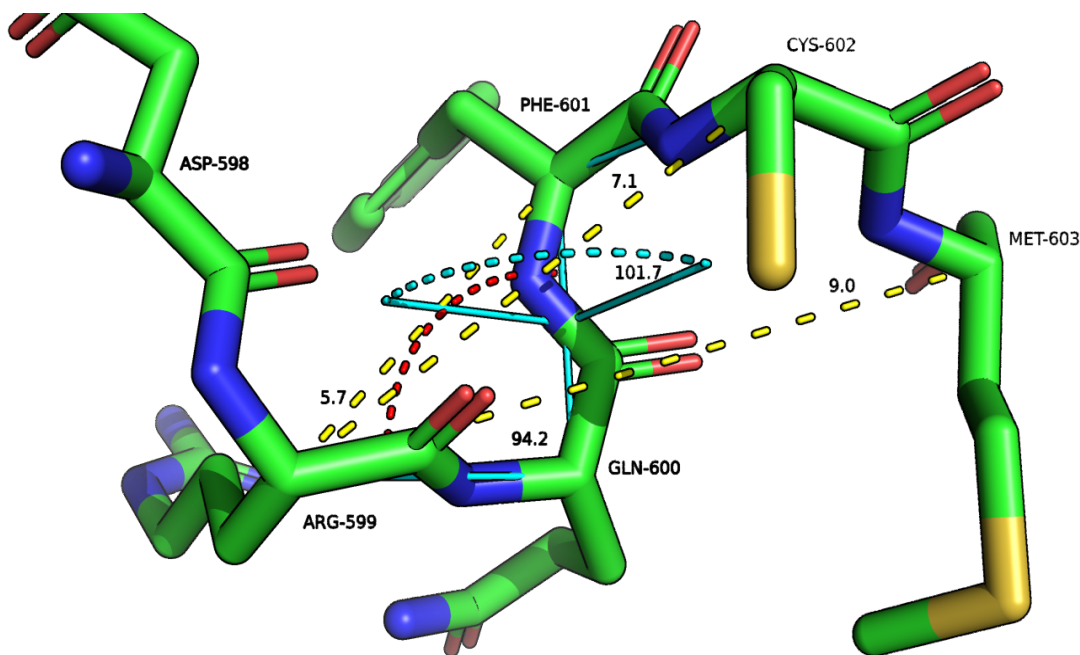


Рисунок 9. Глутамин-600. Дистанции и углы в алгоритме P-SEA.

Алгоритм P-SEA аннотирует изучаемый участок как неструктурированный. Рассмотрим необходимые для этого подхода параметры на примере глутамина-600 (Рисунок 9). Длины связей и значения углов не соответствуют ни одному допустимому интервалу значений как для листов, так и для спиралей, и алгоритм не относит этот участок ни к одному из этих типов.

На данном участке 3 алгоритма предсказали разные типы вторичной структуры. В данном случае я склонен согласиться с аннотацией DSSP. Водородные связи с данной в спирали с достаточно большой энергией, а торсионные углы приближены по значениям к спиральям 3_{10} . Два других алгоритма, ввиду своих особенностей, не могут определить это относительно редкий тип спирали.

Рассмотрев, как данные алгоритмы определяют спирали и тяжи на всей цепи в данной структуре, я не нашел случая, когда для одного участка один алгоритм аннотирует тяж, а другой - спираль. При этом P-SEA, плохо учитывающий взаимодействия между атомами остова и рассматривающий только CA, определяет тяжи там, где другие алгоритмы их не находят - Пример 1. В целом я бы отдал предпочтение DSSP, хотя частные случаи лучше рассматривать вручную.

Задание 2. Разметка вторичной структуры в автоматическом режиме

В данном задании я осуществил аннотацию вторичных структур в наборе из 30 моделей с помощью алгоритма DSSP. Затем по полученным разметкам я рассчитал склонность каждого типа аминокислоты образовывать тот или иной тип вторичной структуры по формуле $P_{ik} = (n_{ik}/n_i) / (N_k/N)$, где P_{ik} - propensity аминокислотного остатка i образовывать тип вторичной структуры j , n_{ik} - количество остатков i в датасете, образующих тип вторичной структуры j , n_i - общее количество остатков i в датасете, N_k - общее количество остатков, образующих тип вторичной структуры j во всем датасете, N - это общее количество остатков в датасете. Классификация вторичной структуры упрощена до β -листов (E), α -спиралей (H), петель (C). Jupyter-notebook с расчетами приложен на странице практикума. Полученные результаты представлены в Таблице 1.

Resn	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y
SS																				
C	0.77	1.28	1.51	0.85	0.78	1.56	1.17	0.64	1.03	0.67	0.74	1.41	1.71	0.77	0.88	1.12	1.06	0.69	0.83	0.80
E	0.81	1.47	0.38	0.68	1.41	0.59	0.69	1.80	0.75	1.05	1.16	0.40	0.46	1.02	1.11	0.88	1.04	1.97	1.65	1.83
H	1.47	0.26	0.69	1.46	1.03	0.46	0.96	0.96	1.13	1.45	1.27	0.82	0.34	1.32	1.10	0.91	0.89	0.78	0.79	0.71

Таблица 1. Склонность остатков образовывать различные типы вторичной структуры.

Петли больше всего склонны образовывать такие остатки, как гибкий глицин и пролин, торсионные углы которого сильно ограничены и могут нарушать регулярность листов и спиралей. Заметно чаще чем петли и спирали, валин и изолейцин образуют β -листы. Возможно, листы, утопленные внутрь глобулы, так формируют гидрофобные ядра с соседними листами и спиральями. Небольшой гидрофобный аланин чаще образует α -спирали.

Источники

1. Bank RPD. 1JZ7. [cited 1 Dec 2021]. Available: <https://www.rcsb.org/structure/1jz7>
2. Juers DH, Heightman TD, Vasella A, McCarter JD, Mackenzie L, Withers SG, et al. A structural view of the action of Escherichia coli (lacZ) beta-galactosidase. *Biochemistry*. 2001;40: 14781–14794.
3. Kabsch W, Sander C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*. 1983;22: 2577–2637.

4. Frishman D, Argos P. Knowledge-based protein secondary structure assignment. *Proteins*. 1995;23: 566–579.
5. Labesse G, Colloc'h N, Pothier J, Mornon JP. P-SEA: a new efficient assignment of secondary structure from C alpha trace of proteins. *Comput Appl Biosci*. 1997;13: 291–295.
6. Wilmot CM, Thornton JM. Beta-turns and their distortions: a proposed new nomenclature. *Protein Eng*. 1990;3: 479–493.
7. Sticke DF, Presta LG, Dill KA, Rose GD. Hydrogen bonding in globular proteins. *J Mol Biol*. 1992;226: 1143–1159.