

**МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ  
имени М.В. ЛОМОНОСОВА**

***ФАКУЛЬТЕТ БИОИНЖЕНЕРИИ И БИОИНФОРМАТИКИ***

**Отчёт по качеству расшифровки РСА структуры 2YA5  
каталитического домена нейраминидазы А из *Streptococcus pneumoniae***

Выполнил студент 4-го курса  
Власов Антон Викторович  
в рамках курса структурной биоинформатики

**Москва, 2021**

# СОДЕРЖАНИЕ

<b>СОДЕРЖАНИЕ</b>	<b>1</b>
<b>АННОТАЦИЯ</b>	<b>1</b>
<b>ВВЕДЕНИЕ</b>	<b>2</b>
<b>РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ</b>	<b>3</b>
Общая информация о модели	3
Значения индикаторов качества модели в целом	4
Список маргинальных остатков	6
Анализ маргинальных остатков	7
PDB и PDB-redo	10
<b>ЗАКЛЮЧЕНИЕ</b>	<b>11</b>
<b>БЛАГОДАРНОСТИ</b>	<b>11</b>
<b>СПИСОК ЛИТЕРАТУРЫ</b>	<b>12</b>

## АННОТАЦИЯ

В представленном отчёте рассмотрена структура каталитического домена нейраминидазы А из *Streptococcus pneumoniae* с PDB кодом 2YA5. Представлена краткая информация о модели: причина изучения фермента, кристаллографические параметры структуры. Также были приведены как общие показатели качества модели, так и частные для отдельных остатков. Были описаны некоторые маргинальные остатки и причины их возникновения в структуре. Наконец, исходную структуру сравнили с улучшенной, полученной из банка PDB-REDO по рефлексам из того же эксперимента.

## ВВЕДЕНИЕ

Нейраминидаза типа А (далее NanA) из *Streptococcus pneumoniae* - ключевой фактор вирулентности патогена, обеспечивающий формирование пневмококком биоплёнки на поверхности дыхательного эпителия человека. Несмотря на вакцинирование детей и взрослых, уровень летальности от воспалительных заболеваний, вызванных *S.pneumoniae*, остаётся высоким для детей младше 5 лет и лиц с ослабленным иммунитетом. Данный фермент является перспективной мишенью для разработки лекарственных средств, направленных на борьбу с респираторными заболеваниями, что объясняет интерес к нему научного сообщества [1]. Было показано, что данный белок состоит из каталитического и лектинового домена, позволяющего ферменту лучше связываться с субстратом - терминальными сиаловыми кислотами гликоконъюгатов. Указанные домены разделены подвижным линкером, который делает невозможным разрешение полной структуры фермента ввиду огромного множества конформаций, возникающих при кристаллизации. Каталитический домен представлен 6-лопастным бета-пропеллером и обладает необходимой гидролитической активностью. Также в его структуре имеется вставка между второй и третьей цепями второго листа бета-пропеллера, которую называют вставочным доменом [2].

В данной работе рассмотрена структура каталитического домена с сиаловой кислотой в активном центре NanA с PDB кодом 2YA5. Она была получена в рамках исследования структурной основы действия ингибиторов нейраминидазы вируса гриппа занамивира и осельтамивира. Помимо указанной структуры, также были получены структура каталитического домена без лиганда и 3 структуры с различными ингибиторами [3].

# РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

## Общая информация о модели

Модель 2YA5 была отправлена и опубликована в PDB в 2011 году Хайнцом Гутом и коллегами. Всего за автором насчитывается 56 различных структур в банке белков. Нескольким моделям каталитического домена NanA с различными лигандами соответствуют статья [3] и структуры с номерами 2YA4 - 2YA8 в банке белков. Ген для экспрессии NanA в *E.coli* был взят из *S.pneumoniae*.

В состав модели 2YA5 входит фрагмент NanA размером 55-КДа с остатками 303-777, включающий в себя каталитический и вставочный (или нерегулярный, как его описывают авторы статьи) домены. В записи представлены цепи А и В. Примечательно, что на самой странице модели в PDB [4] нет указания на доменный состав структуры, там она называется просто кристаллической структурой нейраминидазы А. В самом PDB-файле со структурой в метаданных указано, что описан только каталитический домен с номерами остатков 280-754. Далее следует комментарий, что лучше всего полученной модели соответствуют остатки 303-777 из Uniprot последовательности B2DJD9, ныне описанной как устаревшая [4,5], поэтому далее нумерация в модели соответствует последнему варианту. Также в состав комплекса входит лиганд сиаловая кислота (сокращённо SIA, N-ацетил-альфа-нейраминовая кислота) в активном центре. Также в модели присутствуют ионы хлора и молекулы формиата (как один из компонентов раствора, в котором проводилась кристаллизация) и растворителя (воды).

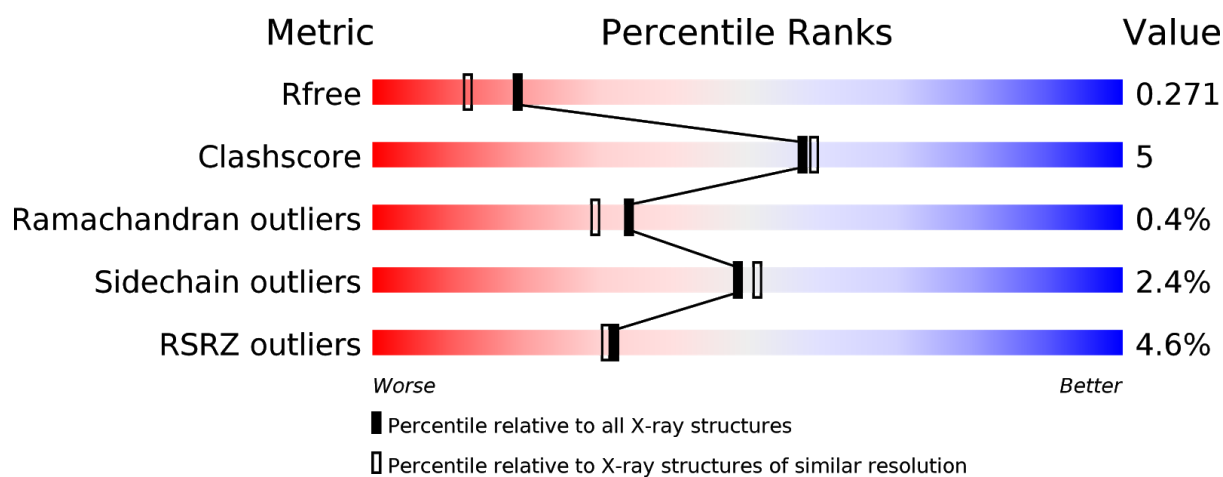
В статье [3] описан следующий подход к решению фазовой проблемы: применяли метод молекулярного замещения с помощью программы Phaser. Гомологичная NanA модель была получена программой MODELLER. Суть метода MODELLER заключается в поиске наиболее оптимальной для предложенной последовательности структуры на основании выравнивания предполагаемых гомологичных структур [3,6].

В файле структурных факторов указано 66156 измеренных рефлексов, из которых 64851 уникальные. Диапазон разрешений составляет 50 - 2 Å, соответственно, заявлено разрешение в 2,00 Å при полноте 97,9%. Дополнительно указана полнота для участка наилучшего разрешения 2,00 - 2,07 Å: 86,4%. Для улучшения модели использовались 63172 рефлекса, остальные 1619 (2,5%) вошли в тестовый набор.

Параметры кристаллической ячейки задаются длинами её сторон a, b, c и углами  $\alpha$ ,  $\beta$  и  $\gamma$  между ними. В файле структурных факторов эти параметры составляют, соответственно, 165.27 Å, 49.25 Å, 125.26 Å, 90.00°, 104.84°, 90.00°. Данной структуре соответствует кристаллографическая группа C 1 2 1. В файле со структурными факторами отсутствуют поля "struct\_ncs\_oper", значит, данных о некристаллографических симметриях нет.

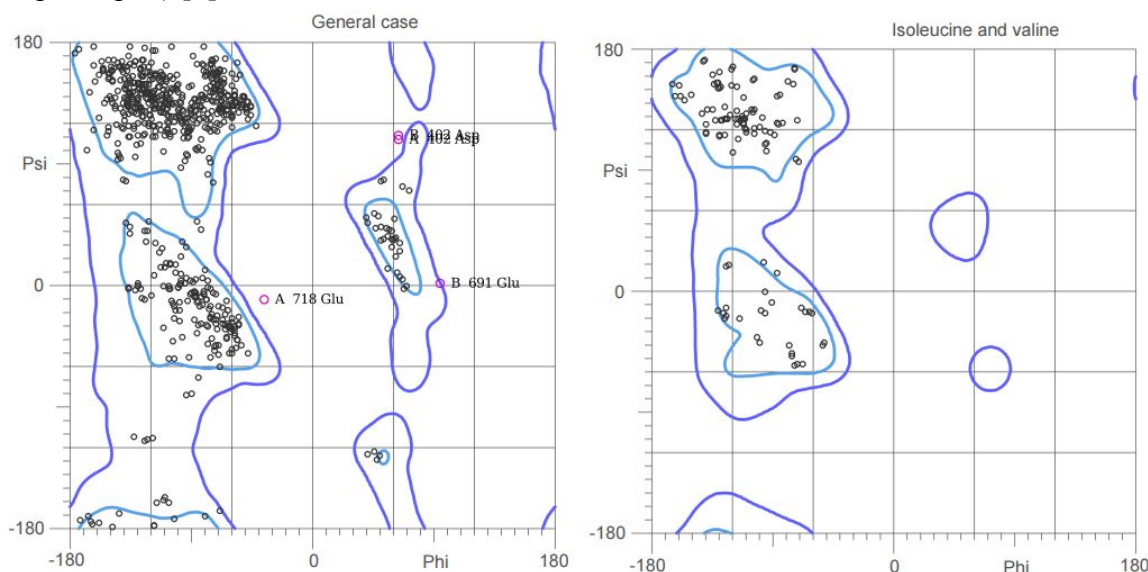
## Значения индикаторов качества модели в целом

Рассмотрим различные критерии качества модели 2YA5. Тестовые рефлексывыбирались из основного множества случайным образом. Значения R-фактора и R-free составили 0,208 и 0,271, соответственно. Вероятно, различия такие различия вызваны оверфиттингом данной модели. Примечательно, что по показателю R-free данная структура значительно хуже аналогичных и в целом по банку белков (рисунок 1).



**Рисунок 1.** Диаграмма валидации структуры сервисом wwPDB, представленная на странице модели [4].

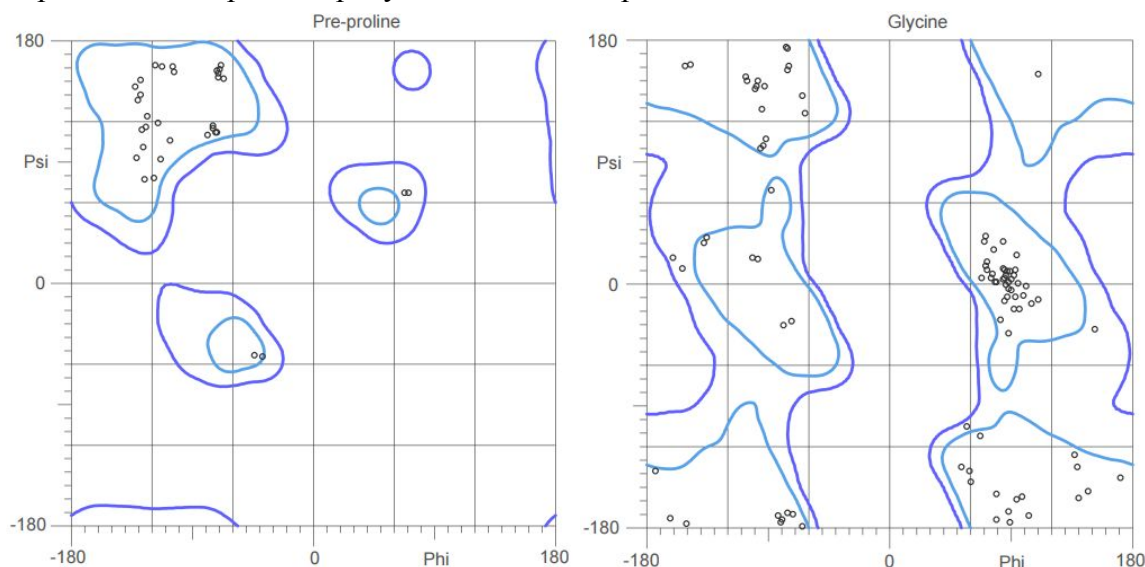
На рисунках 2-4 представлены карты Рамачандрана, полученные с помощью сервиса MolProbity (далее идёт обсуждение выдачи этого сервиса и для других параметров) [7].



**Рисунок 2.** Карты Рамачандрана для большинства остатков (кроме глицина, пролина и препролиновых) - слева и отдельно для изолейцина и валина - справа.

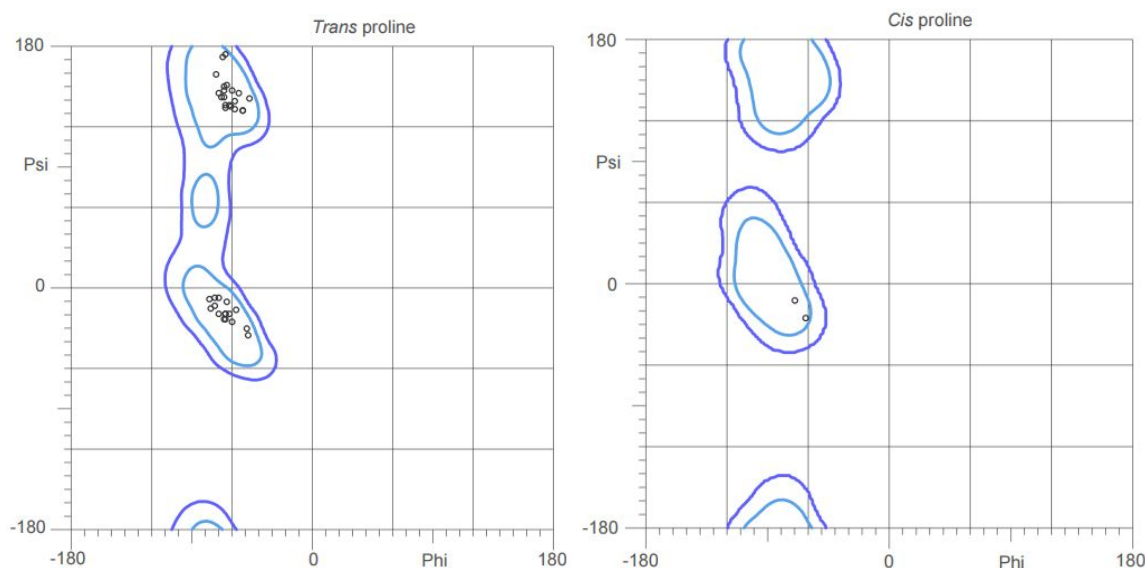
На карте (рисунок 2, слева) представлены 4 резко отличающихся значения: Glu718 (цепь A), Glu691 (цепь B), Asp402 (обе цепи). Поскольку двое выбросов не

совпадают между цепями, можно предположить, что эти остатки оказались аутлаерами по причине не совсем качественной расшифровки электронной плотности, а вот Asp402 вызывает дополнительный интерес. Данная идея будет проверена позднее. Всего в структуре 0,43% аутлаеров, что по данным MolProbity не является критичным. Карты Рамачандрана на рисунках 3-4 не содержат данных об особых остатках



**Рисунок 3.** Карты Рамачандрана для препролиновых остатков (слева) и глицина (справа).

На рисунке 4 справа отмечены 2 цис-пролина, по 1 на каждую цепь, что соответствует “нормальности” ( $\leq 1$  остатка на 1 цепь), но немного больше, чем 5% (5,26%), согласно критериям сервиса MolProbity. А вот цис-не-пролинов в структуре 3, т.е. (0,33 %). Согласно критериям сервиса, хорошо, если их меньше 0,05%.



**Рисунок 4.** Карты Рамачандрана для транс-пролинов (слева) и цис-пролинов (справа).

Также в структуре насчитывается 3 остатка с отклонением C $\beta$  от своего идеального положения более, чем на 0.25Å. Из хорошего про структуру можно сказать,

что MolProbity не обнаружил нетипичных длин и углов связей (а вот согласно отчёту проверки wwPDB есть некоторые нетипичные углы и длины связей у лиганда SIA). Также clashscore (число стерических перекрываний атомов, более чем на 0.4 Å, на 1000 атомов) для структуры составляет 5,16, что соответствует 97-му перцентиллю среди структур, то есть это очень хороший показатель. По общей оценке MolProbity score, включающей в себя оценки стерических перекрываний, качества ротамеров и выбросов карт Рамачандрана, структура 2YA5 показала себя лучше 81% структур.

### Список маргинальных остатков

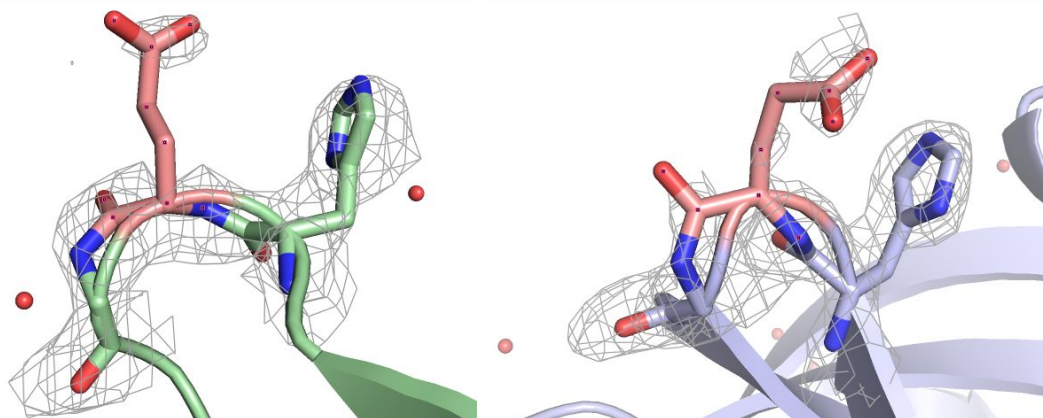
Далее в таблице 1 представлен список выявленных в структуре 2YA5 маргинальных остатков по различным критериям. Информация обо всех факторах бралась из [7], кроме RSR-факторов: для них был использован сервер EDS [8].

**Таблица 1.** Список маргинальных остатков (остатки с \* будут рассмотрены более подробно).

№	Остаток	Цепь	Причина маргинальности
1	Glu718	A	Запрещенная область карты Рамачандрана, неблагоприятная конформация боковой цепи
2	Glu691*	B	Запрещенная область карты Рамачандрана, неблагоприятная конформация боковой цепи, отклонение $C\beta=0.34 \text{ \AA} > 0,25 \text{ \AA}$
3	Asp402*	A, B	Запрещенная область карты Рамачандрана
4	Tyr762*	A, B	Цис-конфигурация
5	Gly746	B	Цис-конфигурация
6	Gly692*	B	Скручивание пептидной цепи не у пролина
7	Lys693	B	Неблагоприятная конформация боковой цепи
8	Val716	A, B	Неблагоприятная конформация боковой цепи
9	Thr631	A	Отклонение $C\beta=0.32 \text{ \AA} > 0,25 \text{ \AA}$
10	Asp497*	A, B	RSR-Z > 2
11	Lys677	A	RSR-Z > 2
12	Asn708	A, B	Подозрение на инверсию бокового радикала
13	His712*	B	Подозрение на инверсию бокового радикала
14	His727*	B	Подозрение на инверсию бокового радикала

## Анализ маргинальных остатков

Далее рассмотрим некоторые маргинальные остатки из перечисленных в таблице 1. На рисунке 5 представлена пара остатков Glu691 из цепей А (для референса) и В (в данной цепи остаток является маргиналом).



**Рисунок 5.** Глутамат 691 (показан розовым) с соседними остатками №№690 и 692 с плотностью вокруг на уровне подрезки 1 и  $\sigma_{cut}=2$  в цепи А (слева) и В (справа).

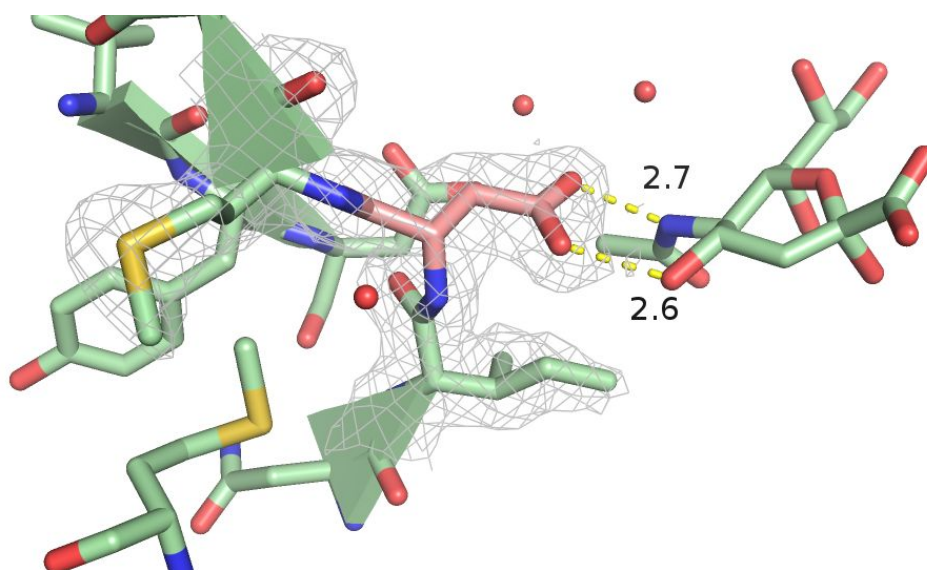
Ранее остаток Glu691 из цепи В был описан как маргинал по причине нахождения в запрещенной области на карте Рамачандрана, неблагоприятной конформации боковой цепи и значительного отклонения  $C\beta$  от оптимального положения. Действительно, это и наблюдается на рисунке справа. В цепи А остаток выглядит совершенно естественно. Заметно, что остов цепи В вылезает за пределы зоны наибольшей электронной плотности, при этом в окружении Glu691 нет групп, с которыми он мог бы вступать во взаимодействия и которые бы объясняли его положение. Можно сделать вывод, что данный остаток некачественно вписан в свою электронную плотность. Из-за неправильного положения глутамата также наблюдается скручивание пептидной связи с 692-ым глицином.

На рисунке 6 изображен аспартат 402, указанный как аутлаер карт Рамачандрана для обеих цепей (преимущественно за счёт увеличения угла  $\psi$ ). Можно заметить, что данный остаток хорошо вписан в свою электронную плотность, которая на уровне подрезки 1 хорошо окружает как остов, так и боковой радикал, и выглядит при этом округло, то есть достаточно подробно. Описанный остаток может являться маргиналом по причине своей биологической значимости для фермента: остаток находится в активном центре и может участвовать в связывании сиаловой кислоты. Такое положение объясняется более удобным расположением для связывания субстрата.

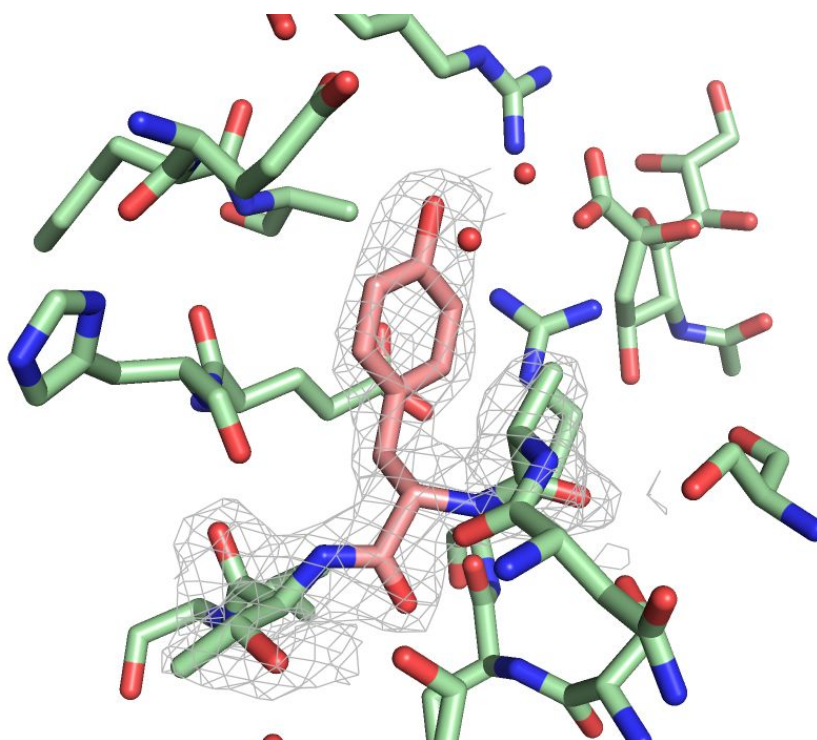
На рисунке 7 показан тирозин 762, который является маргиналом по причине цис-конфигурации. Данное положение справедливо для обеих цепей, и оно хорошо описывается электронной плотностью. Трудно сказать, по какой причине тирозин



находится именно в этом положении, поскольку ему не с чем вступать ни в какие взаимодействия среди соседних остатков, поскольку они расположены слишком далеко. Расстояние до молекул растворителя от -ОН группы тирозина позволяет образовывать с ними водородную связь, но это скорее объясняет положение молекулы растворителя в модели, а не остатка тирозина.



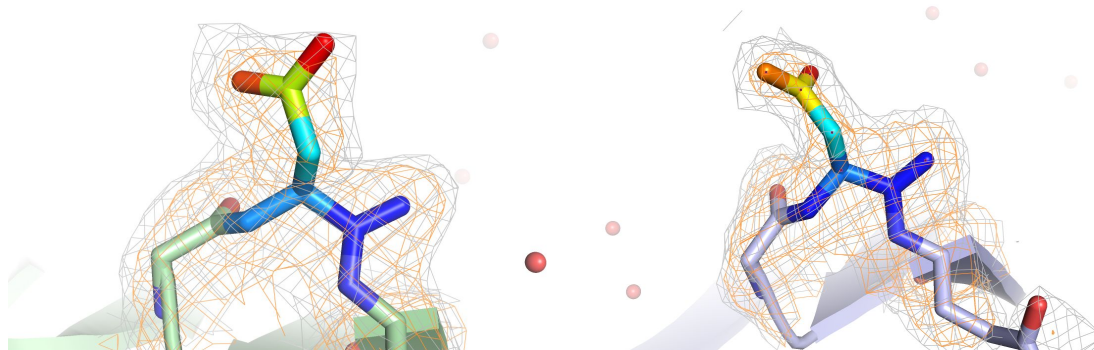
**Рисунок 6.** Аспартат 402 (показан розовым) с окружением в цепи А. Показаны расстояния от боковой карбоксильной группы аспартата до N5 и O4 сиаловой кислоты. Плотность вокруг аспартата и остатков №№401 и 403 показана на уровне подрезки 1 с  $\text{carve}=2$ .



**Рисунок 7.** Тирозин 762 (показан розовым) с окружением на расстоянии 6 Å. Дополнительно показан лиганд, однако он не входит в указанное окружение остатка.

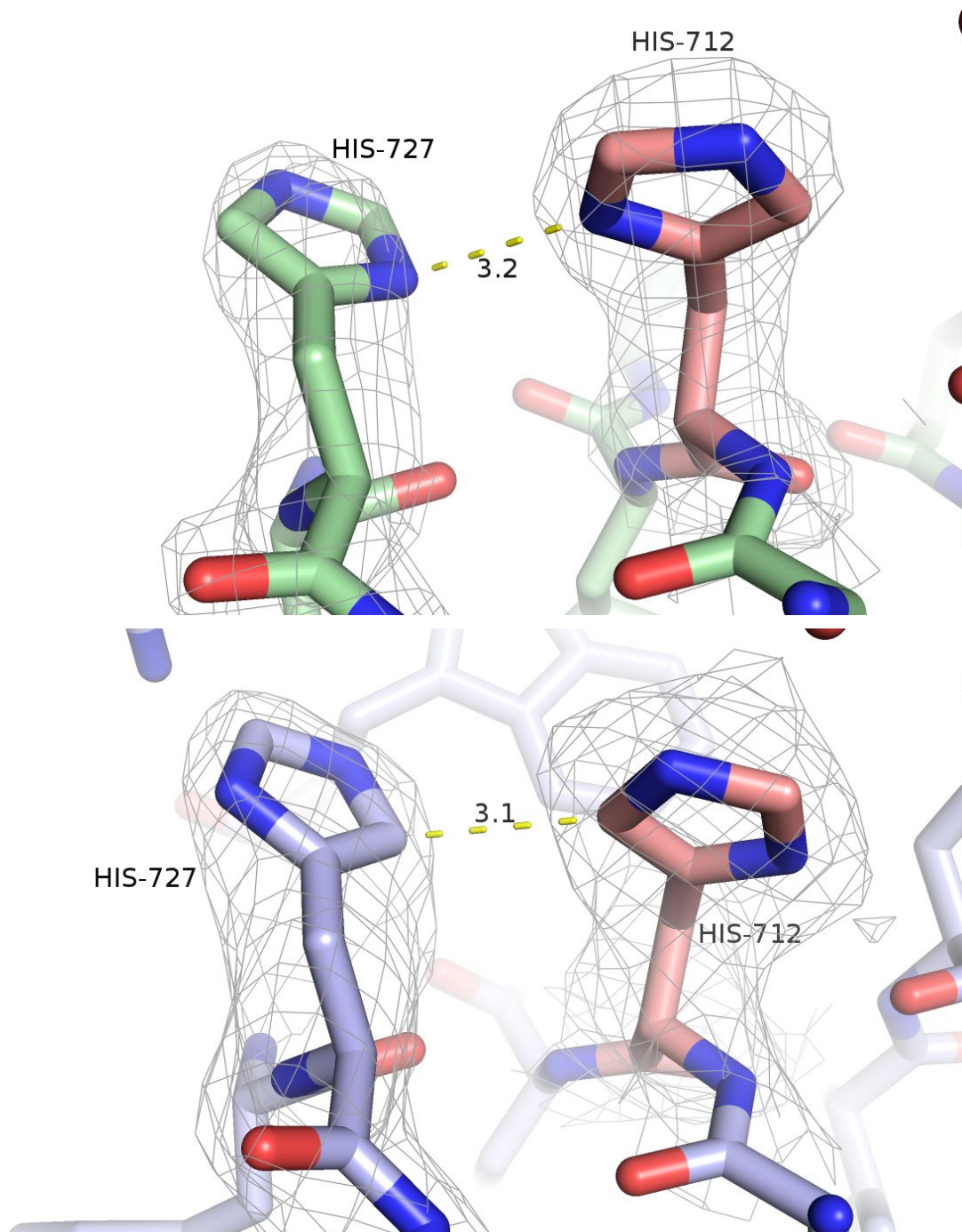
Плотность вокруг тирозина и остатков №№761 и 763 показана на уровне подрезки 1 с  $carve=2$ .

На рисунке 8 показан аспартат, который более, чем на 2 сигма отклоняется от среднего RSR для разрешения 2.00 - 2.20 ангстрем. Действительно, качество электронной плотности вокруг остатка оставляет желать лучшего для обеих цепей. У бокового радикала значительно увеличивается величина В-фактора по мере удаления от остова, что вполне соответствует наблюдаемой картине с качеством плотности. Остаток экспонирован в раствор и не вступает ни в какие нековалентные взаимодействия с соседними остатками. При этом, несмотря на большой RSR-фактор, данный остаток по другим критериям не является маргинальным. С другой стороны, поскольку плотность довольно “смазана”, а “мешающих” соседей нет, авторы структуры и могли вписать аспартат так, чтобы его параметры соответствовали приемлемым значениям. В любом случае, на качество структуры с точки зрения биологии высокий RSR-фактор остатка влияет незначительно.



**Рисунок 8.** Аспартат 497 окрашен по В-фактору для цепей А (слева) и В (справа). Показана электронная плотность с параметром  $carve=2$  на уровнях подрезки 1 (оранжевым) и 0.5 (серым).

Сервис MolProbity [7] предложил сделать разворот боковой группы для нескольких остатков при добавлении в модель водородов. Рассмотрим такую пару гистидинов 712 и 727 (рисунок), которые сервер предложил развернуть для цепи В. По электронной плотности нельзя достоверно сказать, как должно быть правильно ориентировано кольцо. Зато расстояния между ближайшими атомами колец чётко показывают, какое расположение более правдоподобное. В случае с цепью В расстояние между CD2 атомами составляет 3,1 Å. Радиус Ван-дер-Ваальса для углерода составляет 1,7 Å [9]. Таким образом, углероды перекрываются уже сами по себе, а ведь они еще должны иметь по одному водороду, которые однозначно не смогут находиться настолько сближенно. Что касается цепи А, здесь ближайшие атомы колец ND1, расстояние между которыми 3,2 Å. Радиус атома азота составляет 1,55 Å [9], следовательно, никаких стерических затруднений в случае такого расположения колец возникнуть не должно. Вывод: верно положение гистидинов, представленное для цепи А.



**Рисунок 9.** Гистидины 712 и 727 цепей А (сверху, показано расстояние между ND1 атомами) и В (снизу, показано расстояние CD2 атомами). Показана плотность для данных остатков на уровне подрезки 1 с параметром  $\sigma_{cut}=2$ .

### PDB и PDB-redo

Для структуры 2YA5 на сервере PDB-REDO [10] содержится в самом деле улучшенная версия рассмотренной из PDB модели. Показатели R-фактора и R-free стали соответственно 0,167 и 0,206 (ранее было 0,208 и 0,268). При этом структура стала по всем процентильным параметрам лучше, чем ранее. Было заменено 11

ротамеров, а 79 остатков были лучше вписаны в плотность. Что касается карты Рамачандрана, некоторые остатки сместились в более предпочтительные области, а вот судьба трёх изначальных аулаеров оказалась разнообразной: аспартаты 402 для обеих цепей остались в том же положении, что может подтверждать важность такого расположения для них в этом участке структуры. Глутамат 718 цепи А сместился в наиболее предпочтительную зону, а вот глутамат 691 цепи В, наоборот, ещё больше сдвинулся в сторону запрещённой области. Итак, сервер PDB-REDO значительно улучшил рассматриваемую модель практически по всем показателям.

## ЗАКЛЮЧЕНИЕ

В целом, можно утверждать, что структура 2YA5 каталитического домена NanA в комплексе с сиаловой кислотой была сделана достаточно качественно, хотя и на уровне большинства моделей с близким разрешением. Данные структуры получали для исследования действия ингибиторов на NanA и в целом её качество для целей исследования вполне соответствовало [3], как и для других исследований, например, работы, связанной с компьютерным моделированием полной структуры NanA с гибким линкером между доменами [2]. Авторы структуры не учли некоторые моменты при улучшении модели, такие, как инверсии боковых радикалов или же качество описания положения некоторых остатков плотностью. Однако часть “особенных” остатков действительно оказалась такой по биологическим причинам, как, например, аспартат 402. Отдельно стоит отметить качество проведённого самого PCA-эксперимента и полноту полученных рефлексов, которое позволило сервису PDB-REDO по имеющимся данным значительно улучшить данную структуру.

## БЛАГОДАРНОСТИ

Я бы хотел поблагодарить своего научного руководителя Суплатова Д.А. за исследование, которое я уже ранее проделывал со структурой NanA, поэтому было приятно работать с уже хорошо знакомой структурой. Также я благодарен преподавателям курса биоинформатики за возможность (и необходимость) изучить другие аспекты, казалось бы, очень подробно рассмотренного ранее объекта исследования.

## СПИСОК ЛИТЕРАТУРЫ

1. Xu Z. et al. Sequence diversity of NanA manifests in distinct enzyme kinetics and inhibitor susceptibility // *Sci. Rep.* 2016. Vol. 6. P. 25169.
2. Sharapova Y., Suplatov D., Švedas V. Neuraminidase A from *Streptococcus pneumoniae* has a modular organization of catalytic and lectin domains separated by a flexible linker // *The FEBS Journal.* 2018. Vol. 285, № 13. P. 2428–2445.
3. Gut H. et al. Structural basis for *Streptococcus pneumoniae* NanA inhibition by influenza antivirals zanamivir and oseltamivir carboxylate // *J. Mol. Biol.* 2011. Vol. 409, № 4. P. 496–503.
4. Bank R.P.D. 2YA5 [Electronic resource]. URL: <https://www.rcsb.org/structure/2YA5>
5. B2DJD9 in UniParc [Electronic resource]. URL: <https://www.uniprot.org/uniparc/?query=B2DJD9&sort=score>.
6. Sali A., Blundell T.L. Comparative protein modelling by satisfaction of spatial restraints // *J. Mol. Biol.* 1993. Vol. 234, № 3. P. 779–815.
7. <http://molprobity.biochem.duke.edu/>
8. EDS - Uppsala Electron Density Server [Electronic resource]. URL: <http://eds.bmc.uu.se/eds/>
9. Rowland R.S., Scott Rowland R., Taylor R. Intermolecular Nonbonded Contact Distances in Organic Crystal Structures: Comparison with Distances Expected from van der Waals Radii // *The Journal of Physical Chemistry.* 1996. Vol. 100, № 18. P. 7384–7391.
10. <https://pdb-redo.eu/db/2ya5>