

Вторичная структура

Введение

Я взяла PDB-структуру из практикума 6 - [1YWR](#). Это модель структуры комплекса домена MAP киназы P38 с моноциклическим пиразолоном, выступающим в роли ингибитора. MAPK14 (Mitogen-activated protein kinase 14) – одна из четырех MAP киназ P38. Она играет роль в клеточном ответе, вызванном внеклеточным стимулом (провоспалительными цитокинами или клеточным стрессом) и ведущим к прямой активации транскрипционных факторов [1, 2].

Работа с разметкой вторичной структуры в ручном режиме

Для структуры [1YWR](#) была сгенерирована аннотация с помощью [2Struc](#). Я рассмотрела генерации, выданные алгоритмами DSSP и STRIDE.

Алгоритм DSSP рассчитывает энергии водородных связей для пар остовных атомов азота и кислорода и размечает водородные связи с достаточной энергией на повороты и мостики. Затем находятся паттерны поворотов, которые объединяются в спирали, и паттерны мостиков, которые объединяются в лестницы и листы.

Алгоритм STRIDE тоже вначале производит расчет энергий водородных связей для пар атомов N и O остова (но отличным от DSSP способом). STRIDE отличается от DSSP еще тем, что он основан на статистике экспертных аннотаций. На основе этих аннотаций в алгоритме распределены веса для отнесения остатков к определенным элементам вторичной структуры.

2Struc выдал последовательность белка, в которой размечено, какому элементу вторичной структуры принадлежит тот или иной аминокислотный остаток. Картинка с полной последовательностью размещена по [ссылке](#).

В некоторых позициях оценка принадлежности остатка к тому или иному элементу вторичной структуры различалась у разных алгоритмов. Я рассмотрела несколько примеров таких различий.

Пример 1. Здесь я рассмотрела сразу два случая расхождений в выдаче двух алгоритмов. Остатки в позициях 33-35 алгоритм DSSP выдал как три идущих подряд излома. STRIDE же совсем не выдал изломов (он это не делает в принципе), но оценил остатки как β -лист и два идущих за ним поворота (см. Рис. 1).

| | | | | |
|-----------|-------------------------------|-------------|-----|------------|
| RESNUM | 5 | 25 | 35 | |
| SEQ | TIWEVPERYQNLSPVGSGAYGSVCAAFDT | | | |
| CONSENSUS | i EEEEEET | T EEEEEEEEE | XXX | EEEEEEEEET |
| DSSP | i EEEEEET | T EEEEEEEEE | SSS | EEEEEEEEET |
| STRIDE | i EEEEEET | T EEEEEEEEE | TT | EEEEEEEEET |

Рис. 1. Несовпадение разметки вторичной структуры алгоритмами DSSP и STRIDE в позициях 33-35

Рассмотрим остаток Gly-33. Как показано на Рис. 2, между остовным атомом азота Gly-33 и остовным атомом кислорода Gly-36 есть водородная связь (пусть и слабаватая). К тому

же, торсионные углы ϕ (169.1) и ψ (-133.9) соответствуют β -листу. Таким образом, можно «поверить» алгоритму STRIDE и отнести 33 остаток к β -листу. Скорее всего, алгоритм DSSP просто не присвоил метку « β -лист» глицину 33 как терминальному остатку вторичной структуры.

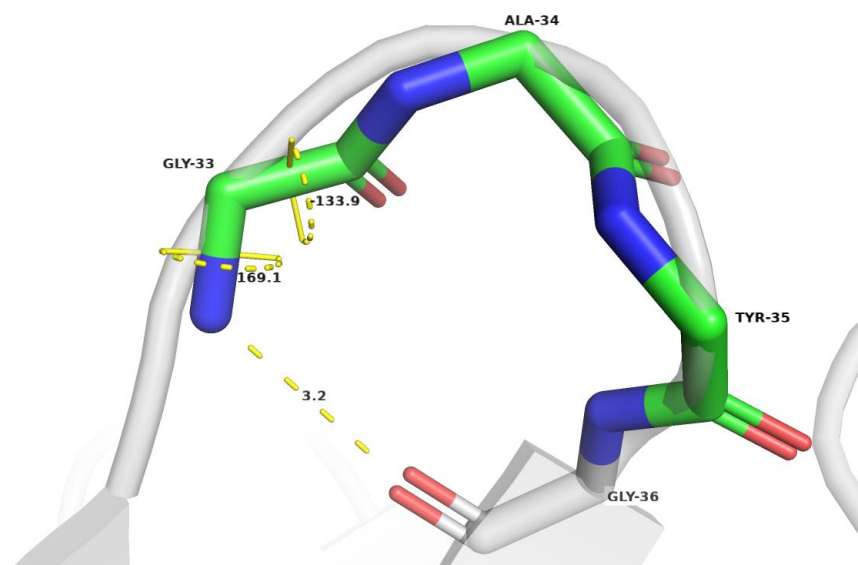


Рис. 2. Глицин 33, отнесенный алгоритмами DSSP и STRIDE к излому и β -листу, соответственно. Остатки Ala-34 и Tyr-35 STRIDE посчитал за повороты, а DSSP – за изломы. Я больше склоняюсь ко второму варианту, так как остов изменяет направление более, чем на 70° (см. Рис. 3)

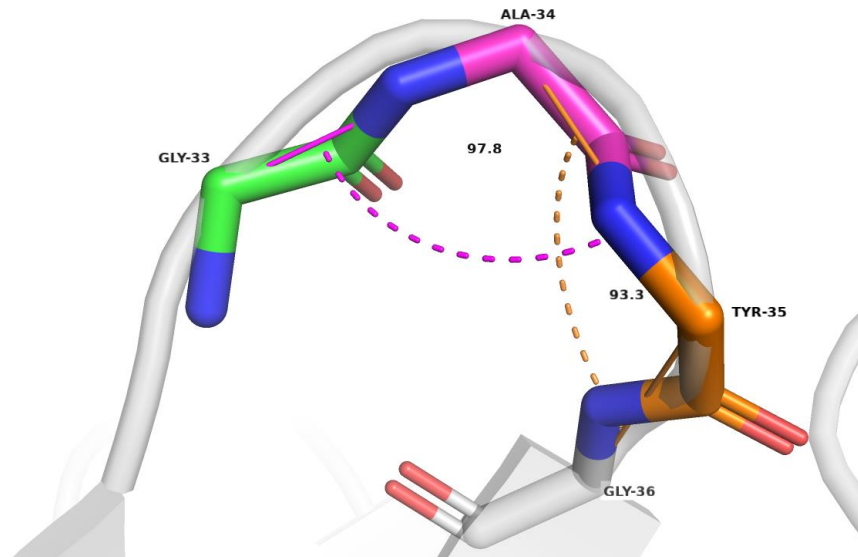


Рис. 3. Аланин 34 и тирозин 35, отнесенные алгоритмами DSSP и STRIDE к изломам и поворотам, соответственно

Пример 2. Остаток Ala-144 алгоритм DSSP расценил как поворот, однако STRIDE выдал остаток за α -спираль (см. Рис. 4).

| | | | | |
|-----------|---|-------------------------------|----------|-------------|
| RESNUM | | 136 | 146 | 156 |
| SEQ | | IYQILRGLKYIHSADIIHRDLKPSNLAVN | | |
| CONSENSUS | i | HHHHHHHHHHHHHH | XX00XXXX | GGGEEEX |
| DSSP | i | HHHHHHHHHHHH | TT000SS | OOGGEEEO |
| STRIDE | i | HHHHHHHHHHHH | CCC | TTTCGGGEEET |

Рис. 4. Несовпадение в оценке принадлежности к элементу структуры остатка Ala-144

Скорее всего, DSSP не оценил остаток как принадлежащий к α -спирали, потому что этот алгоритм не присваивает терминальным остаткам метку вторичной структуры. Как видно на Рис. 5, остаток Ala-144 связан с Tyr-140 водородной связью, слабой, но соответствующей паттерну α -спирали. К тому же, торсионные углы остатка аланина ϕ (-98.4) и ψ (11.9) соответствуют α -спирали. Так что в этот раз алгоритм STRIDE выдал более правильный вариант разметки.

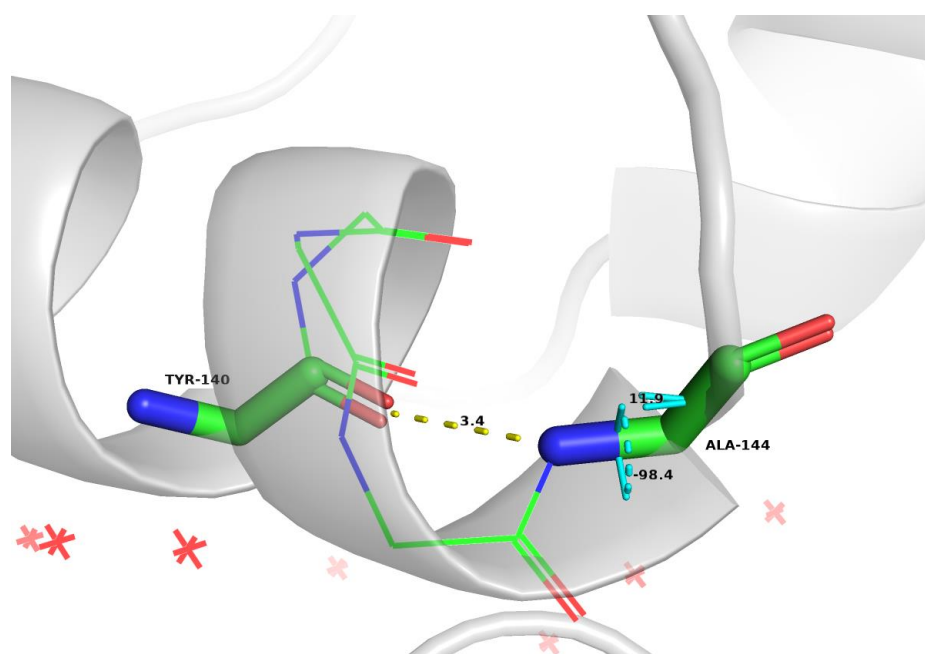


Рис. 5. Аланин-144, помеченный алгоритмами DSSP и STRIDE как поворот и α -спираль, соответственно

Пример 3. Остатки в позициях 314 – 316 были размечены алгоритмом DSSP как 3_{10} -спираль, а алгоритмом STRIDE – как повороты (см. Рис. 6).

| | | | |
|-----------|---|---------------------------|--------------------|
| RESNUM | | 305 | 315 |
| SEQ | | AAQALAHAYFAQYHDPDDEPVADPY | |
| CONSENSUS | i | HHHHXXX | GGGTTTXXXXXX000000 |
| DSSP | i | HHHHTTS | GGGTTT00GGG0000000 |
| STRIDE | i | HHHHHCC | GGGTTTTTTTTTCCCCC |

Рис. 6. Несовпадение в разметке вторичной структуры в позициях 314-316

На участке 314-315 есть паттерн водородных связей между остовными атомами азота и кислорода, однако одна из этих связей показалась мне уж очень слабой (длина этой связи 4.2 Å). Я рассмотрела связи между боковыми атомами остатков, возможно, они бы «поддержали» 3_{10} -спираль (если она есть). Действительно, между атомом кислорода Asp-313 боковой цепи и остовным атомом азота Asp-315 имеется водородная связь, а между

атомом азота His-312 и ионогенным радикалом Glu-317 боковой цепи имеется соляной мостик (см. Рис.7).

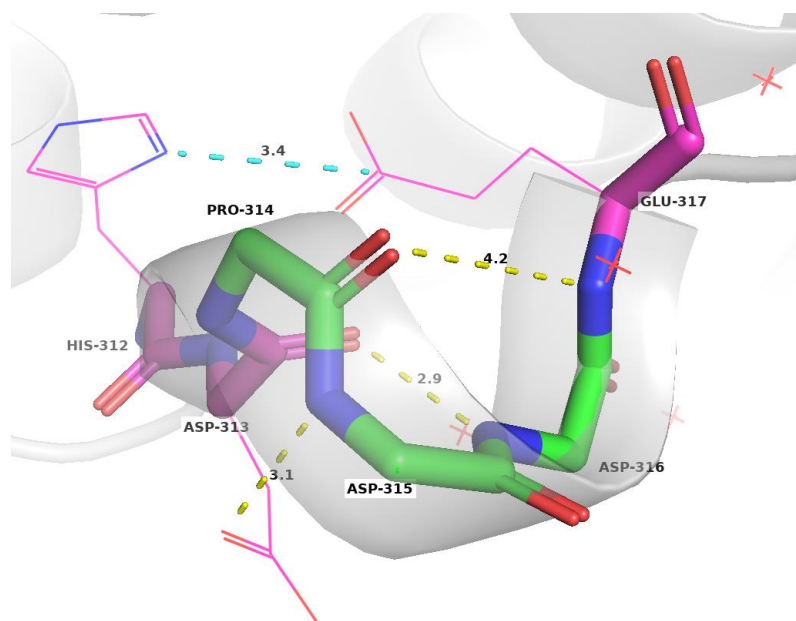


Рис. 7. Участок 314-316, помеченный алгоритмами DSSP и STRIDE как 3_{10} -спираль и повороты, соответственно. Водородные связи обозначены желтым пунктиром, соляной мостик – голубым.

Рассмотрела также торсионные углы ϕ и ψ на участке 314-316. Они соответствуют 3_{10} -спирали для Pro-314 и Asp-315, но ϕ и ψ напоминают углы, соответствующие 3_{10} -спирали лишь отдаленно.

Таким образом, остатки Pro-314 и Asp-315 соответствуют участкам 3_{10} -спирали, однако остаток Asp-316 соответствует участку 3_{10} -спирали с большой натяжкой. STRIDE, возможно, не принял Asp-316 за 3-поворот из-за чего длина 3_{10} -спирали оказалась недостаточной, и STRIDE пометил участок 314-316 как идущие подряд повороты. Выдаче какого алгоритма в данном случае верить, мне кажется, вопрос спорный, но все-таки считаю этот участок маленькой 3_{10} -спиралью, потому что данная вторичная структура поддерживается связями между атомами боковых радикалов.

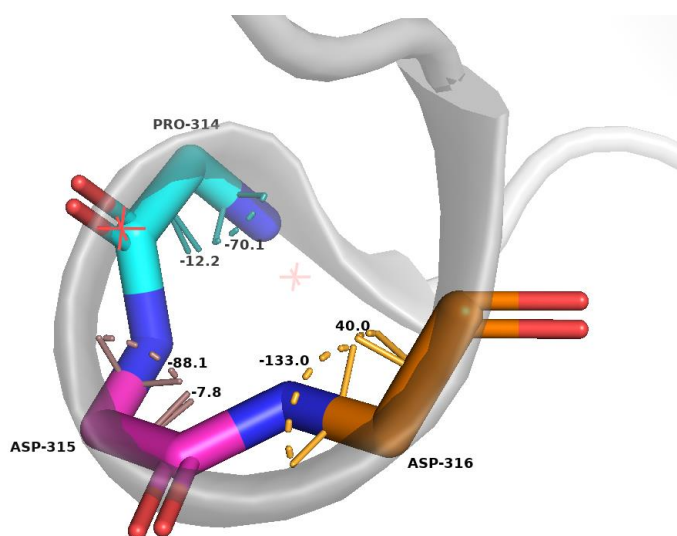


Рис. 8. Торсионные углы ϕ и ψ на участке 314-316.

Замечу, что, в целом, DSSP нашел больше «других» (не относящихся к α -спиралям и β -тяжам) вторичных структур* (см. Рис. 9). Думаю, так может получиться из-за того, что после расчета энергий водородных связей для пар остовных атомов N и O алгоритм STRIDE для каждого остатка по значениям торсионных углов ϕ и ψ рассчитывает вероятность принадлежности остатка α -спиралям и β -тяжу (то есть, сначала идет отбор на α -спираль и β -тяжи, а затем – на остальные структуры). Только потом идет детекция «зародышей» элементов вторичной структуры. Алгоритм DSSP же минует стадию отбора на α -спираль и β -тяжи и сразу ищет «зародышей» элементов вторичной структуры.

Summary of Whole Chain

| Method | H | E | O | X |
|-----------|-------|-------|-------|-------|
| CONSENSUS | 0.344 | 0.147 | 0.317 | 0.192 |
| DSSP | 0.356 | 0.147 | 0.497 | 0.000 |
| STRIDE | 0.375 | 0.150 | 0.475 | 0.000 |

Рис. 9. Подсчет размеченных элементов вторичной структуры для всей цепи, полученных с помощью алгоритмов DSSP и STRIDE для модели структуры 1YWR.

* С целью проверить, не случайно ли так получилось, я также посмотрела суммарное количество размеченных алгоритмами DSSP и STRIDE элементов вторичной структуры для моделей [1ATK](#) (см. Рис. 10) и [6R3V](#) (см. Рис. 11). Получила тот же результат: у DSSP больше «других» структур, чем у STRIDE, а α -спиралей и β -лизов – меньше. Скорее всего, это не случайность.

Summary of Whole Chain

| Method | H | E | O | X |
|-----------|-------|-------|-------|-------|
| CONSENSUS | 0.279 | 0.181 | 0.298 | 0.242 |
| DSSP | 0.279 | 0.191 | 0.530 | 0.000 |
| STRIDE | 0.321 | 0.200 | 0.479 | 0.000 |

Рис. 10. Подсчет размеченных элементов вторичной структуры для всей цепи, полученных с помощью алгоритмов DSSP и STRIDE для модели структуры 1ATK.

Summary of Whole Chain

| Method | H | E | O | X |
|-----------|-------|-------|-------|-------|
| CONSENSUS | 0.285 | 0.000 | 0.631 | 0.084 |
| DSSP | 0.294 | 0.000 | 0.706 | 0.000 |
| STRIDE | 0.312 | 0.000 | 0.688 | 0.000 |

Рис. 11. Подсчет размеченных элементов вторичной структуры для всей цепи, полученных с помощью алгоритмов DSSP и STRIDE для модели структуры 6R3V.

Работа с разметкой вторичной структуры в автоматическом режиме

В [папке](#) мне был дан набор из 30 белков. Находясь в рабочей папке практикума в домашней директории, провела аннотацию вторичной структуры для каждого белка при помощи DSSP. Я использовала следующие команды:

```
for file in $(ls /srv/www/htdocs/FBB/year_18/sbinf2021/2021.T8/spyro); do
mkdssp -i /srv/www/htdocs/FBB/year_18/sbinf2021/2021.T8/spyro/$file -o
./${file}.dssp
done
```

Далее при помощи предоставленного в задании [скрипта](#) остатки группировались по трем типам вторичной структуры: α -спиралям («Н»), β -листам («Е») и петлям («С»).
Использовала такую команду:

```
for file in *.dssp; do python3 parse_dssp.py $file out_$file; done
```

Далее объединила выходы скрипта в один файл:

```
cat out* > out.txt
```

Полученный текстовый [файл](#) обрабатывался на Python. Тетрадка с кодом расположена по [ссылке](#). В Табл. 1 приведен результат разметки вторичной структуры алгоритмом DSSP.

Табл. 1. Склонности образования аминокислотными остатками трех типов вторичной структуры (Н - α -спирали, Е - β -листы, С - петли)

| | A | C | D | E | F | G | H | I | K | L | M | N | P | Q | R | S | T | V | W | Y |
|---|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| Н | 1.393 | 1.005 | 0.818 | 1.284 | 1.094 | 0.411 | 0.779 | 1.149 | 1.022 | 1.332 | 1.402 | 0.782 | 0.415 | 1.180 | 1.271 | 0.859 | 0.798 | 0.736 | 1.162 | 1.107 |
| Е | 0.897 | 1.220 | 0.562 | 0.655 | 1.202 | 0.733 | 1.151 | 1.688 | 0.851 | 1.212 | 0.898 | 0.422 | 0.495 | 0.635 | 0.871 | 0.677 | 1.144 | 2.058 | 1.255 | 1.240 |
| С | 0.747 | 0.896 | 1.338 | 0.940 | 0.836 | 1.571 | 1.100 | 0.574 | 1.051 | 0.650 | 0.739 | 1.429 | 1.677 | 1.028 | 0.852 | 1.255 | 1.089 | 0.720 | 0.761 | 0.809 |

Из таблицы видно, что неструктурированные области в белке чаще всего образованы глицинами и пролинами. У остатков глицина слабо ограничены торсионные углы, что может мешать образованию листов и спиралей. Если остатки глицинов слишком гибкие для образования вторичных структур, то остатки пролинов, наоборот, слишком жесткие, что тоже нарушает регулярность вторичных структур, таких как α -спирали и β -листы.

Аланин – самый «спиралеобразующий» остаток, судя по таблице. Валин – наиболее склонен образовывать β -листы. Не могу назвать точную причину этого явления. Возможно, данные аминокислотные остатки достаточно компактны, а также имеют ограничения на торсионные углы в области, доступной для образования α -спиралей и β -листов.

Источники

1. <https://www.uniprot.org/uniprot/P47811>
2. Golebiowski, Adam & Townes, Jennifer & Lauffersweiler, Matthew & Brugel, Todd & Clark, Michael & Clark, Cynthia & Djung, Jane & Laughlin, Steven & Sabat, Mark & Bookland, Roger & VanRens, John & De, Biswanath & Hsieh, Lily & Janusz, Michael & Walter, Richard & Webster, Mark & Mekel, Marlene. (2005). The development of monocyclic pyrazolone based cytokine synthesis inhibitors. Bioorganic & medicinal chemistry letters. 15. 2285-9. 10.1016/j.bmcl.2005.03.007.