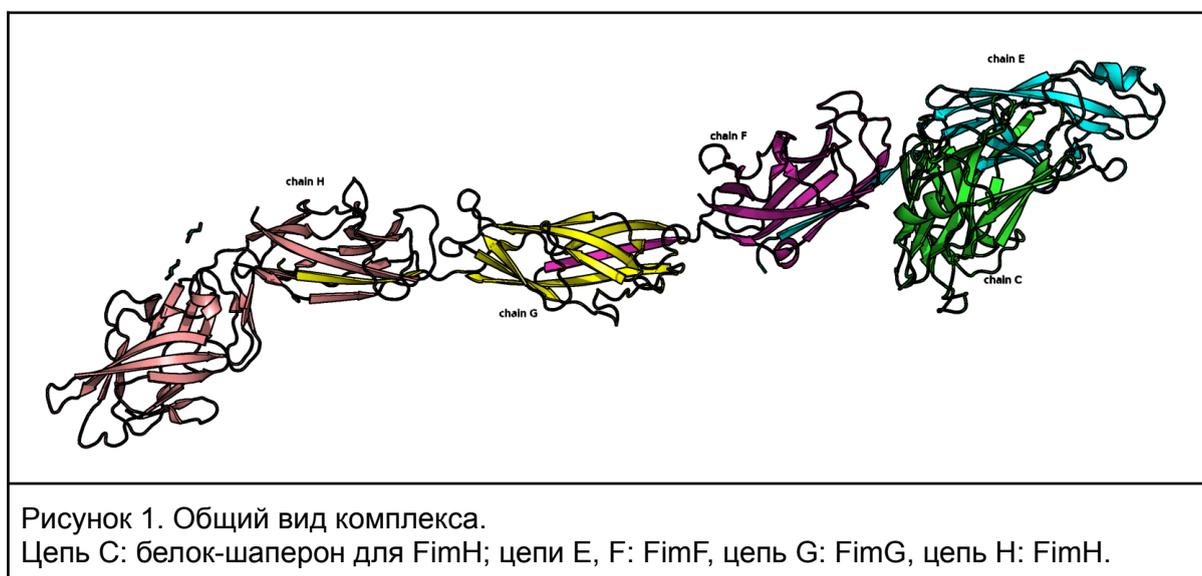


## Введение

Мультидоменные белки адгезии широко распространены среди прокариот и эукариот. Обычно один из доменов принимает участие в связывании целевой молекулы, а другой встроен в мембрану клетки и/или органеллы адгезии. Взаимодействие с закоривающим доменом могут модулировать связывание лиганда. Некоторые из них могут работать по принципу “catch bonds”, т.е. усиливать силу связывания при механическом растяжении. Такие связи более устойчивы к водорастворимым ингибиторам (Yakovenko et al. 2008).

Одним из таких белков является комплекс кишечной палочки FimH, который помогает бактериям удерживаться на различных эпителиях, а также неорганических поверхностях. Установление структуры данного комплекса позволило уточнить механизм работы данного комплекса и показало, что он действительно функционирует по механизму “catch bonds” (Le Trong et al. 2010).



|        |      |
|--------|------|
| PDB ID | 3jwn |
|--------|------|

## Разметка вторичной структуры вручную

Используя веб-сервис 2Struc вручную проанотируем элементы вторичной структуры для некоторых цепей данного белка и рассмотрим те элементы, которые были по-разному проанотированы алгоритмами DSSP и STRIDE.

|           |  |            |        |            |            |            |            |
|-----------|--|------------|--------|------------|------------|------------|------------|
| RESNUM    | 1  | 51         | 61     | 71         | 81         | 91         | 101        |
| SEQ       | ATTPVVPFRILLSPCGNAVSAVKVGF TGVADSHNANLLALENTVSAAGLGIQLLNEQQNQIIF |            |        |            |            |            |            |
| CONSENSUS | 1  | XXXXXXXXXX | 000000 | XXXXXXXXXX | 0000000000 | XXXXXXXXXX | 0000000000 |
| DSSP      | 1  | 000000     | 000000 | 000000     | 0000000000 | 0000000000 | 0000000000 |
| STRIDE    | 1  | EE0000     | 000000 | 0000000000 | 0000000000 | 0000000000 | 0000000000 |

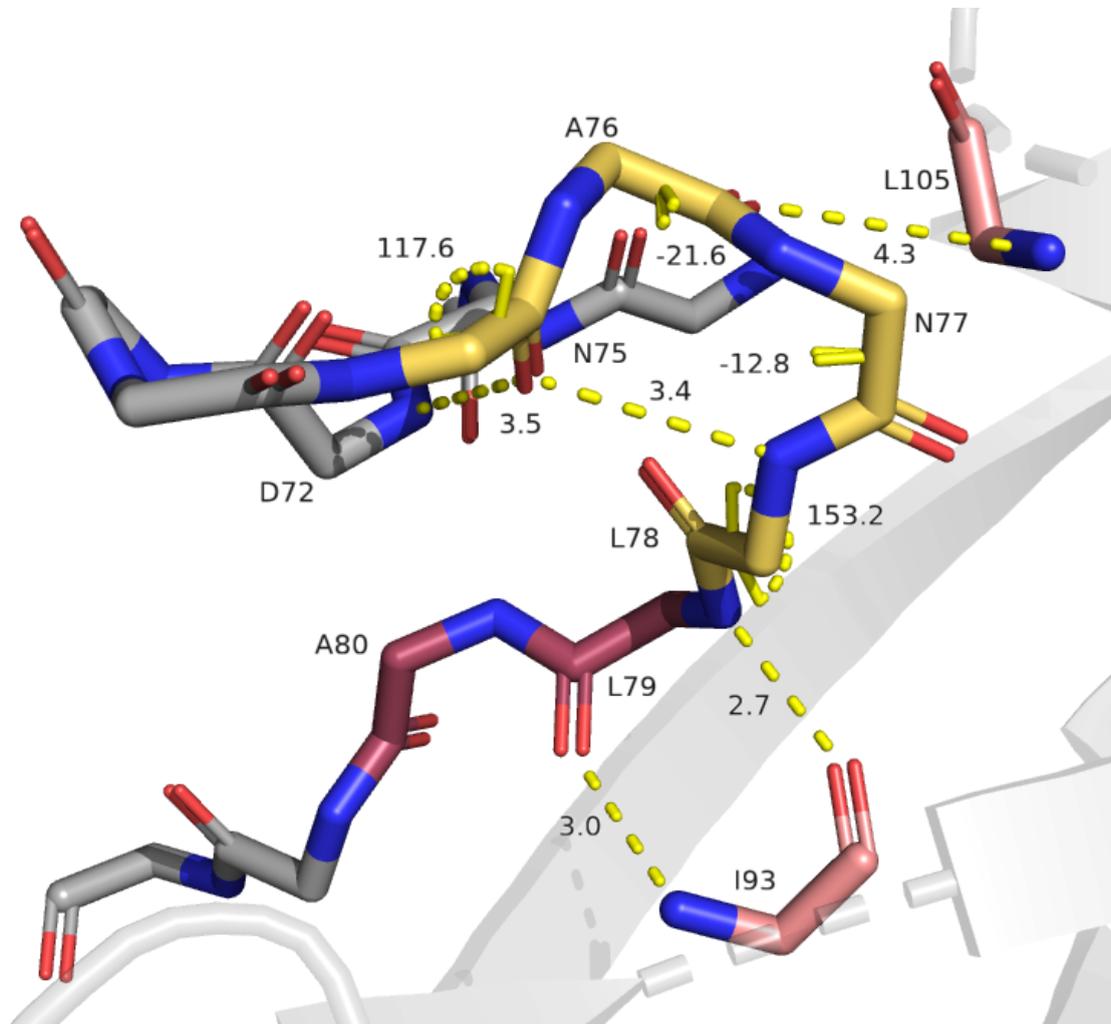


Рисунок 2. Аннотация вторичной структуры, полученная различными алгоритмами для участка цепи F. Ручное определение углов и длин связей в данной структуре. Винным покрашен участок, одинаково аннотированный обоими алгоритмами, желтым -- участок, аннотированный DSSP как  $\beta$ -лист.

Для данного участка цепи, на мой взгляд, алгоритм STRIDE определяет тип вторичной структуры лучше, чем алгоритм DSSP. Мы видим, что торсионные углы данной петли не соответствуют теоретическим торсионным углам, которые наблюдаются для  $\beta$ -листов. Данный участок немного похож на  $\beta$ -петлю (геометрия не оптимальна) и в данном случае аннотацию STRIDE можно считать более верной.

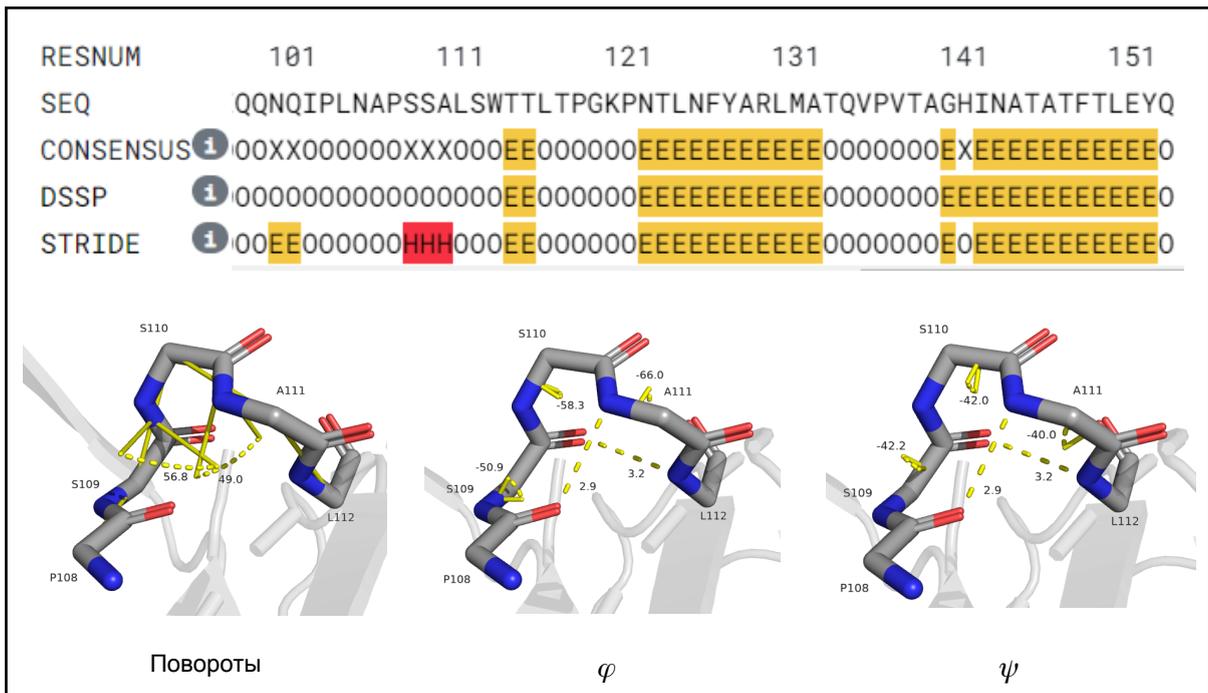
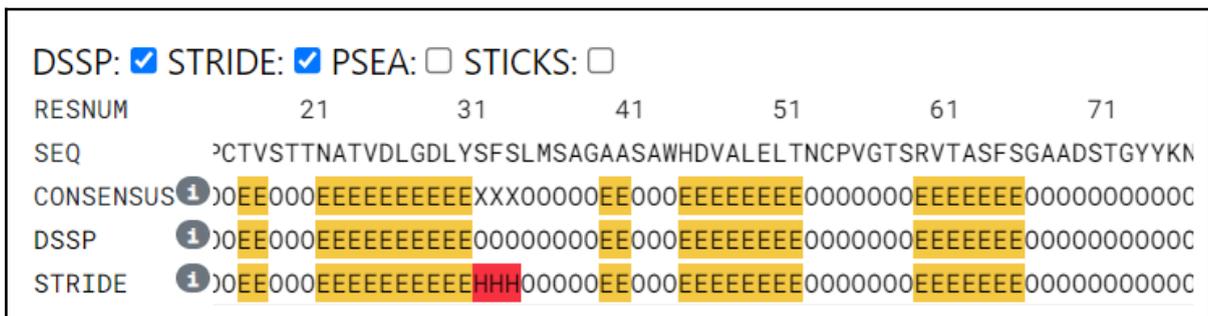


Рисунок 3. Аннотация вторичной структуры, полученная различными алгоритмами для участка цепи F. Ручное определение углов и длин связей в данной структуре. Данный участок аннотируется алгоритмом STRIDE как  $\alpha$ -петля (конкретно, 3-10- $\alpha$ -спираль), алгоритмом DSSP -- как участок без определенной структуры.

Для данного участка цепи, как и в предыдущем случае, лучше работает алгоритм STRIDE. Для данного участка торсионные углы  $\varphi$  и  $\psi$  попадают в области, характерные для  $\alpha$ -спиралей; помимо этого мы видим достаточно высокую степень закрученности (меньше, чем для 4-10- $\alpha$ -спиралей, что мы и ожидали для 3-10- $\alpha$ -спирали). Расположение водородных связей также соответствует ожидаемому для спирали. Данный участок имеет очень небольшую длину (аннотируются 3 аминокислоты, в состав спирали, видимо, входят пять), что также ожидается для данного типа вторичной структуры.





вторичной структуры, что также может влиять на определение небольших участков (таких как 3-10- $\alpha$ -повороты). Алгоритм STRIDE же, помимо паттерна водородных связей, определяет еще и торсионные углы для аминокислот и для каждого остатка определяет вероятность того, что он принадлежит  $\alpha$ -спирали или  $\beta$ -листу. На основании этих данных собираются уже зародыши элементов вторичных структур (а не паттернов водородных связей); терминальные остатки могут быть включены в паттерн при соблюдении определенных условий. Это, на мой взгляд, позволяет алгоритмам одинаково работать при определении протяженных участков вторичной структуры и по-разному при работе с различными мелкими паттернами.

Рассмотрим еще два алгоритма, которые нам предлагает данный веб-сервис: P-SEA и STICKS. Алгоритм P-SEA использует информацию только о положениях  $C_\alpha$ -атомов, по которым считает расстояния между остатками  $(i - 1)$  и  $(i + 1)$ ,  $(i + 2)$ ,  $(i + 3)$  и углы  $\tau_i$  и  $\alpha_i$  между  $C_\alpha$ -атомами  $(i - 1, i, i + 1)$  и  $(i - 1, i, i + 1, i + 2)$ . Вторичная структура аннотируется исходя из данных значений следующим образом:

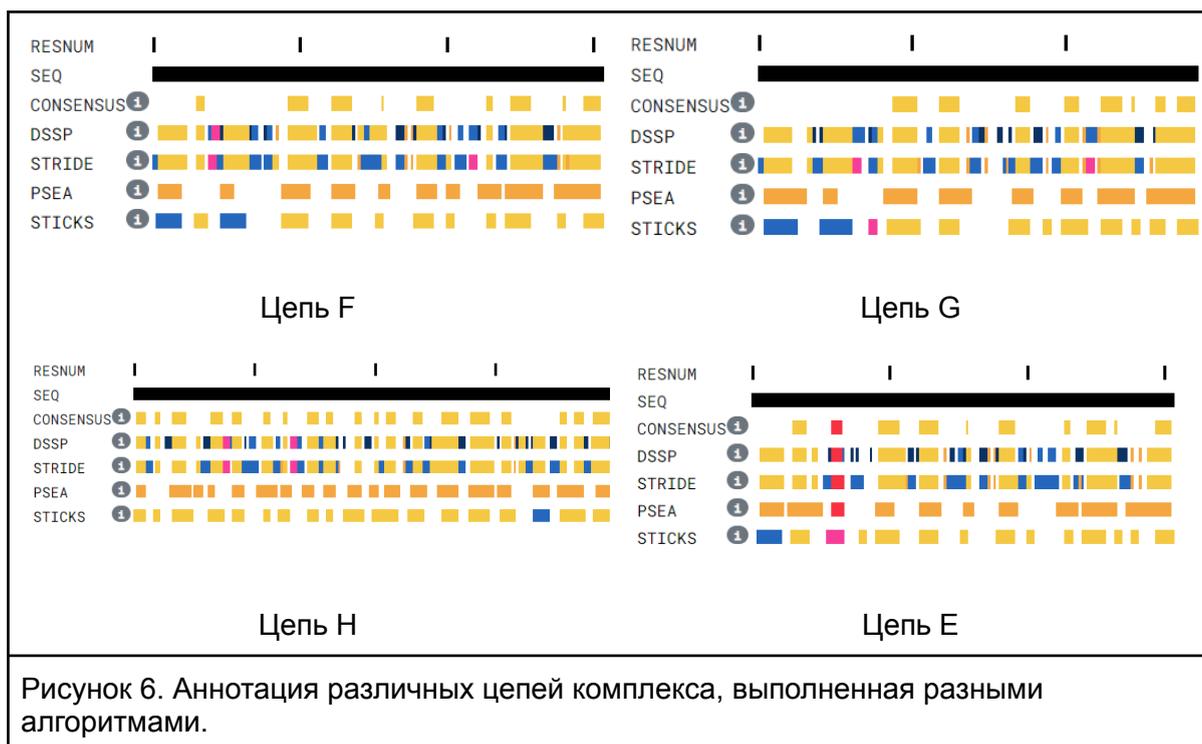
| Assignment parameters                | Secondary structure |                |
|--------------------------------------|---------------------|----------------|
|                                      | Helix               | Strand         |
| Angle $\tau$ ( $^\circ$ )            | $89 \pm 12$         | $124 \pm 14$   |
| Dihedral angle $\alpha$ ( $^\circ$ ) | $50 \pm 20$         | $-170 \pm 45$  |
| Distance d2 ( $\text{Å}$ )           | $5.5 \pm 0.5$       | $6.7 \pm 0.6$  |
| Distance d3 ( $\text{Å}$ )           | $5.3 \pm 0.5$       | $9.9 \pm 0.9$  |
| Distance d4 ( $\text{Å}$ )           | $6.4 \pm 0.6$       | $12.4 \pm 1.1$ |

Таблица 1. Значения параметров, соответствующие элементам вторичной структуры.

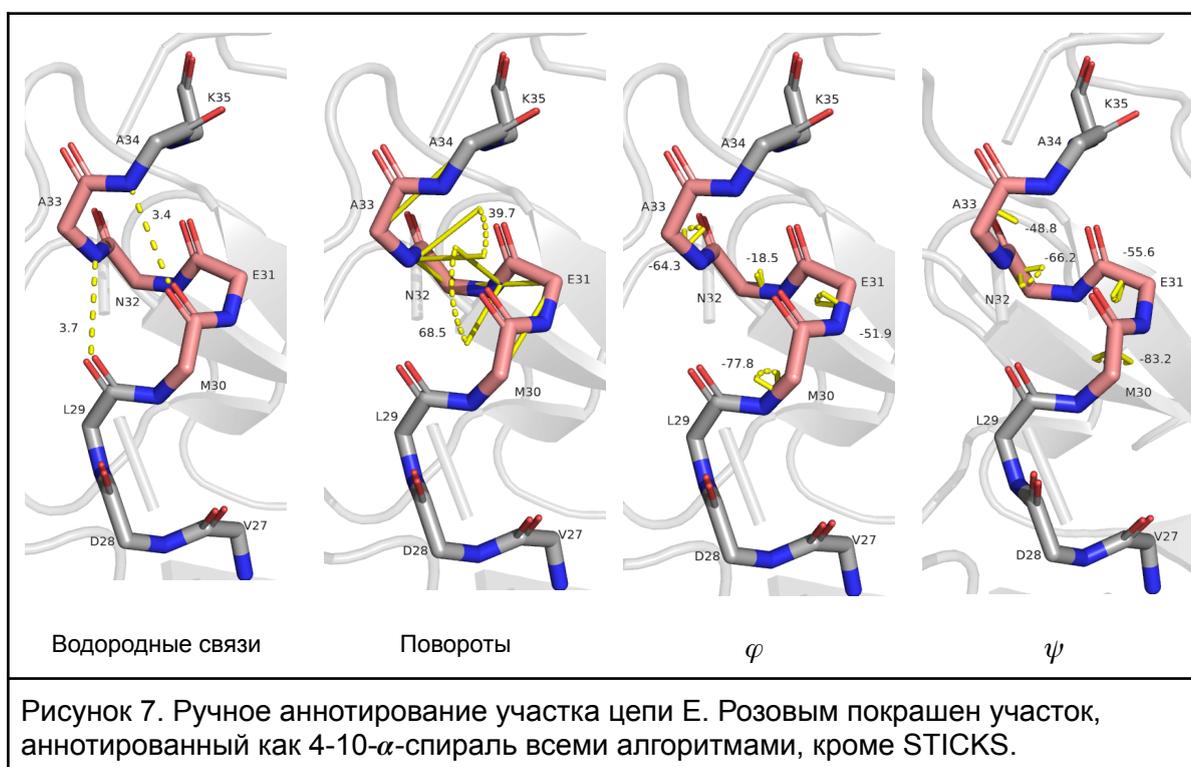
В целом, данный алгоритм аннотирует вторичную структуру также, как и DSSP/STRIDE (Labesse et al. 1997).

Иначе работает алгоритм STICKS, который предсказывает элементы вторичной структуры не в дискретном виде, что, по мнению авторов, позволяет проще находить более редкие и короткие по длине элементы, такие как 3-10- $\alpha$ -спирали. В ходе данного алгоритма структура разбивается на независимые друг от друга “линейные” участки. Каждый такой участок характеризуется распределением длин связей и двугранных углов; на основании данных величин участок аннотируется как какой-то элемент вторичной структуры. Для расчетов используются сглаженные координаты  $C_\alpha$ -атомов (Labesse et al. 1997; Taylor 2001).

Рассмотрим, как все алгоритмы аннотируют цепи данного комплекса:



Алгоритм P-SEA не умеет размечать в структуре никакие элементы кроме  $\alpha$ -спиралей и  $\beta$ -листов. Вместе с этим, алгоритм STICKS, который должен хорошо предсказывать короткие нестандартные элементы, часто с этим не справляется. Рассмотрим один из участков, которые данный алгоритм аннотирует как 3-10- $\alpha$ -спираль, в то время как все остальные – как 4-10- $\alpha$ -спираль.



В данном случае мы видим очень явную 4-10- $\alpha$ -спираль, причем все параметры для спирали в целом лежат в допустимых значениях. STICKS также расширяет границы этой спирали на остатки 29 и 28, что очевидно противоречит модели. В целом, данный алгоритм работает наиболее плохим образом среди использованных.

## Разметка в автоматическом режиме

Аннотируем 30 структур различных белков с помощью алгоритма DSSP, разметив в них участки, принадлежащие  $\alpha$ -спиралям,  $\beta$ -листам или петлям, а затем соберем все полученные данные в один датасет и определим, какие аминокислотные остатки склонны образовывать названные типы вторичных структур. Для этого посчитаем следующую величину:

$$P_{ik} = \frac{n_{ik} \cdot N}{n_i \cdot N_k}, \text{ где}$$

$n_{ik}$  – 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,

$N$  – общее количество остатков в датасете.

|     | Е     | Н     | С     |
|-----|-------|-------|-------|
| Res |       |       |       |
| M   | 1.305 | 1.032 | 0.853 |
| V   | 2.107 | 0.772 | 0.707 |
| N   | 0.618 | 0.755 | 1.327 |
| A   | 0.745 | 1.479 | 0.770 |
| I   | 1.880 | 0.959 | 0.669 |
| R   | 0.796 | 1.282 | 0.887 |
| T   | 1.188 | 0.795 | 1.066 |
| P   | 0.354 | 0.622 | 1.527 |
| D   | 0.427 | 0.812 | 1.365 |
| Q   | 0.728 | 1.119 | 1.028 |
| F   | 1.549 | 0.975 | 0.793 |
| S   | 0.981 | 0.958 | 1.037 |
| L   | 1.116 | 1.365 | 0.698 |
| Y   | 1.344 | 0.933 | 0.906 |
| G   | 0.568 | 0.381 | 1.608 |
| H   | 0.827 | 0.991 | 1.077 |
| E   | 0.698 | 1.384 | 0.855 |
| C   | 1.289 | 0.618 | 1.149 |
| W   | 2.034 | 0.808 | 0.712 |
| K   | 0.669 | 0.910 | 1.198 |

В процессе анализа с использованием python была получена следующая таблица частот встречаемости аминокислотных остатков, где 'E' соответствует  $\beta$ -листу, 'H' –  $\alpha$ -спирали, а 'C' – петлям.

Стоит отметить, что в состав  $\alpha$ -спиралей наиболее часто входят остатки аланина, лейцина и глутамата, в состав  $\beta$ -листов – остатки валина, триптофана и изолейцина, в состав петель – остатки глицина и пролина.

Остатки, наиболее часто встречающиеся в составе петель, были ожидаемы: глицин из-за отсутствия хиральности и высокой подвижности может немного нарушать геометрию более структурированных элементов вторичных структур, а пролин, как аминокислота изгибающая белковый остов, не может (как минимум, достаточно часто) входить в состав структурированных элементов вторичной структуры.

- Labesse, G., N. Colloc'h, J. Pothier, and J. P. Mornon. 1997. "P-SEA: A New Efficient Assignment of Secondary Structure from C Alpha Trace of Proteins." *Computer Applications in the Biosciences: CABIOS* 13 (3): 291–95.
- Le Trong, Isolde, Pavel Aprikian, Brian A. Kidd, Manu Forero-Shelton, Veronika Tchesnokova, Ponni Rajagopal, Victoria Rodriguez, et al. 2010. "Structural Basis for Mechanical Force Regulation of the Adhesin FimH via Finger Trap-like Beta Sheet Twisting." *Cell* 141 (4): 645–55.
- Taylor, W. R. 2001. "Defining Linear Segments in Protein Structure." *Journal of Molecular Biology* 310 (5): 1135–50.
- Yakovenko, Olga, Shivani Sharma, Manu Forero, Veronika Tchesnokova, Pavel Aprikian, Brian Kidd, Albert Mach, Viola Vogel, Evgeni Sokurenko, and Wendy E. Thomas. 2008. "FimH Forms Catch Bonds That Are Enhanced by Mechanical Force due to Allosteric Regulation." *The Journal of Biological Chemistry* 283 (17): 11596–605.