

# Практикум № 1. Рунol

Иосиф Финкельберг

## 1 Введение

В этом практикуме мы работали со структурами, имеющими PDB-id **6TRB** и **6NFR**. Это структуры белка CopC грамотрицательной бактерии *Pseudomonas fluorescens* (далее белок мы будем называть *Pf-CopC*), полученные с помощью ЯМР (6TRB) [1] и РСА (6NFR) [2]. Бактериальные белки семейства CopC являются медь-связывающими периплазматическими белками, участвующими в том числе в детоксикации меди в случае повышения ее концентрации в окружающей среде [3]. В Cop-C белках имеются отдельные сайты для связывания Cu(I) и Cu(II) (см. рис 1). При этом, как недавно оказалось, некоторые белки обладают лишь сайтом связывания Cu(II) [1, 3]. К таким белкам относится, в частности, *Pf-CopC*.

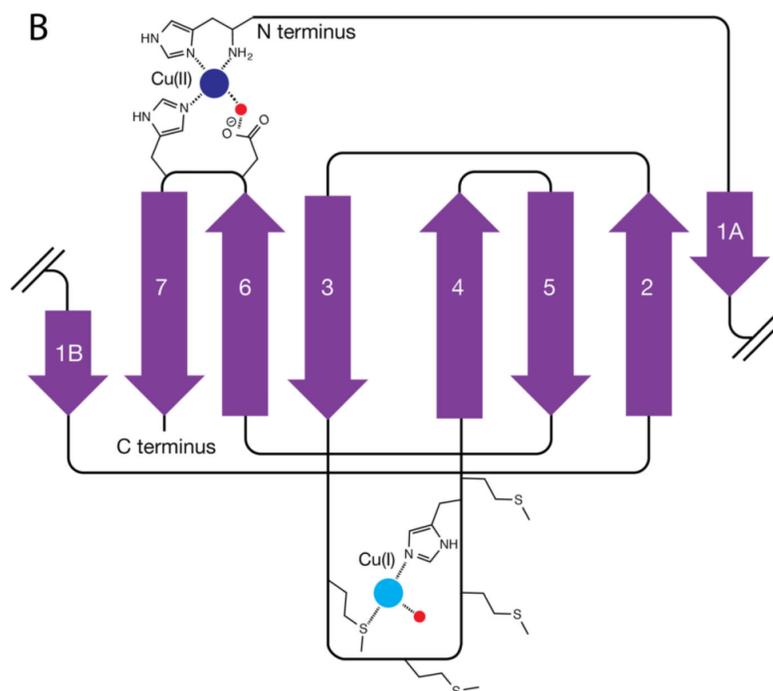


Рис. 1: Схема строения белков из семейства CopC (*Pss-CopC* и *Esc-PcoC*), связывающих медь Cu(I), Cu(II). Темно-синим показаны ионы Cu(II), голубым – ионы Cu(I), красные точки обозначают воду. Ионы Cu(II) координируются за счет двух гистидинов (один из них N-концевой) и аспартата через молекулу воды. Ионы Cu(I) связываются остатками гистидина и метионина (схема взята из статьи [3]).

## 2 Задание 1 (Вводное)

Как указано на сайте PDB, разрешение PCA-модели 6NFR равняется 1.00 Å; в ЯМР-структуре 6PTV имеется 20 моделей. Сперва мы расположили рядом PCA и одну из моделей ЯМР структуры и показали обе структуры с одного ракурса (см. рис 2). На рисунке видно, что вторичная структура белка состоит из бета-листов, которые складываются в мотивы, напоминающие Greek key. Ясно также, что имеются различия в ЯМР и PCA-структурах. Первое из них состоит в окружении молекул белков. Авторы PCA-структуры использовали для кристаллизации сульфат аммония (действительно, в состав структуры попали три иона  $SO_4^{2-}$ ) [1]. Кроме того, в PCA-структуру попали молекулы воды. Что касается крупных различий в самой структуре, то уже на этом этапе заметно, что в ЯМР-структуре имеется дополнительные по сравнению с PCA бета-лист (см. рис 2: мы постарались подписать это место в структуре). Кроме того, по картинке ка-

жется, что некоторые бета-листы в РСА являются более протяженными по сравнению с их аналогами в ЯМР.

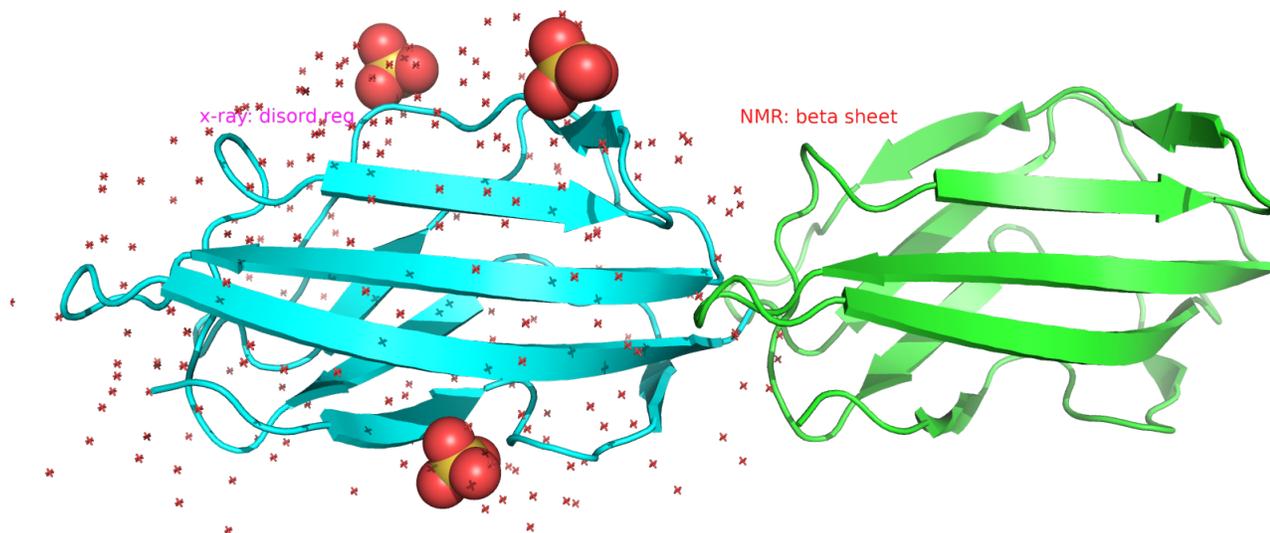


Рис. 2: Сравнение РСА (слева, голубой цвет) и ЯМР (справа, зеленый цвет) структур. Подписано одно и то же с точки зрения последовательности место в структуре: в РСА это неупорядоченный участок, в ЯМР – бета лист.

Затем мы убрали ионы  $SO_4^{2-}$  и воду из РСА-структуры, чтобы не загромождать картинку, и наложили все 20 ЯМР моделей друг на друга и на РСА-структуру (см. рис 3).

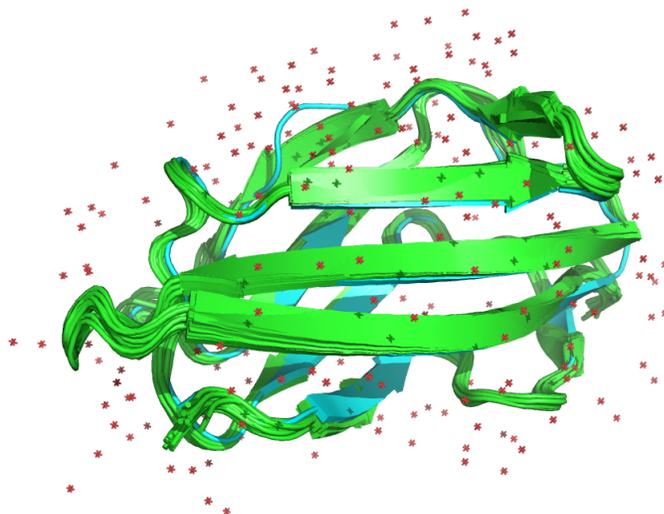


Рис. 3: Выравненные RSA-структура (голубой цвет) и ЯМР-структура из 20 моделей (зеленый цвет).

Пока что мы видим хорошо то же самое отличие, что и на предыдущей картинке: на месте одного из бета-листов ЯМР-структуры в RSA имеется неупорядоченный участок (кстати, как нам говорили на лекции по ЯМР, обычно происходит обратное: в RSA кристаллизуются в упорядоченные структуры участки, которые *in vivo* являются неупорядоченными). Если посмотреть на структурное выравнивание под другим ракурсом (см. рис 4), то заметно также и то, что длины бета-листов разнятся в RSA и ЯМР-моделях (длиннее бывают и там, и там: не исключено, что здесь дело просто в разметке вторичной структуры).

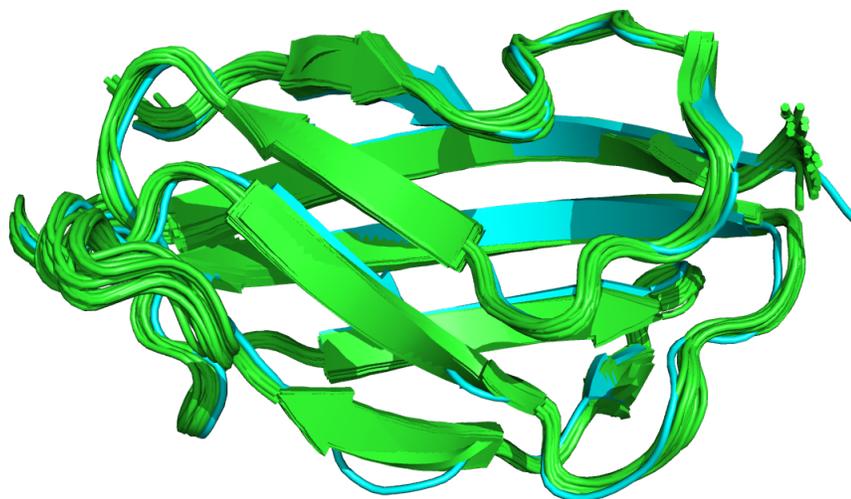


Рис. 4: Выравненные RCSA-структура (голубой цвет) и ЯМР-структура из 20 моделей (зеленый цвет).

Теперь мы решили посмотреть на микро-различия: отличия в положениях остатков (и, возможно, в самих остатках). Так как с концентрацией и внимательностью у нас все плохо, мы, во-первых, оставили только одну ЯМР-модель в структурном выравнивании, и, во-вторых, убрали остовы структуры, оставив только боковые радикалы (см. рис. 5).

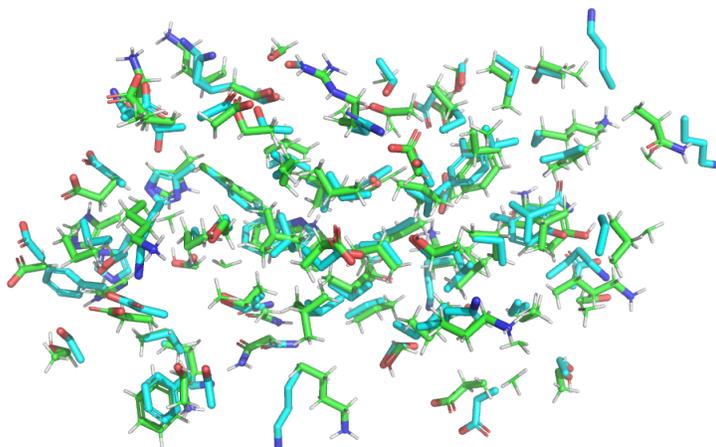


Рис. 5: Выравненные РСА-структура (голубой цвет) и ЯМР-структура (зеленый цвет). На ЯМР-структуре видны атомы водорода.

Заметно, что структуры в принципе сильно отличаются по положениям остатков, а также, что в ЯМР-структуре имеются атомы водорода (впрочем, последнее верно для любой ЯМР структуры, и, наоборот, в РСА никогда экспериментально определить водород невозможно). Кроме того, интересно, что есть отличия даже не в положении остатков, а в самих остатках (см. рис 6). Действительно, например, там, где в последовательности РСА находится фенилаланин, в ЯМР мы видим гистидин. Оказывается, авторы структуры 6NFR сделали РСА для мутированного белка (His109->Phe109 в нумерации, учитывающей сигнальную последовательность пептида) [3].

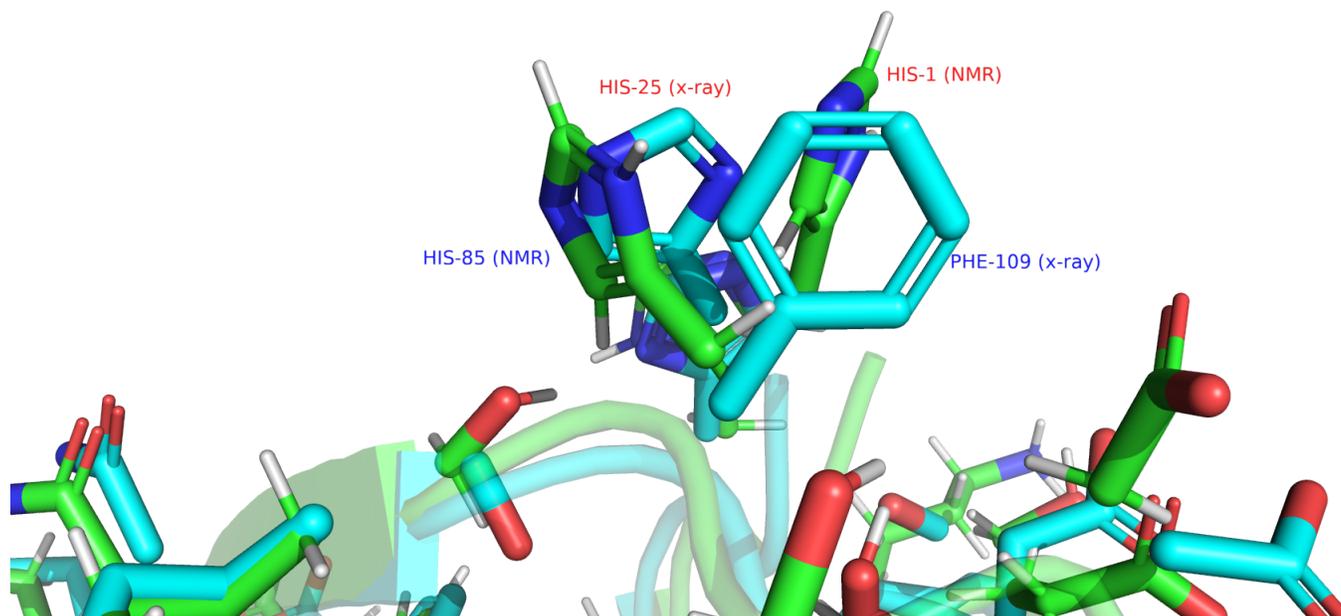


Рис. 6: Выравненные PCA-структура (голубой цвет) и ЯМР-структура (зеленый цвет). Нумерация атомов в PCA начинается с 25, так как авторы учитывают в нумерации сигнальную последовательность белка [3]. Одинаковым цветом подписаны остатки, соответствующие друг другу по последовательности. Видим, что фенилаланину в PCA структуре соответствует гистидин в ЯМР. В WT-белке на этом месте располагается именно гистидин.

### 3 Задание 2. RMSF

В этом задании мы попытались выяснить, отражает ли ансамбль моделей в ЯМР-структуре реальную подвижность атомов в белке. Договоримся, что наша нулевая гипотеза: модели ЯМР-структуры отражают подвижность атомов белка. Альтернативная гипотеза, соответственно, состоит в том, что ансамбль ЯМР-моделей для нашей структуры появляется вследствие экспериментального шума, вызывающего неполноту в данных об ограничениях на расстояния и ковалентные связи в белке.

В качестве референсной меры подвижности мы использовали значения B-фактора для атомов белка, взятые из PCA-структуры. В качестве меры, взятой из ЯМР-структуры, которая в рамках нулевой гипотезы выражает подвижность атомов, мы использовали **RMSF**.

Сперва мы решили приблизительно оценить наличие корреляции между средними значениями RMSF и B-фактора для остатков нешкго белка (см.

рис. 7). Видим, что значения RMSF и B-факторов положительно скоррелированы. На основании этих данных мы не можем отвергнуть нулевую гипотезу о том, что RMSF отражает подвижность атомов.

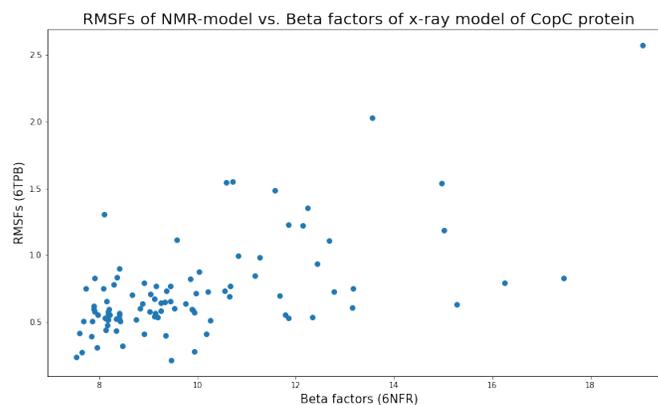


Рис. 7: Корреляция между средними по остаткам значениями RMSF (посчитаны по 6TPB) и B-факторов (посчитаны по 6NFR). Коэффициент корреляции Пирсона равняется 0.60, p-value = 7.63e-11 (посчитано в **предположении**, что значения обеих мер распределены нормально)

Заметим однако, что в данном случае мы рассматривали приближение, в котором мы усредняли значения наших мер по атомам одного остатка. Если же мы перейдем к рассмотрению каждого атома отдельно, то у нас будет более надежный способ проверить истинность нулевой гипотезы. Действительно, согласно уравнению (1) из статьи [4] в случае выполнения нашей нулевой гипотезы зависимость между RMSF и Beta-факторами для атомов белка должна быть следующей:

$$|1| \text{ RMSF}_i^2 = \frac{3 \cdot B_i}{8 \cdot \pi^2} \iff \text{RMSF}_i = \sqrt{\frac{3 \cdot B_i}{8 \cdot \pi^2}}$$

Поскольку не все атомы белка присутствуют в обеих моделях одновременно, мы оставили только общие для моделей атомы (например, не учитывали атомы водорода ЯМР-моделей): нам удалось соотнести 689 атомов белка в двух моделях. Затем мы построили реальную зависимость RMSF от Beta-факторов, и ту зависимость, которую мы ожидали бы увидеть в рамках нулевой гипотезы (уравнение |1|): см. рис 8.

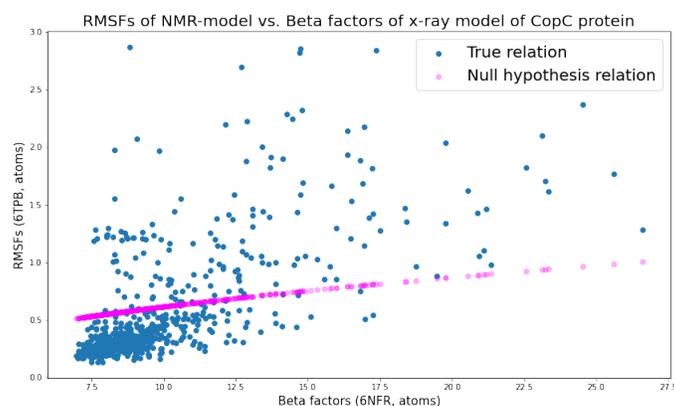


Рис. 8: Корреляция между значениями RMSF (посчитаны по 6TPB) и B-факторов (посчитаны по 6NFR) для 689 атомов белка *Pf*-CopC. Коэффициент корреляции Пирсона равняется 0.67,  $p\text{-value} = 1.36e-90$  (посчитано в **предположении**, что значения обеих мер распределены нормально). Розовая кривая отражает теоретическую зависимость при условии нулевой гипотезы: см. уравнение [1]

Теперь мы видим, что нулевую гипотезу следует отвергнуть, так как теоретическая зависимость никаким образом не отражает реальную (кстати, забавно отметить, что значения обеих мер при этом достаточно сильно скоррелированы).

## 4 Задание 3 (Водородные связи)

В этом задании мы находили в PCA-структуре потенциальные водородные связи, а затем смотрели, как они выглядят в соответствующей ЯМР-структуре. Первую водородную связь мы выбрали между остовами ароматических остатков, входящих в состав бета-листа (связь между двумя антипараллельными бета-нитями): см. рис. 9.

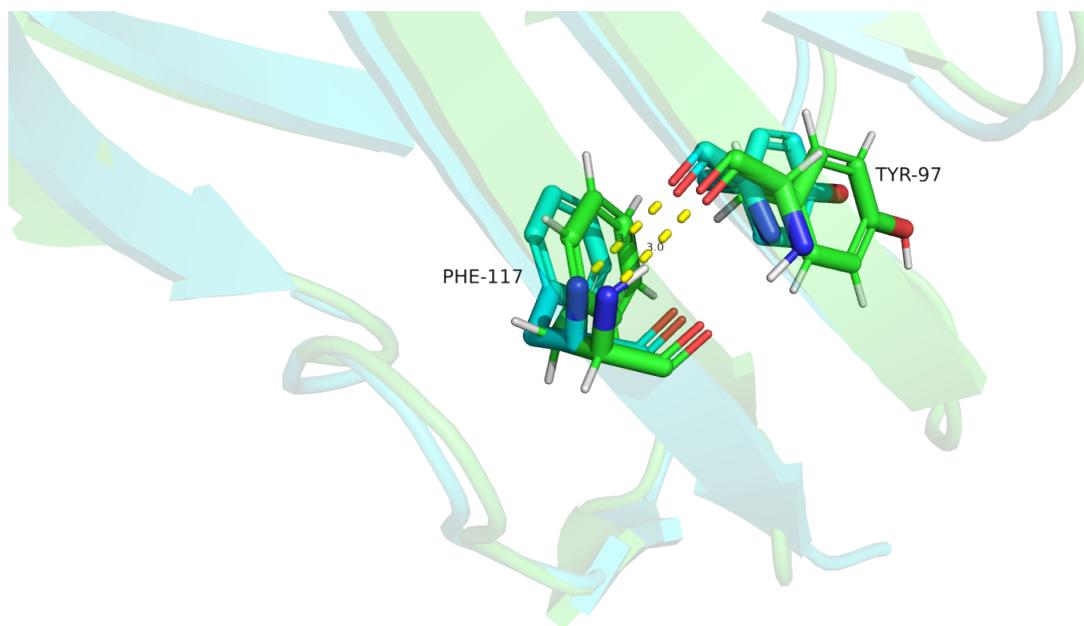


Рис. 9: Показаны выравненные РСА (голубой цвет) и ЯМР (зеленый цвет) структуры. Водородная связь между остовами ароматических остатков, находящихся на двух антипараллельных бета-нитях. Указана нумерация остатков из РСА-структуры.

Вторую водородную связь мы нарисовали между боковым остатком треонина и остовом валина, недалеко от центра белка. Данные аминокислоты находятся в неупорядоченной части структуры, но близко к бета-листам (рис 10).

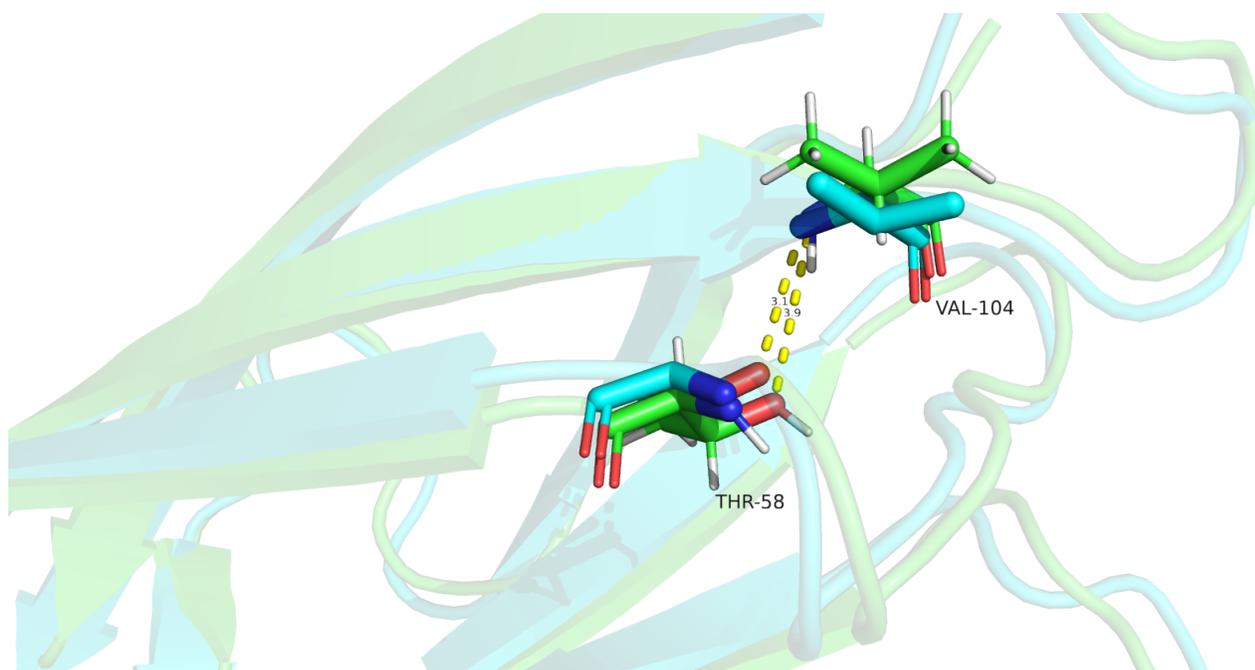


Рис. 10: Показаны выравненные РСА (голубой цвет) и ЯМР (зеленый цвет) структуры. Водородная связь между азотом остова остатка валина и кислородом боковой цепи остатка треонина. Валин и треонин находятся рядом с бета-листами, недалеко от центра белка. Указана нумерация остатков из РСА-структуры.

Третью водородную связь мы изобразили между боковыми цепями остатков аспарагина и аспартата, которые находятся в неупорядоченной периферийной части белка. В данном случае, имелось расхождение между ЯМР и РСА-структурами: в ЯМР-структуре расстояние между соответствующими атомами не позволяло образование водородной связи между ними (рис 11).

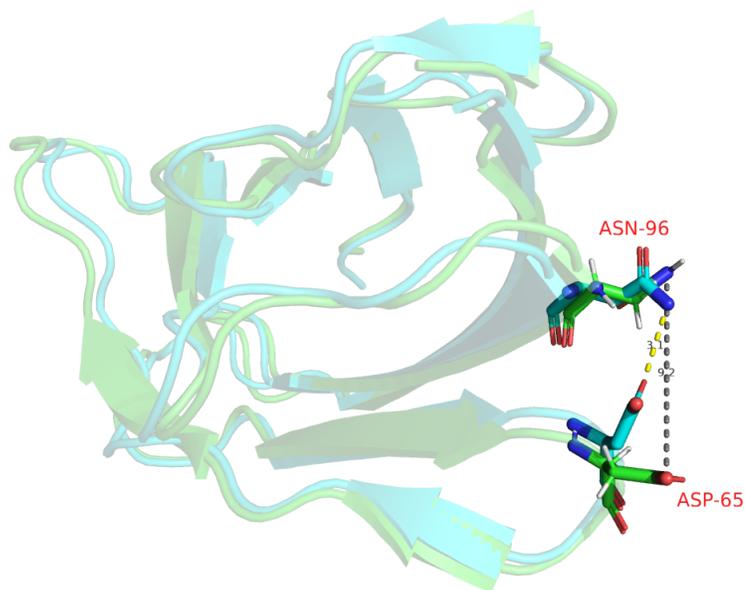


Рис. 11: Показаны выравненные РСА (голубой цвет) и ЯМР (зеленый цвет) структуры. Водородная связь между боковыми цепями остатков аспарагина и аспартата. Эти аминокислоты находятся в периферийной, подвижной части белка. Указана нумерация остатков из РСА-структуры.

Так как наша ЯМР-структура представлена 20 моделями, для которых расстояния и углы предполагаемой водородной связи могут различаться, мы посчитали для каждой из моделей эти два параметра, воспользовавшись python-модулем prody (рис 12).

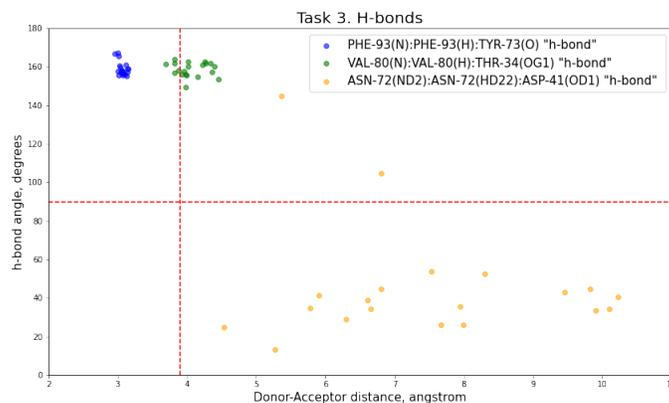


Рис. 12: Углы донор-водород-акцептор и расстояния донор-акцептор предполагаемой водородной связи для трех обозначенных нами выше пар атомов. Нумерация атомов согласно ЯМР-структуре. Одной точкой изображается одна модель. Верхний левый угол, который отрезают красные пунктирные линии – 'типичные' для водородных связей значения углов и расстояний.

Результаты достаточно хорошо согласуются с тем, что мы видим в `runol`: водородная связь между остовами бета-нитей имеет типичные параметры (угол, расстояние) во всех 20 ЯМР-моделях. Водородная связь между боковой цепью треонина и остовом валина в неупорядоченной части структуры недалеко от центра белка имеется уже не во всех ЯМР-моделях, расстояния между донорным и акцепторным атомами для нее лежат на границе допустимых значений. Водородная связь между боковыми цепями остатков, находящихся в периферийной неупорядоченной части белка, которую позволяет предположить РСА-структура, не возможно с точки зрения ЯМР-моделей. В таблице 1 мы привели отчет по каждой из трех выбранных связей.

Интересно было бы понять, что бы мы делали, не будь у нас ЯМР-структуры: правда ли, что для нас все три отмеченные водородные связи имели бы одинаковое право на существование? Чтобы проверить это, мы раскрасили нашу РСА-структуру по В-фактору. Сперва мы сделали это только для остова структуры (см. рис 13). Мы видим, что остов предположительно наиболее подвижен (имеет наибольшие значения В-фактора) для остатка аспарагиновой кислоты. Именно для пары ASP-ASN водородная связь была невозможна с точки зрения ЯМР-структуры. Таким образом, уже раскрасив по В-фактору остов, мы можем сделать вывод о (не)надежности гипотезы о существовании водородной связи.

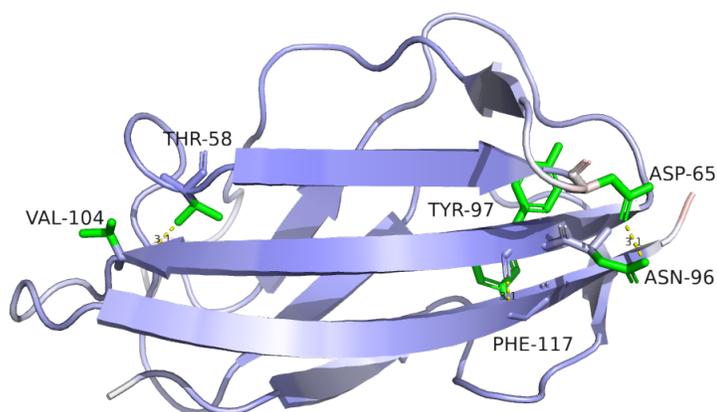


Рис. 13: PCA-структура 6NFR. Остов раскрашен по значениям B-фактора (синий – низкие значения, красный высокие). Зеленым показаны пары остатков, между которыми мы предположили наличие водородной связи. Нумерация остатков дана с точки зрения PCA-структуры.

Наконец, мы раскрасили по B-фактору уже сами остатки (и их остовы): см. рис. 14. Тут уже окончательно становится ясно, что водородной связи, которую мы изначально предположили между остатками ASP-65 и ASN-96 не существует. Таким образом, можно сделать вывод о том, что ЯМР-структура полезна для верификации положения остатков, однако в данном случае достаточно было бы иметь и одну только PCA-структуру.

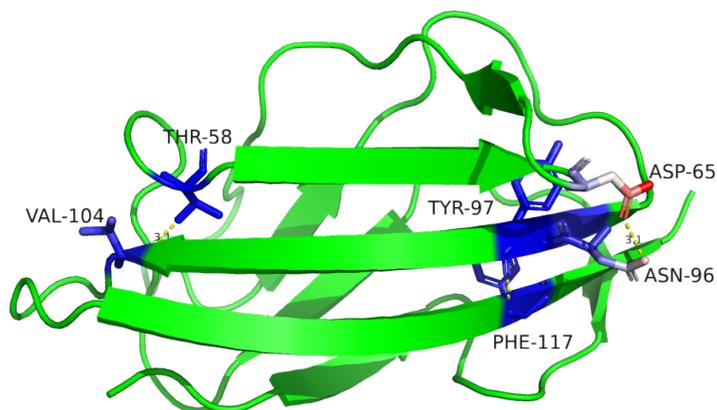


Рис. 14: РСА-структура 6NFR. Остов показан зеленым. По значениям В-фактора (синий – низкие значения, красный высокие) раскрашены пары остатков, между которыми мы предположили наличие водородной связи. Нумерация остатков дана с точки зрения РСА-структуры.

## 5 Jupyter-notebook и таблица с отчетом

<https://kodomo.fbb.msu.ru/~osyafinkelberg/term7/prac5/task2.ipynb>  
<https://kodomo.fbb.msu.ru/~osyafinkelberg/term7/prac5/table1.pdf>

## 6 Сессии PyMol

<https://kodomo.fbb.msu.ru/~osyafinkelberg/term7/prac5/session0.pse>  
<https://kodomo.fbb.msu.ru/~osyafinkelberg/term7/prac5/session1.pse>  
<https://kodomo.fbb.msu.ru/~osyafinkelberg/term7/prac5/session2.pse>  
<https://kodomo.fbb.msu.ru/~osyafinkelberg/term7/prac5/session3.pse>  
<https://kodomo.fbb.msu.ru/~osyafinkelberg/term7/prac5/session4.pse>  
<https://kodomo.fbb.msu.ru/~osyafinkelberg/term7/prac5/session5.pse>  
<https://kodomo.fbb.msu.ru/~osyafinkelberg/term7/prac5/session6.pse>

## 7 ССЫЛКИ

[1] Udagedara et al., 'The crystal structure of the CopC protein from *Pseudomonas fluorescens* reveals amended classifications for the CopC protein family', 2019 *Journal of Inorganic Biochemistry*

[2] Persson et al., 'NMR structure of the apo-form of *Pseudomonas fluorescens* CopC', 2021 to be published

[3] Lawton et al., 'The CopC Family: Structural and Bioinformatic Insights into a Diverse Group of Periplasmic Copper Binding Proteins', 2016 *Biochemistry*

[4] Kuzmanic and Zagrovic, 'Determination of Ensemble-Average Pairwise Root Mean-Square Deviation from Experimental B-Factors', 2010 *Biophysical Journal*