

## Разметка вторичной структуры

### Задание 1. Знакомство с укладками

Для выполнения задания было получено 10 записей PDB, содержащих структуры, имеющие по последовательностям не более 40% попарного сходства. Выравнивание всех структур на одну (с ID 1vx4q01) показано на рис.1. Заметна общая схожесть пространственной организации, несколько  $\alpha$ -спиралей хорошо накладываются во всех структурах. Тем не менее есть и значимые отличия, так, например, участок между 180 и 190 аминокислотными остатками начальной структуры у 1vx4q01 и еще нескольких структур является  $\beta$ -тяжом, а у части — состоит из  $\alpha$ -спирали и  $\beta$ -тяжа или не структурирован вовсе. Также хорошо видны два участка, не имеющие аналогов в других структурах. Это С-конец 4hubG01 (голубого цвета на рис.1), состоящий, к тому же полностью из неизвестных аминокислотных остатков и N-конец 4v19J00 (светло-зеленого цвета на рис.1), он уже состоит из известных аминокислот.

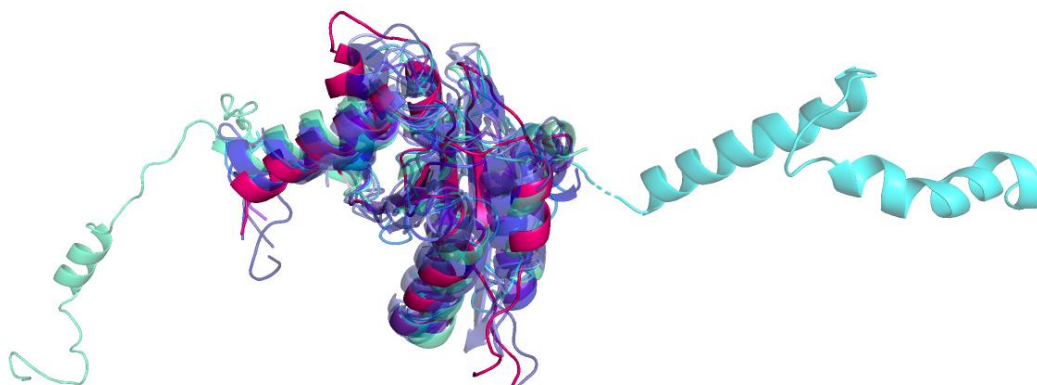


Рис.1 Выравнивание всех структур на 1vx4q01 (розового цвета).

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

Был рассмотрен более подробно участок, разметка которого сделанная Rm01 по умолчанию (с использованием dssp) различается между структурами 1vx4q01, где участку не приписана определенная вторичная структура, и 1zavA01, где он является  $\alpha$ -спиралью (рис.2). Пространственно этот участок находится в одном месте у обеих структур, однако различается по последовательности: в 1vx4q01 он находится примерно с 59 по 62 аминокислотные остатки, а в 1zavA01 — с 60 по 63 остатки.

На рис.3 и рис.4 показаны водородные связи в названном участке в структурах 1vx4q01 и 1zavA01 соответственно. В первой структуре возможно предположить продление спирали,

начинающейся от 63 а.о только на одно снование вниз, образование же водородных связей, обеспечивающих данную конформацию с атомами 59-61 а.о невозможно. Также можно заметить забавную конструкцию, образуемую THR 59 и MET 61, возможно там была водородная связь, которую при расшифровке приняли за ковалентную (к вопросу о разметке это не очень относится). Во второй же структуре все атомы, размеченные как  $\alpha$ -спираль, могут образовывать необходимые водородные связи. Таким образом, алгоритм сработал, как мне кажется, верно, но к  $\alpha$ -спирали в структуре 1vx4q01 можно «добавить» 62 ARG.

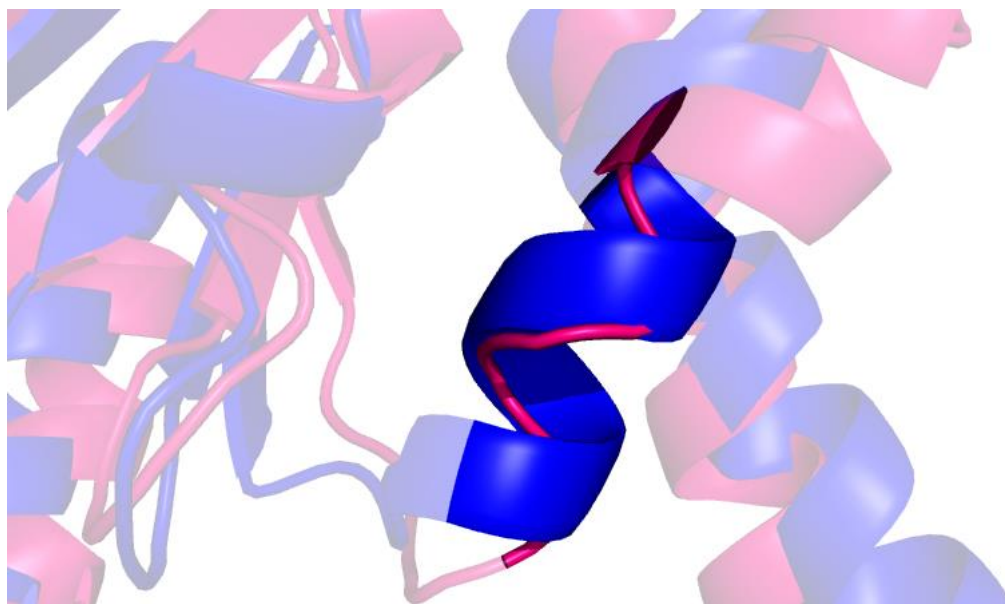


Рис.2 Участок разметка которого сделанная Rm01 по умолчанию (с использованием dssp) различается между структурами 1vx4q01 (розовая) и 1zavA01 (синяя). Участок выделен цветом.

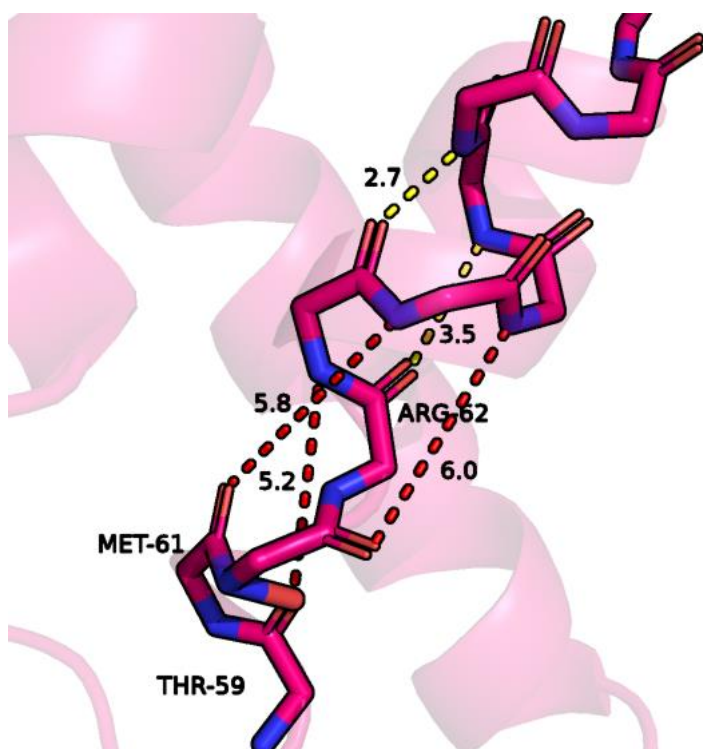


Рис.3 Водородные связи с атомами 59-62 а.о в 1vx4q01 (водородные связи показаны желтым, их отсутствие красным).

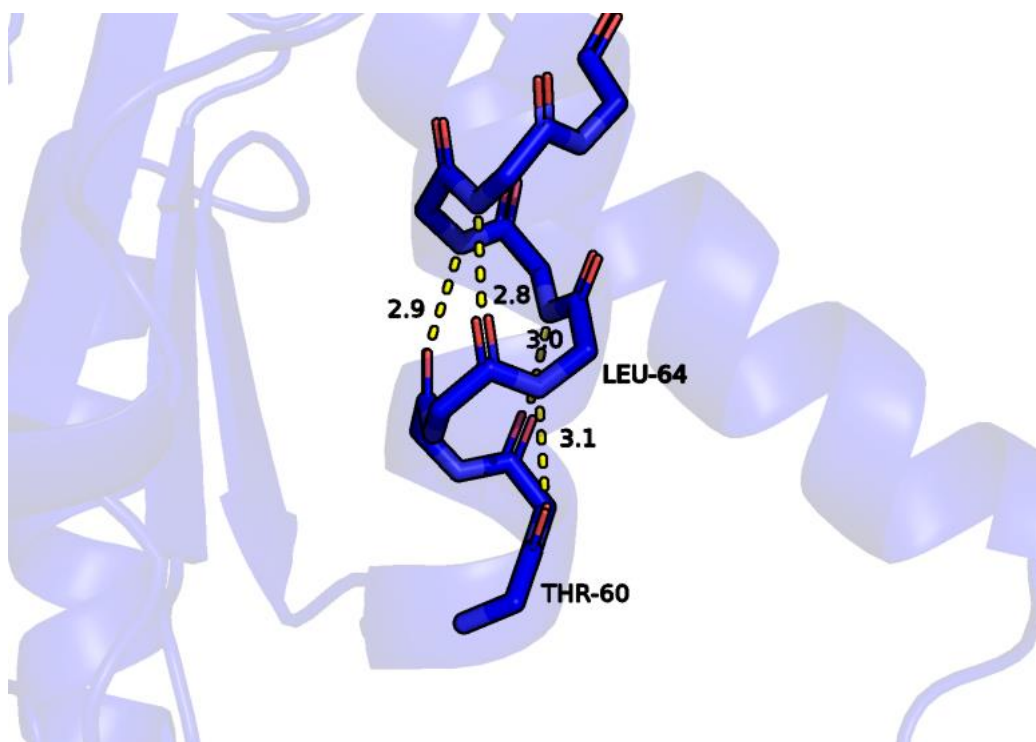


Рис.4 Водородные связи с атомами 60-64 а.о в 1zavA01 (водородные связи показаны желтым).

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

Для каждой записи была получена разметка вторичной структуры с помощью dssp. Для каждого типа аминокислотных остатков была рассчитана склонности образовывать тот или иной тип вторичной структуры по формуле:

$$P_{ik} = (n_{ik}/n_i) / (N_k/N)$$

$P_{ik}$  — 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

К количествам определенных остатков, образующих определенный тип структуры, были добавлены псевдокаунты = 1, чтобы избежать нулевых значений в итоговой таблице. В таблице 1 приведены результаты расчётов по данной выборке. Согласно полученным данным аргинин и глутамат чаще образуют альфа-спирали, чем бета-листы и петли. Глицин заметно чаще образует бета-листы, чем альфа-спирали и петли. Пролин заметно чаще образует петли, чем альфа-спирали и бета-листы

Таблица 1: склонности образовывать типы вторичной структуры для различных аминокислот.

а.о	Петля	β-лист	α-спираль
A	1,00	1,75	2,70
C	0,70	0,79	0,07
D	1,06	0,34	0,86
E	1,00	0,54	1,71
F	0,74	1,27	0,56
G	1,31	3,31	1,14
H	1,20	0,45	0,27
I	0,73	1,25	0,97
K	0,94	0,48	0,86
L	0,94	2,05	2,43
M	0,79	0,64	0,87
N	1,07	0,31	0,80
P	1,56	0,29	0,63
Q	0,72	0,56	1,12
R	0,91	0,40	1,45
S	1,24	0,39	0,65
T	1,15	1,29	0,40
V	0,85	3,76	2,86
W	0,44	0,57	0,41
Y	1,05	3,18	3,27