

МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ИМЕНИ М.В. ЛОМОНОСОВА ФАКУЛЬТЕТ  
БИОИНЖЕНЕРИИ И БИОИНФОРМАТИКИ

ОТЧЕТ ПО КАЧЕСТВУ РАСШИФРОВКИ СТРУКТУРЫ 4С4М,  
ПОЛУЧЕННОЙ МЕТОДОМ RSA

Работу выполнила  
студентка 4 курса  
ФББ МГУ им.Ломоносова  
Корягина Алена

Москва, 2016

## Аннотация

Данный отчет представляет собой анализ качества расшифровки структуры 4с4т на основе определенных показателей, таких как параметры кристаллической решетки, R- фактор, R\_free, карта Рамачандрана, и др., которые могут быть получены с помощью различных сервисов.

## Оглавление

|  |    |
|--|----|
| Аннотация                                  | 2  |
| Оглавление                                 | 2  |
| Введение                                   | 3  |
| Результаты и обсуждения                    | 3  |
| Общая информация о модели                  | 3  |
| Индикаторы качества модели                 | 4  |
| Примеры маргинальных остатков              | 6  |
| Анализ маргинальных остатков               | 6  |
| Сравнение модели из PDB с моделью PDB_redo | 10 |
| Заключение                                 | 11 |
| Список литературы и сервисов               | 12 |

## Введение

«Сверхзвуковой ежик» (sonic hedgehog, Shh) — один из трех белков семейства hedgehog (Hh-семейство), названный так в честь героя популярной видеоигры «Sonic the Hedgehog», вышедшей в 1990-е годы. Впервые этот белок был открыт у плодовой мушки *Drosophila melanogaster* в 1980 году [1].

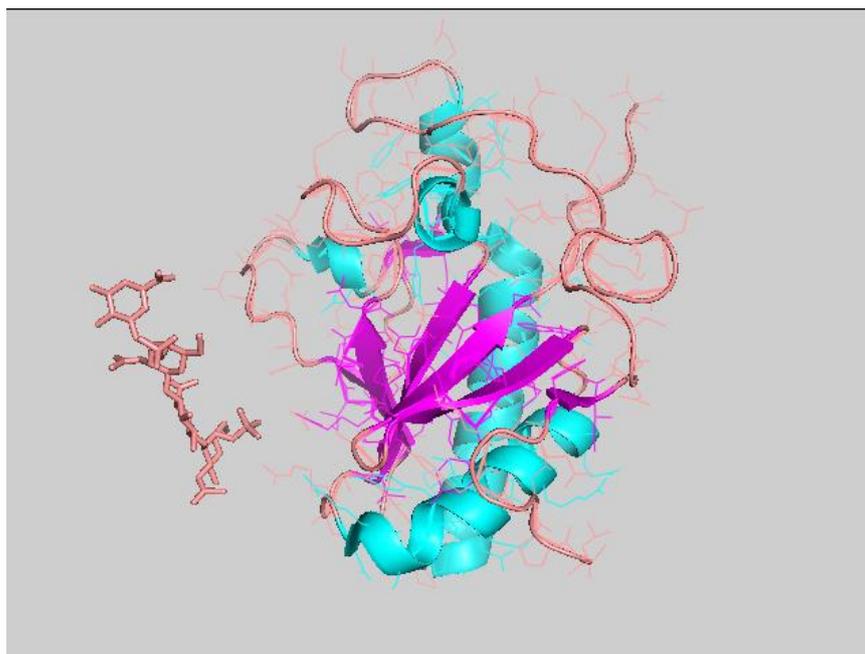
Белки этого семейства принимают участие в Hh-сигнальном пути (hedgehog signaling), который отвечает за дифференцировку клеток во время эмбрионального развития [2]. Во взрослом организме этот путь также имеет место быть и различные дисфункции Hh-пути приводят к образованию различных дефектов развития и болезням, в том числе к раковым заболеваниям. Именно Shh у млекопитающих играет ключевую роль в органогенезе и отвечает за развитие пальцев на конечностях и организацию мозга [3].

Существует много различных клеточных рецепторов, регулирующих работу Hh-сигнального пути. Среди них есть гликозаминогликановые (ГАГ) цепи протеогликанов, которые формируют градиент Hh-белков и осуществляют сигнальную трансдукцию. Ранее не был определен сайт связывания Hh с ГАГ, и целью расшифровки этой структуры было именно определение этого сайта.

## Результаты и обсуждение

### Общая информация о модели

Структура 4с4m представляет собой комплекс белка SHH с хондроитин-4-сульфатом (рис.1). Она была получена в 2013 году методом рентгеноструктурного анализа (РСА). Авторами являются Daniel M. Whalen, Tomas Malinauskas, Robert J. C. Gilbert и and Christian Siebold, из отдела структурной биологии Оксфордского Университета, Великобритания.



**Рис.1. Структура белка sonic hedgehog (4с4m).**  
Изображение получено с помощью программы rutil.

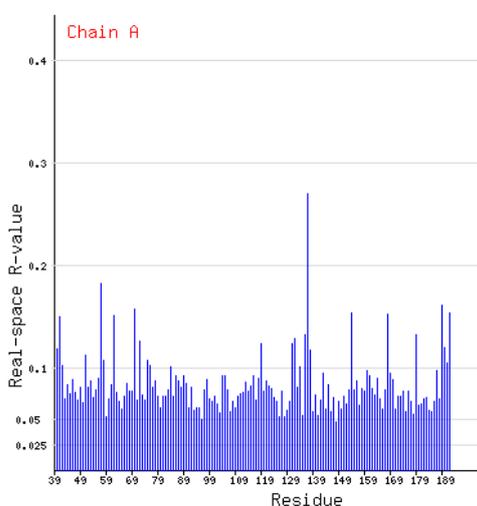
Некоторые характеристики модели структуры 4с4m, полученные из файла структурных факторов, а также с сайта ESD [4], приведены в таблице 1.

**Таблица 1.** Основные характеристики модели структуры 4c4m

| Параметр                       | Значение   |             |
|--------------------------------|--|-------------|
| Метод решения фазовой проблемы | Молекулярное замещение, используя структуру 1VNH |             |
| Число измеренных рефлексов     | 16 993   |             |
| Разрешение                     | 1,74 Å   |             |
| Диапазон разрешений            | 43,84-1,74 Å                                     |             |
| Полнота данных                 | 99,9 %   |             |
| Кристаллографическая группа    | P 21 21 21                                       |             |
| R-фактор                       | 0,162  |             |
| Свободный R-фактор             | 0,193  |             |
| Параметры ячейки               | Стороны (Å)                                      | Углы (°)    |
|                                | a= 40,35   | $\alpha=90$ |
|                                | b=55,66  | $\beta=90$  |
|                                | c= 71,10   | $\gamma=90$ |

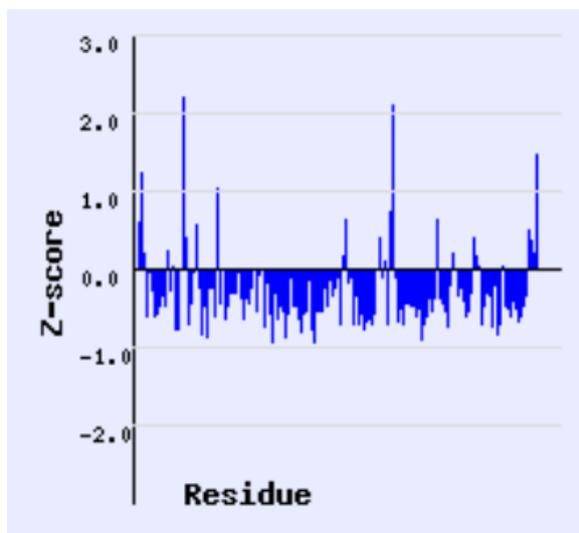
### Индикаторы качества модели

R-фактор и свободный R-фактор являются одними из главных индикаторов качества модели в целом. Согласно эмпирическому правилу, R-фактор хорошей модели не должен превышать значение разрешение/10. В полученной модели R-фактор равен 0,162, что меньше 0,174. Также  $(R - R\text{-free}) = 0,031$  ( $<0,1$ ), что также указывает на то, что структура достаточно хороша для своего разрешения. Помимо этого, существует параметр RSR (real-space R-value), значения которого выше 0,2 говорят о плохом соответствии экспериментальной ЭП модельной ЭП. Для полученной структуры  $RSR = 0,088$ , такое значение указывает на хорошее соответствие экспериментальной модельной ЭП. С помощью сервера EDS [4] был визуализирован RSR для всех остатков исследуемого белка (рис.2). Видно, что для одного остатка (Glu 37) значение RSR превышает 0,2.

**Рис.2.** Значения RSR для всех остатков структуры 4c4m

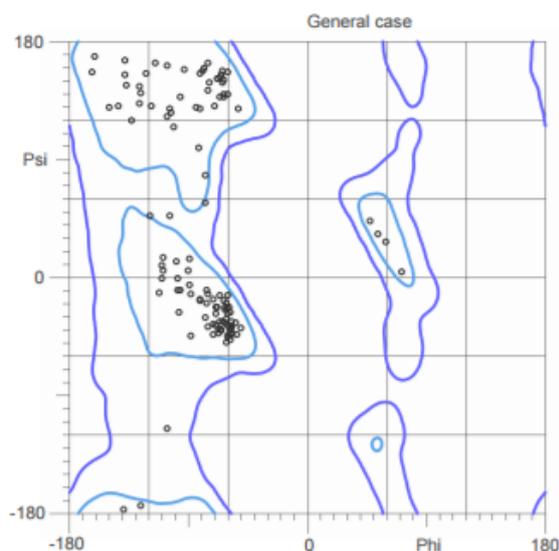
Также была проверена еще одна оценка качества расшифровки структуры – RSR-Z [4]. Это относительная оценка RSR для отдельных остатков. Для вычисления Z для какого-либо остатка проводится сравнение его RSR со средним RSR для того же типа остатков по выборке из PDB с примерно таким же разрешением структуры. Высокие положительные значения Z ( $> 2$ ) для остатка

свидетельствуют о том, что он плохо вписан в электронную плотность по сравнению с другими структурами с тем же разрешением, поэтому такой остаток можно рассматривать как маргинальный. Для структуры 4с4т было определено 2 остатка (Leu 37, Glu 137) для которых RSR-Z принимает значение больше 2 (2,188 и 2,103, соответственно). Визуализация RSR-Z представлена на рисунке 3.



**Рис.3. Значения RSR-Z для всех остатков структуры 4с4т**

С помощью сервиса MolProbity [5] к структуре были добавлены водороды, определены стерически-затрудненные боковые цепи Asn, Gln, His. После редакции всех инверсий радикалов, был проведен полный анализ взаимодействий атомов в структуре. Clashscore структуры составил 1,19, что является 99% персентилью среди моделей с уровнем разрешения хуже 1,74 Å. Это говорит о том, что структура хорошая для этого разрешения. Полных маргиналов не было обнаружено в данной структуре. Всего было найдено два остатка с маргинальными по отклонению от ротаметров боковыми цепями, что составляет 1,53% от всех остатков. В предпочитаемых областях карт Рамачандрана находится 98,01% всех остатков (148 остатков), еще три остатка находятся в позволенной области. Также было определено, что 4 ковалентных связи из 1260 существенно отклоняются от теоретических, три из которых находятся в структуре лиганда. На рисунке 4 приведены карты Рамачандрана для исследуемой структуры.



**Рис.4. Карта Рамачандрана для структуры белка 4с4т.**

Затем с помощью сайта WHAT\_CHECK [6] опцией Fine Packing Quality Control была оценена комфортность окружения для всех остатков. Не найдено ни одного остатка с параметром меньше -3, что говорит о том, что в структуре нет маргиналов по окружению, а также нет даже сомнительных остатков, то есть модель является хорошей по окружению. Среднее значение Z-score = -1,04, что соответствует хорошей структуре. Все оценочные суждения основаны на инструкции в конце выдачи программы

### Примеры маргинальных остатков

В таблице 2 приведены 12 примеров остатков, которые можно считать маргинальными. Маргинальность остатков предположена на основании выдачи программ EDS, MolProbity и WHAT\_CHECK.

**Таблица 2.** Маргинальные остатки в структуре белка sonic hedgehog

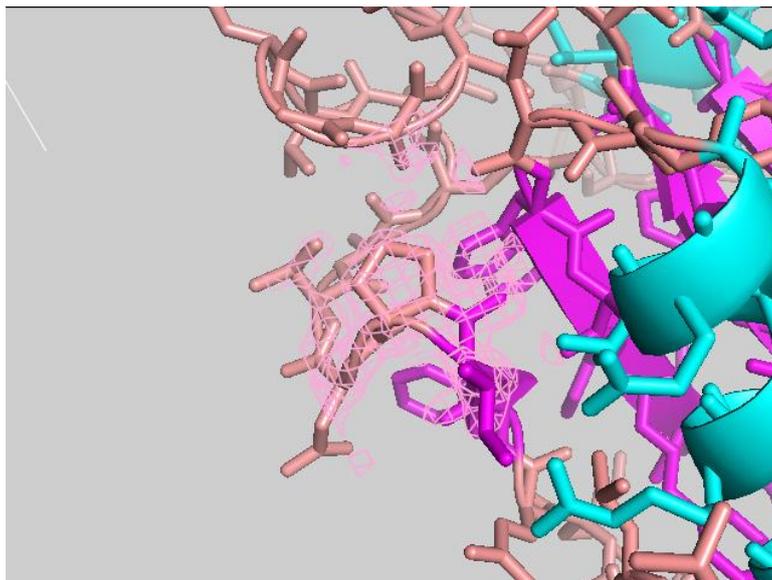
| #  | Остаток | Причина маргинальности   |
|----|---------|--|
| 1  | Leu 40  | Перекрытие 0.57Å N с A2001 НОН О   |
| 2  | Pro 50  | Остаток в цис положении<br>Cis PRO omega= -3.06  |
| 3  | Leu 57  | Высокое значение RSR-Z 2.188   |
| 4  | Als 59  | Находится в допустимой, а не предпочитаемой области на карте Рамачандрана (0.27%) углы phi и psi: -107.9,-115.9  |
| 5  | Ser 60  | Перекрытие 0.49Å<br>О с A 145 ARG NH2  |
| 6  | Asn 70  | Находится в допустимой, а не предпочитаемой области на карте Рамачандрана (0.82%) углы phi и psi: -105.8, 48.9   |
| 7  | Thr 78  | Отклонение от ротамеров боковыми цепями<br>Allowed (1.7%) t chi angles: 173.3  |
| 8  | Glu 137 | Высокое значение RSR 0,27<br>Высокое значение RSR-Z 2,103  |
| 9  | Glu 138 | Перекрытие 0.66Å<br>OE1 с A2022 НОН О<br>И находится в допустимой, а не предпочитаемой области на карте Рамачандрана (1.54%) углы phi и psi: -120.8,48.9 |
| 10 | Arg 145 | Перекрытие 0.49Å<br>NH2 с A 60 SER О   |
| 11 | Tyr 176 | Отклонение от ротамеров боковыми цепями<br>Allowed (0.8%) t80 chi angles: 192.7,336  |
| 12 | Val 192 | Выброс по длине связи C--N: 6.8 σ  |

### Анализ маргинальных остатков

Далее был проведен более подробный анализ пяти маргинальных остатков. Выбранные остатки выделены голубым цветом в таблице 2.

- 1) Pro 50 находится в цис положении (omega= -3.06°). Этот остаток располагается в повороте после β-тяжа (рис.5) и вписан в электронную плотность хорошо. Он располагается в предпочитаемой области на карте Рамачандрана, его цис-конформация не является удивительной, потому что в поворотах он часто располагается именно в цис-конформации, так как именно цис-пролин позволяет осуществить быстрый изгиб пептидной цепи. Остаток Pro 50 следует считать маргинальным, его конформация

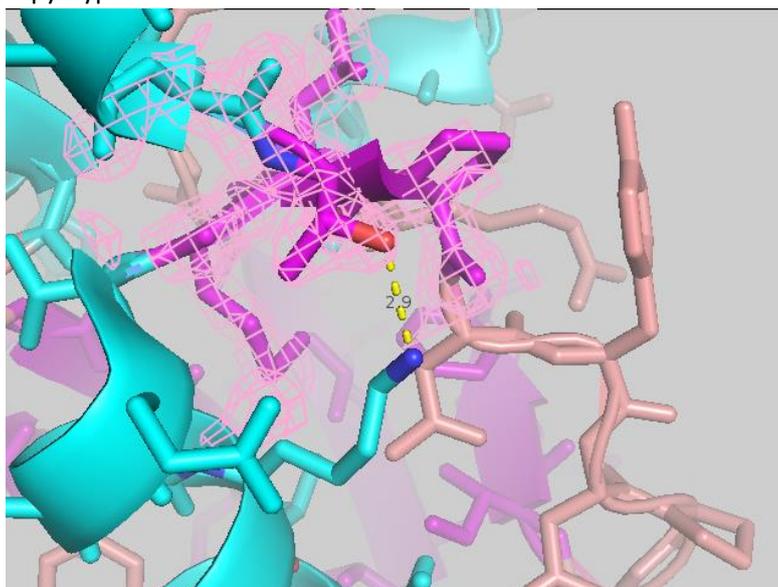
является особенностью элемента структуры – поворота.



**Рис.5. Pro 50 из структуры 4c4m.**

Изображение получено с помощью программы rutil. Уровень подрезки электронной плотности 1,5  $\sigma$ .

- 2) Thr 78 отклоняется от конформации боковыми цепями: плохой угол  $\chi_1$  173.3°. Треонин хорошо вписан в электронную плотность. В данном случае этот остаток является маргинальным, он образует дополнительную водородную связь с остатком лизина (Lys 104) благодаря изменению угла  $\chi_1$ . Эта водородная связь стабилизирует такое расположение боковой цепи треонина (рис.6). Отклонение боковой цепи Thr 78 от конформации можно считать особенностью структуры.

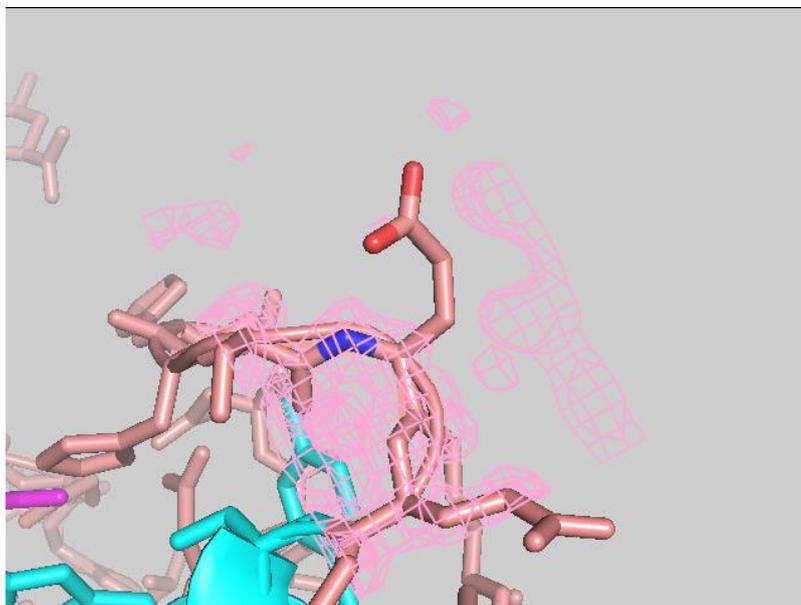


**Рис.6. Thr 78 из структуры 4c4m.**

Изображение получено с помощью программы rutil. Уровень подрезки электронной плотности 1,5 $\sigma$ .

- 3) Glu 137 имеет высокие значения RSR и RSR-Z, 0,27 и 2,103 соответственно. Высокие значения этих параметров говорят о том, что этот остаток плохо вписан в экспериментальную электронную плотность. При визуализации электронной плотности на уровне подрезки 1 $\sigma$  вокруг этого остатка (рис.7) действительно видно, что боковая цепь

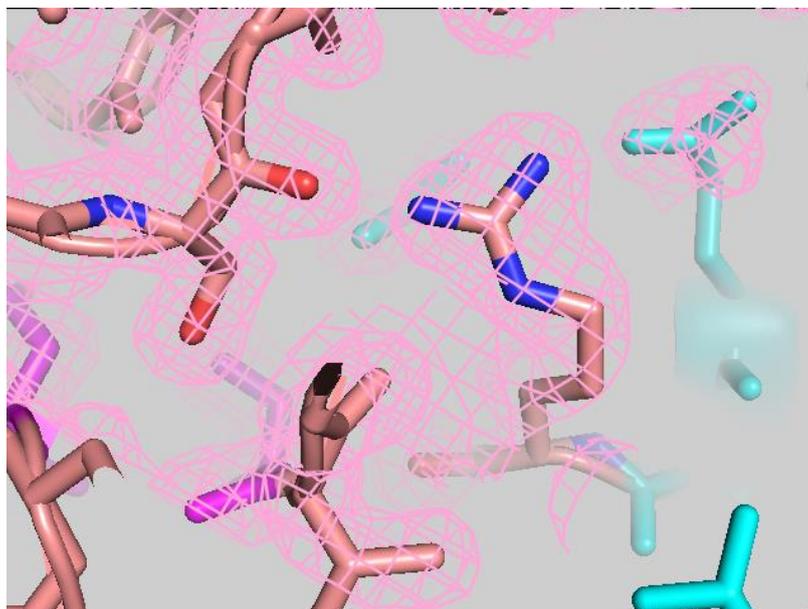
абсолютно не вписана в электронную плотность. Электронного облака для боковой цепи Glu 137 вообще нет, что очень странно, потому что глутамат имеет COO-группу, кислороды которой электроотрицательны. Отсутствие электронного облака для боковой группы глутамата является странным, а учитывая тот факт, что разрешение структуры считается хорошим (1,74 Å), это ставит под сомнение тот факт, что этот аминокислотный остаток является глутаматом. Я бы предположила, что на этом месте скорее всего расположился бы глицин. Определенно остаток Glu 137 можно считать маргинальным из-за ошибки расшифровки.



**Рис.7. Glu 137 из структуры 4c4m.**

Изображение получено с помощью программы rutil. Уровень подрезки электронной плотности 1 $\sigma$ .

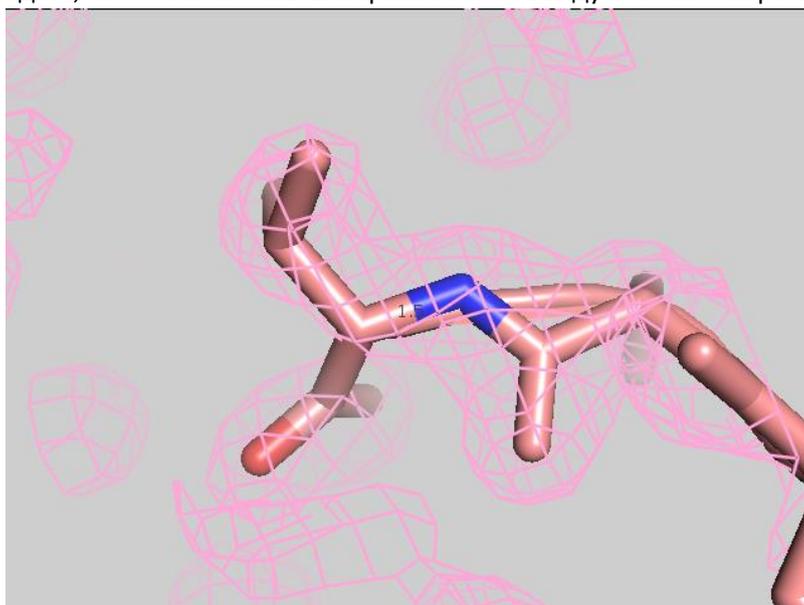
- 4) Arg 145 имеет перекрытие 0.49Å между своей NH<sub>2</sub>-группой и OH-группой серина 60. При визуализации электронной плотности вокруг остатков Arg 145 и Ser 60 на уровнях подрезки 2, 1, 0.5 я не заметила перекрытие электронных облаков от NH<sub>2</sub>-группы аргинина и OH-группы серина. Возможно, это связано с тем, электронная облако не покрывает атомы водорода, потому что он имеет всего 1 электрон. И поэтому после восстановления атомов водорода и дальнейшего анализа структуры сервисом MolProbity, это сервис определил наличие перекрытия атомов водорода от NH<sub>2</sub>-группы аргинина и OH-группы серина. Я считаю, что оба остатка можно считать маргинальными из-за неточности расшифровки (не учтено расположение атомов водорода).



**Рис.8. Arg 145 и Ser 60 из структуры 4c4m.**

Изображение получено с помощью программы rutil. Уровень подрезки электронной плотности 1 $\sigma$ .

- 5) Val 192 имеет связь C—N, длина которой отличается от теоретической. Валин 192 является C-терминальным остатком в структуре белка sonic hedgehog. Действительно связь C—N больше теоретической пептидной связи (1,33 Å) и составляет 1,5 Å. Это связано с неточностью расшифровки структуры, потому что можно заметить, что, если немного укоротить эту связь, остаток будет лучше вписан в электронную плотность. Можно утверждать, что остаток Val 192 маргинальный ввиду неточности расшифровки.



**Рис.9. Val 192 из структуры 4c4m.**

Изображение получено с помощью программы rutil. Уровень подрезки электронной плотности 1 $\sigma$ .

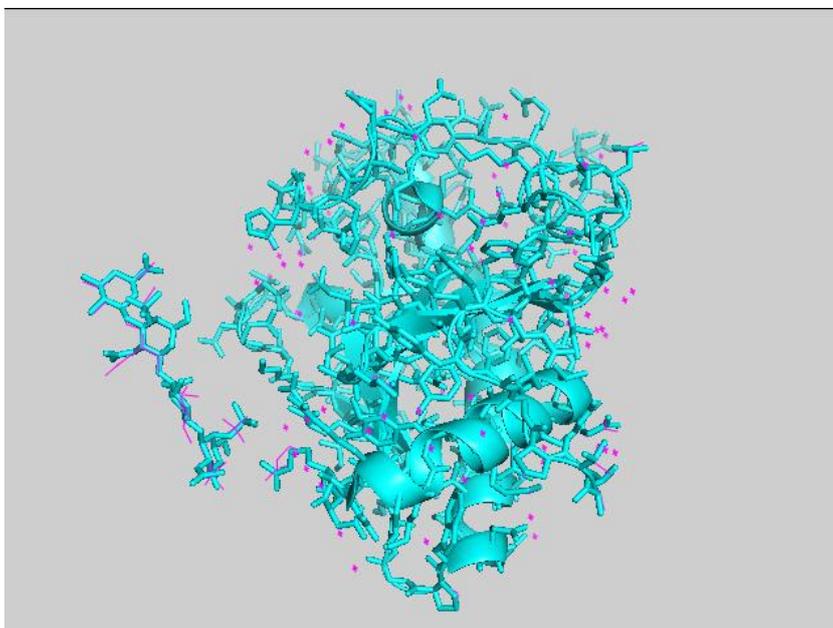
## Сравнение модели из PDB с моделью PDB\_redo

PDB\_redo [7] позволяет оптимизировать модель, используя различные параметры уточнения, выбирая между изо- и анизотропным B-фактором, перестраивая конформации ротамеров, вращая боковые цепи остатков для оптимизации сети водородных связей, пересматривая молекулы воды и конформации лигандов.

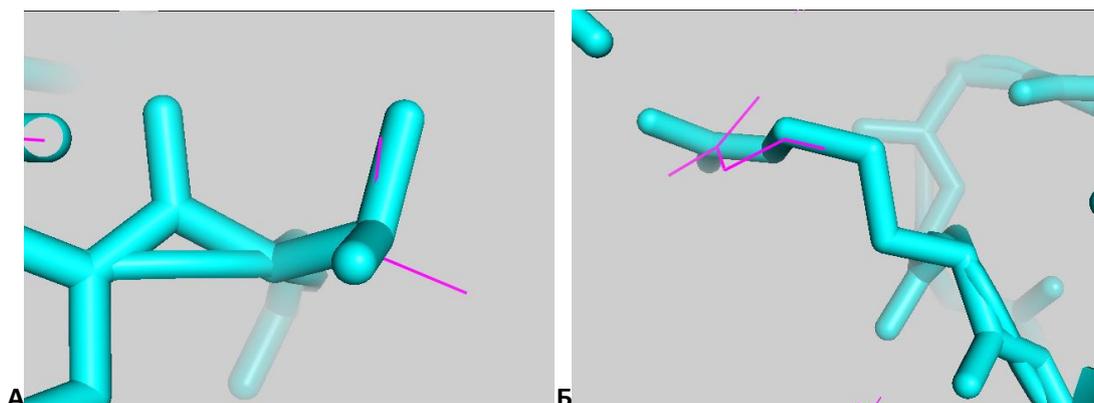
В результате оптимизации структуры 4с4т были улучшены параметры R, R-free до 0,1566 и 0,1859, а также RSR для 16 остатков и RSR-Z для 18. Но также произошло ухудшение RSR для 5 остатков.

При одновременной визуализации первоначальной модели и оптимизированной модели не заметно никаких кардинальных улучшений (рис.10). Достаточно много изменений наблюдается в структуре хондроитин-4-сульфата, но я считаю их значимыми, так как оптимизация проходила так, если бы это была часть структуры белка, а это лиганд, структура которого важна именно четко восстановленная из электронной плотности. Также есть небольшие изменения в расположении боковых цепей для некоторых остатков, например, Glu 168, Leu 57, Leu 40, Lys 55 и др. Более значительные изменения для остатков Thr 41 (рис.11, А) и Arg 154 (рис.11, Б).

Но все же, я думаю, что эта оптимизация не так существенна, потому что модель 4с4т достаточно хорошо расшифрована первоначально.



**Рис.10.** Первоначальная структура 4с4т (голубой цвет) и оптимизированная PDB\_redo модель (розовый). Изображение получено с помощью программы rutil.



**Рис.11. Первоначальная структура 4c4m (голубой цвет) и оптимизированная PDB\_redo модель (розовый). А – Thr 41, Б – Lys 154.**

Изображение получено с помощью программы rutil.

## Заключение

Данная структура определена хорошо. Потому что была необходима точная расшифровка структуры для определения сайта взаимодействия лиганда с белком, что и было в результате достигнуто[3].

Такое заключение о качестве расшифровки сделано на основании проанализированных факторов: значения R, R free, RSR, RSR-Z достаточно маленькие, маргинальных остатков не много, нет полных маргиналов. Кроме того, в PDB\_redo в структуры было внесено достаточно небольшое число изменений, что также может говорить о ее качестве и об отсутствии грубых дефектов структуры. Есть некоторые неточность в расшифровке структуры, но они не являются значительными.

## Список литературы и сервисов

- [1] Nüsslein-Volhard C, Wieschaus E (October 1980). "Mutations affecting segment number and polarity in *Drosophila*". *Nature*. 287 (5785): 795–801.
- [2] Ingham, P. W.; McMahon, AP (2001). "Hedgehog signaling in animal development: Paradigms and principles". *Genes & Development*. 15 (23): 3059–87.
- [3] Whalen, Daniel M., et al. (2013) "Structural insights into proteoglycan-shaped Hedgehog signaling." *Proceedings of the National Academy of Sciences* 110.41: 16420-16425.
- [4] «EDS» [В Интернете]. Available: <http://eds.bmc.uu.se/eds/>
- [5] «MolProbity» [В Интернете]. Available: <http://molprobity.biochem.duke.edu>
- [6] «WHAT\_CHECK» [В Интернете]. Available: <http://swift.cmbi.ru.nl/servers/html/index.html>
- [7] «Выдача pdb\_redo» [В Интернете]. Available: [http://www.cmbi.ru.nl/pdb\\_redo/c4/4c4m/index.html](http://www.cmbi.ru.nl/pdb_redo/c4/4c4m/index.html)