

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

**Отчет по качеству расшифровки структуры белка альдегид
декарбонилазы (PDB ID 4quw), полученной методом
рентгеноструктурного анализа**

Пиунова Ульяна,
4 курс, 402 группа

Москва, 2019

Аннотация

В данном отчете представлен анализ качества расшифровки структуры фермента альдегид-деформирующей оксигеназы, или альдегид декарбонилазы, с PDB ID 4quw, полученной методом рентгеноструктурного анализа. Были рассмотрены параметры качества модели, а также проанализированы отдельные маргинальные остатки.

Введение

Согласно международной иерархической классификации ферментов альдегид декарбонилаза (АД) имеет номер ЕС:4.1.99.5. Первая цифра (4) относит белок к лиазам – ферментам, катализирующие разрыв С-О, С-С, С-N и других связей, а также обратимые реакции отщепления различных групп негидролитическим путем. Подкласс (1) включает углерод-углерод лиазы, а номер подподкласса (99) и порядковый номер (5) однозначно определяет белок.

Цианобактериальная АД является мономером и принадлежит суперсемейству ферритин-подобных металлопротеинов и содержит центр связывания 2х ионов железа (Fe^{2+}). АД катализирует деформирование альдегида с длинной цепью (C_n) с образованием алкана (C_{n-1}) и формиата:

$R-CHO + O_2 + 2 NADPH \rightleftharpoons R + HCOO^- + H_2O + 2 NADP^+$, где R – жирный алкан. Реакция является нейтральной с точки зрения изменения окислительно-восстановительного потенциала и кажется простым гидролизом, однако в действительности механизм более сложный. Молекула кислорода взаимодействует с лигандом, в результате чего образуется пероксид железа. Пероксид нуклеофильно атакует субстрат, и образуется полуацеталь (назван так в статье [1], хотя это скорее «металлоперекись алкила»), то есть Fe^{2+} оказывается связано через 2 кислорода с карбонильным углеродом. Затем связь между С1-С2 углеродами разрывается и образуется формиат (один атом кислорода из субстрата, второй – из O_2 ; второй атом из O_2 - в молекулу воды) и алкан. Также АД необходимы 2 молекулы NADPH, как доноры протонов.

Через вышеописанную реакцию происходит биосинтез алканов у некоторых цианобактерий. Кроме того, в промышленности фермент может быть полезным при производстве биотоплива.

АД из *Synechococcus elongatus* (штамм PCC 7942) представлена одной цепью, состоящей из 231 а.о (рис. 1). Глобула белка образована из 8 альфа-спиралей. 6 консервативных аминокислот из 4х альфа-спиралей образуют пучок, в котором расположен центр связывания ионов Fe^{2+} .

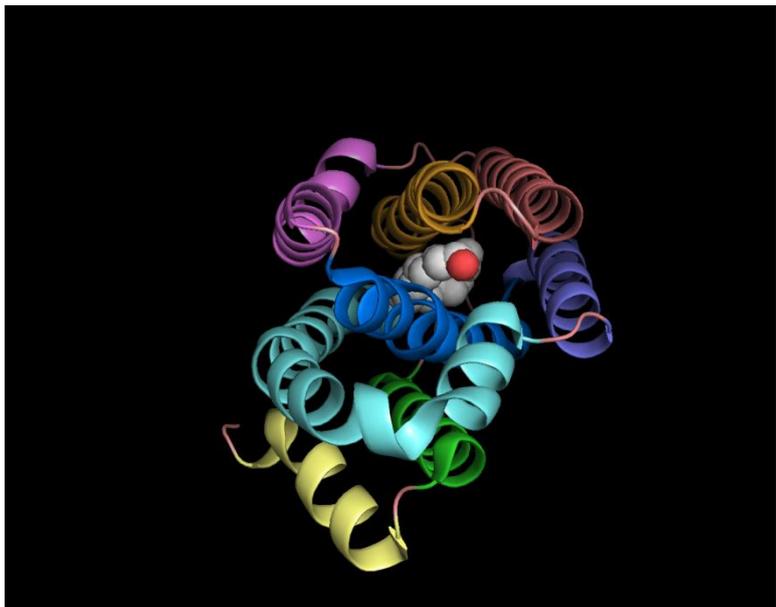
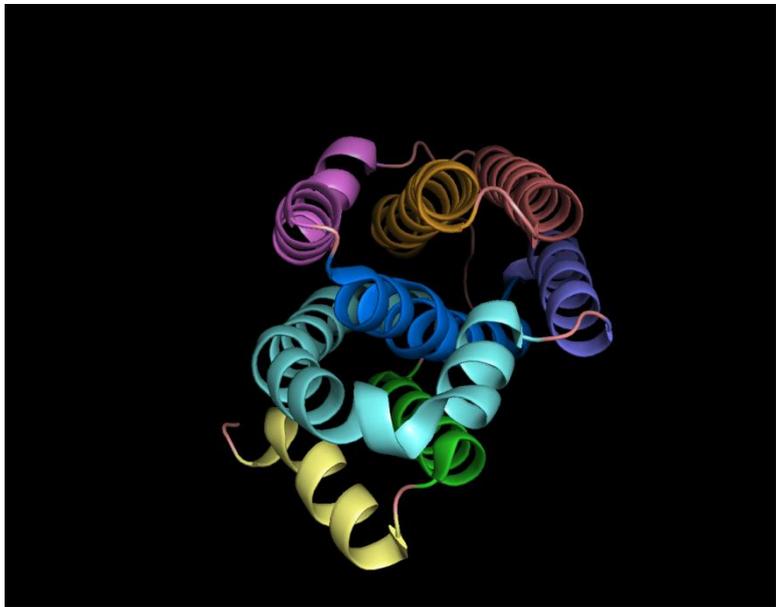
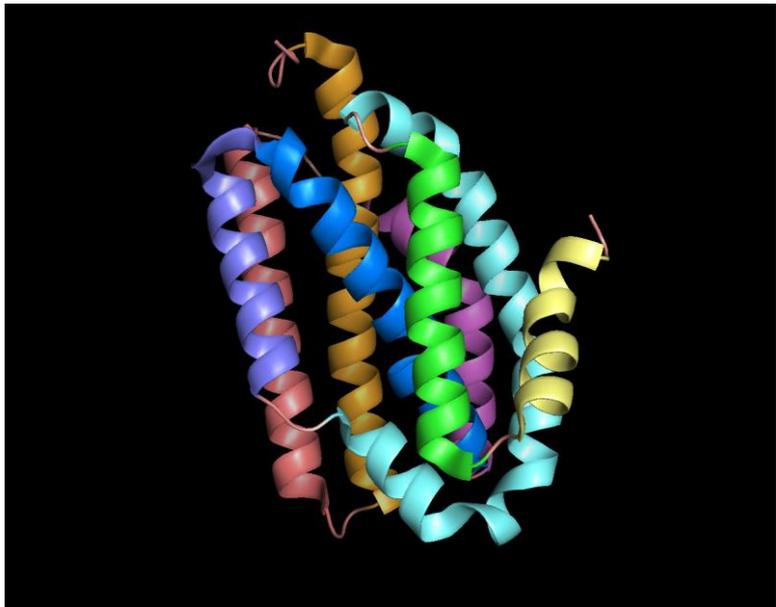


Рисунок 1. Структура альдегид декарбонилазы. Альфа-спирали выделены разными цветами. Сверху и посередине – молекула в разных ракурсах; снизу – субстрат (показан шариками) связан в гидрофобном канале белка

Гидрофобный канал, в котором субстрат удерживается и через который проходит к активному центру, образован всеми альфа-спиралями, но одна из них (обозначена желтым на рис. 1) вносит небольшой вклад.

Результаты и обсуждение

Общая информация о модели представлена в таблице №1 [2].

Таблица №1. Информация о модели	
Параметр	Описание
Состав комплекса	Макромолекулы: АД, цепь А, 231 а.о.
	Другие молекулы: C ₁₆ H ₃₄ O (гексадеканаль-1)
Год	2014
Авторы	Jia, C.J., Li, M., Chang, W.R.
Метод решения фазовой проблемы	Молекулярное замещение
Число измеренных рефлексов	10556
Разрешение	2.26 Å
Полноту набора структурных факторов	99.7 %
Диапазон разрешений структурных факторов	41.15 - 2.26 Å
Параметры кристаллографической ячейки	a=61.68 Å, b=61.68 Å, c=110.46 Å; alpha=90.00, beta=90.00, gamma=90.00
Тип кристаллографической симметрии	P 4 ₁ 2 ₁ 2
Наличие некристаллографических симметрий в асимметрической ячейке	Нет

Значения индикаторов качества модели в целом

В качестве индикаторов качество модели обычно используют R-фактор и R_{free}.

R-фактор является мерой согласованности между моделью и экспериментальными данными. Оптимизация заключается в минимизации R-фактора:

$$R = \frac{\sum |F^{obs} - F^{calc}|}{\sum |F^{obs}|}$$

F^{obs} - модули структурных факторов, полученные в эксперименте, F^{calc} - модули структурных факторов, посчитанные по построенной модели, \sum - суммирование по всем (h, k, l).

Для данной модели R-фактор = 0.227 по аннотации авторов и R-фактор = 0.225 по данным программы DCC. В обоих случаях значения факторов меньше 0.25, а значит модель можно считать хорошей.

R_{free} рассчитывают аналогично R-фактору, но на небольшом количестве (по 505 рефлексам, 4.81%, в данной модели) случайно выбранных рефлексов, не участвовавших в оптимизации. Таким образом производится контроль переоптимизации. Для данной модели $R_{\text{free}} = 0.266$ по аннотации авторов и $R_{\text{free}} = 0.261$ по данным программы DCC. Эти значения выше 0.25, но не выше 0.4, а значит их можно считать достаточно хорошими. Более того, значения $(R_{\text{free}} - R) \ll 0.1$, что также характеризует модель как хорошую.

Для оценки конформации остова часто используют карты Рамачандрана. Для остатка (или для всего белка) строят карту допустимых торсионных углов ϕ и ψ , и каждой точке на карте сопоставляют энергию конформации остатка с такой парой углов. Области низкой энергии будут соответствовать предпочтительным и часто встречающимся конформациям. Если остатки не попадают в предпочитаемые области, то их считают маргинальными, и по их числу оценивают качество структуры (структуру считают хорошей, если >90% остатков находятся в областях с низкой энергией).

При помощи сервиса EDS[3] была построена карта Рамачандрана для рассматриваемой структуры (рис. 2).

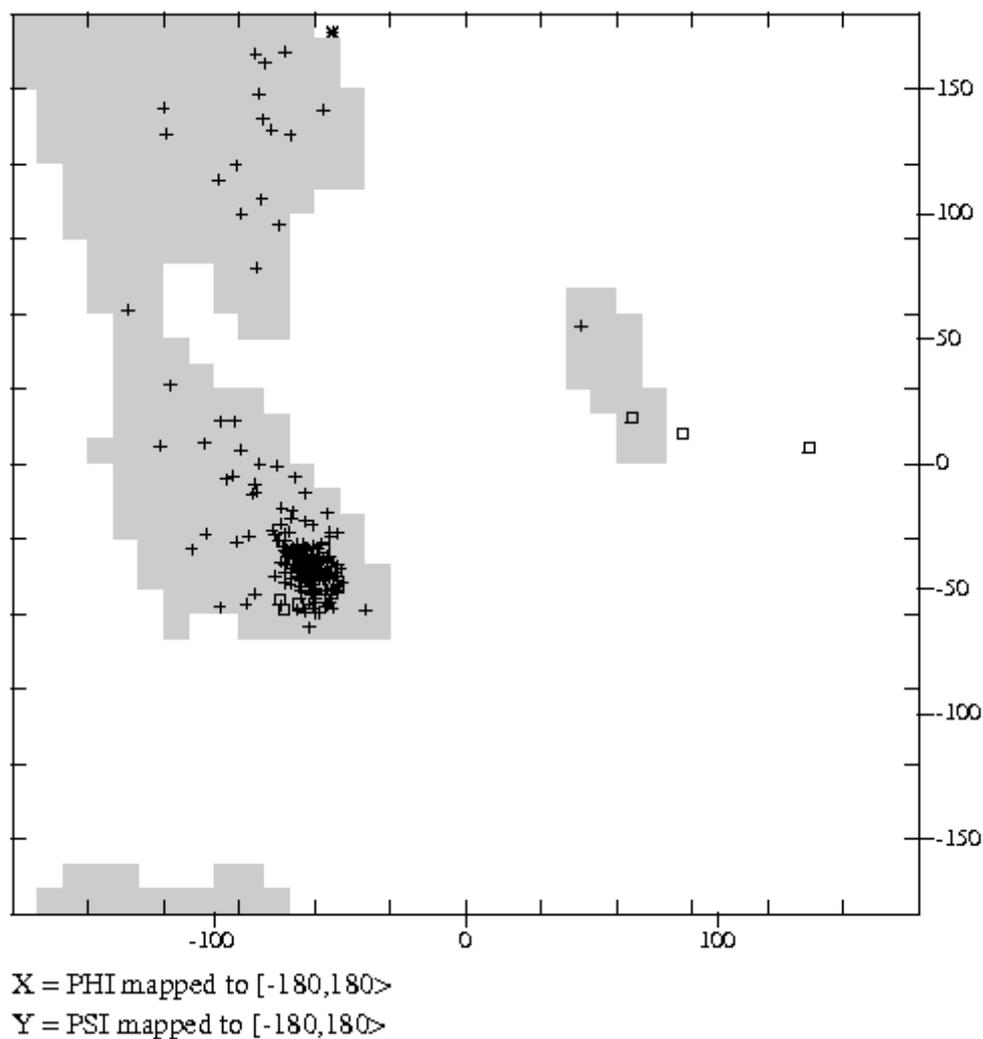


Рисунок 2. Карта Рамачандрана для структуры с PDB ID 4quw

По данным программы EDS процент магринальных остатков составляет 0.5% (всего 1 остаток: SER`14), что можно считать очень хорошим результатом, так как ожидалось 0-5%.

Полученная с помощью сервиса MolProbity[4] карта представлена на рисунке 3. Эта программа тоже нашла всего один маргинальный остаток по карте Рамачандрана (SER`14). Процент остатков в предпочитаемой области составляет 98.19%, в идеале таких должно быть > 98%, то есть результат хороший.

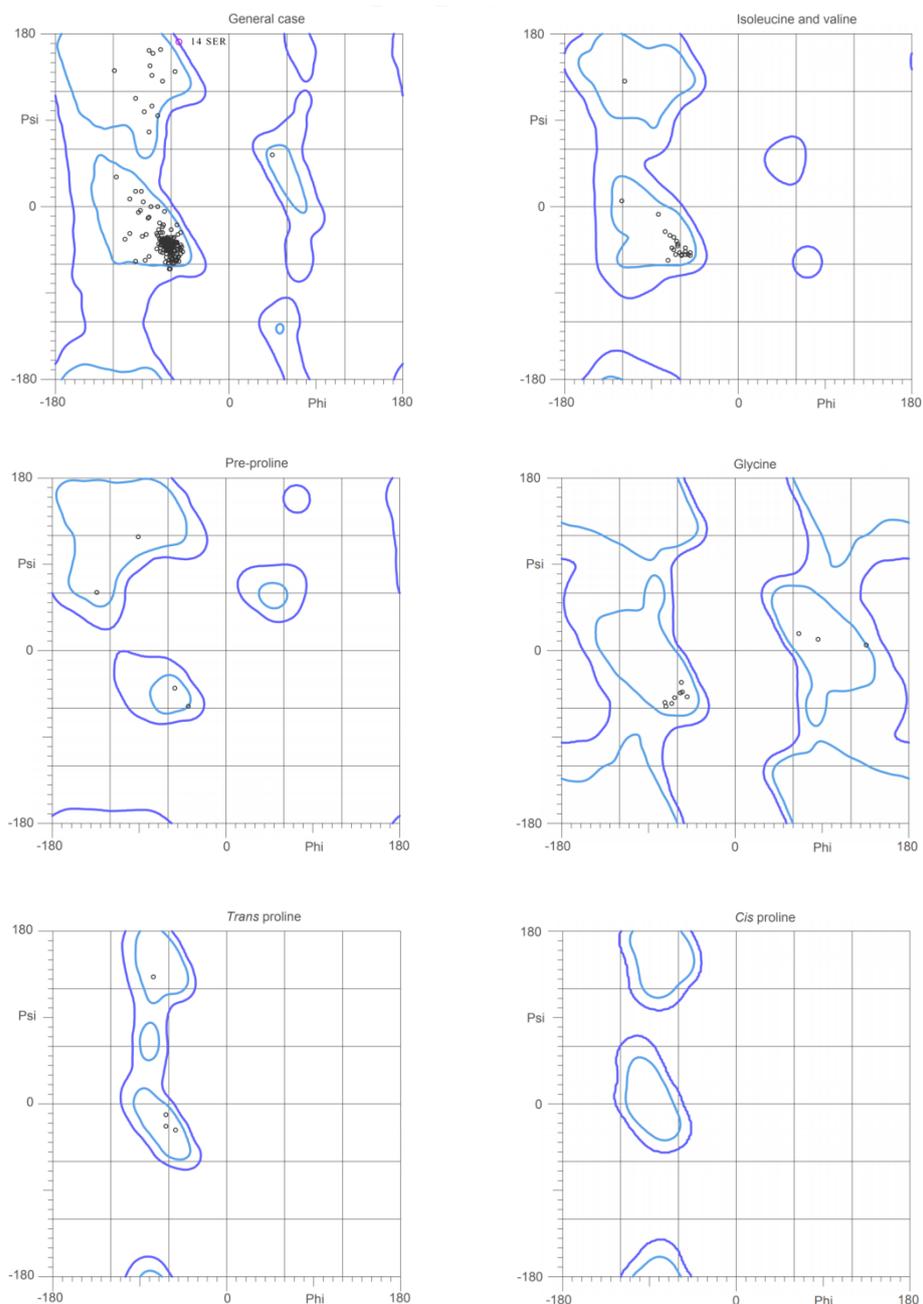


Рисунок 3. Карты Рамачандрана, полученные с помощью программы MolProbity

Ротамеры - это боковые цепи в типичных для данного остатка конформациях. Для каждого остатка есть список характерных ротамеров. Если углы, которые получились в модели не совпадают ни с одним ротамером, их считают маргинальными. Программой MolProbity было найдено 5 маргинальных остатков по положению боковых цепей: LEU`74, VAL`76, LYS`85, VAL`177, VAL`231. Всего таких маргиналов 2.72%, что является не очень хорошим, но не самым плохим значением (для плохих структур – около 7%) [2].

Пространственный R-фактор (RSR) также применяют для выявления маргинальных остатков. Этот метод характеризует соответствие функции электронной плотности,

полученной из модели функции, полученной в эксперименте. Z-score (RSRZ) показывает, насколько RSR для остатка отличается от среднего RSR для такого же типа остатков в структурах PDB со сходным разрешением и используется для оценки RSR. Маргинальным считают остаток, для которого $RSRZ > 2$. В данной структуре 8 таких остатков: ALA`43, ASP`48, LEU`53, TYR`145, ARG`50, GLU`15, PRO`47, LYS`55.

Выше не было сказано только про ClashScore, как показатель качества модели. Это число пар атомов, расположенных слишком близко или наложившихся друг на друга, в расчете на 1000 атомов. На рисунке 4 представлены основные характеристики качества модели [5].

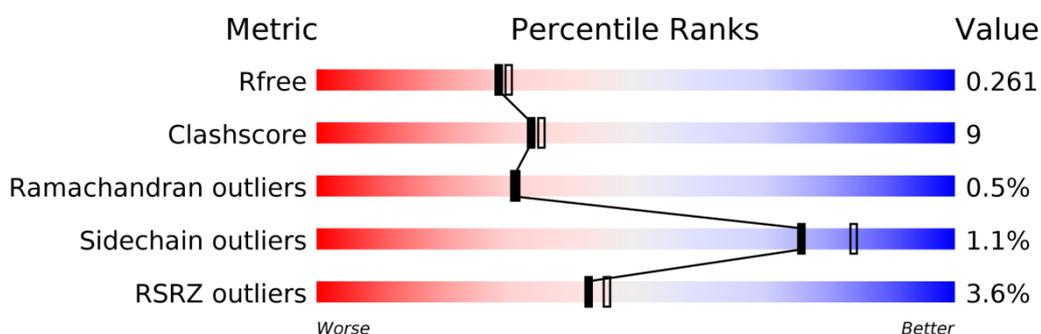


Рисунок 4. Основные показатели качества модели. Черным отмечены перцентили, относительно всех PCA моделей; белым – относительно PCA моделей близкого разрешения

Анализ маргинальных остатков

В таблице №2 перечислены все маргинальные остатки и причины их таковыми считать.

Таблица №2. Список маргинальных остатков	
Остаток	Параметр
SER`14	Из запрещенной области карты Рамачандрана
LEU`74	С неблагоприятной конформацией боковых цепей
VAL`76	
LYS`85	
VAL`177	
VAL`231	
ALA`43	
ASP`48	
LEU`53	
TYR`145	
ARG`50	
GLU`15	
PRO`47	
LYS`55	

Из таблицы №2 видно, что нет остатков, маргинальных сразу по нескольким параметрам. Были выбраны 5 остатков, информация о которых приведена в таблице №3.

Таблица №3. 5 маргинальных остатков	
Остаток	Параметр
SER`14	φ : -52.7, ψ : 172.5
LEU`74	χ : 281.8, 133.8
VAL`177	χ : 319
ALA`43	RSRZ: 3.5
LYS`55	RSRZ: 2.1

Далее рассмотрим подробнее эти а.о.

SER`14

Этот остаток имеет недопустимые значения торсионных углов, то есть лежит в запрещенной области карты Рамачандрана. На рисунке 5 видно, что на уровне подрезки = 2σ основная часть а.о. уже не совпадает со своей электронной плотностью (за счет нехарактерных торсионных углов). Остаток находится в

начале цепи – в участке без регулярной структуры, не связан с активным центром или гидрофобным каналом, поэтому вряд ли можно говорить о функциональной значимости. Скорей всего это ошибка расшифровки.

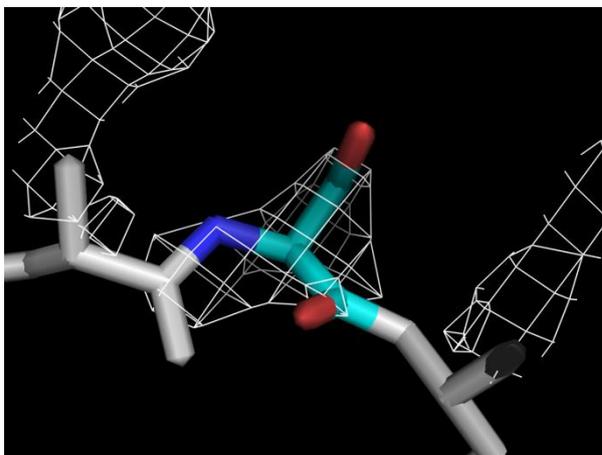
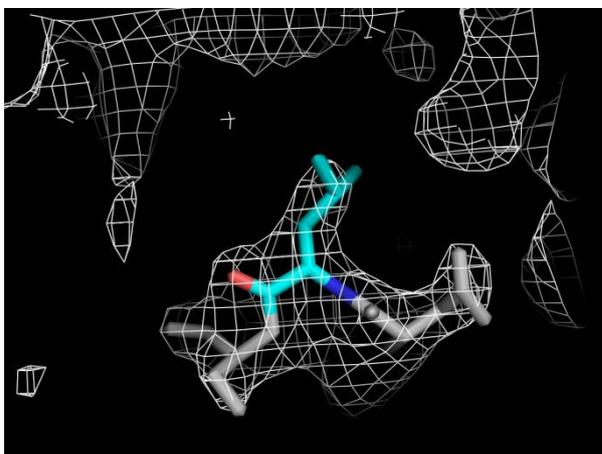


Рисунок 5. Остаток SER`14 и электронная плотность вокруг него на уровне подрезки = 2 σ

LEU`74

Согласно выдаче программы MolProbity боковая цепь LEU`74 находится в неблагоприятной конформации. Действительно, уже при уровне подрезки = 1 σ остаток не совпадает со своей электронной плотностью, что свидетельствует в пользу ошибки расшифровки (рис. 6, сверху). Однако, LEU`74 находится на входе в гидрофобный канал фермента (рис. 6, снизу), и его радикал направлен внутрь структуры. Согласно данным, этот а.о., вместе с несколькими другими, образуя водородные связи, способен закрывать проход в гидрофобный канал и регулировать поступление и выход субстрата и молекул кислорода. Поэтому, возможно, нестандартная конформация боковой цепи а.о. имеет здесь биологический смысл.



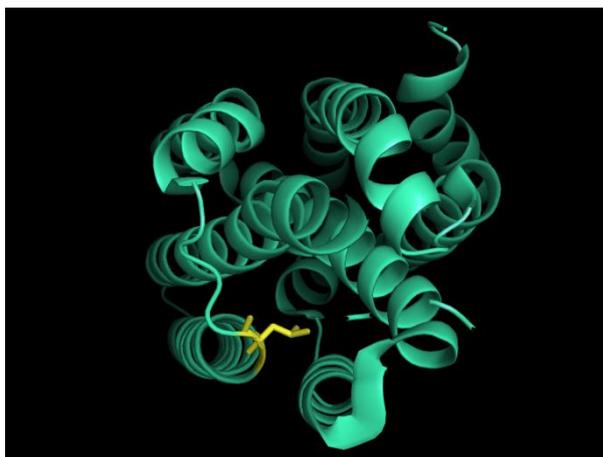


Рисунок 6. Остаток LEU`74 и электронная плотность вокруг него на уровне подрезки = 1 σ (сверху); структура АД с LEU`74 (выделен желтым) на входе в гидрофобный канал (снизу)

VAL`177

Боковая цепь VAL`177 также находится в неблагоприятной конформации, если верить программе MolProbity. Этот остаток обращен радикалом внутрь глобулы белка, но не участвует в формировании функционально значимых элементов. На рисунке 7 видно, что даже на уровне подрезки = 0.5 σ боковая цепь валина не целиком совпадает с электронной плотностью.

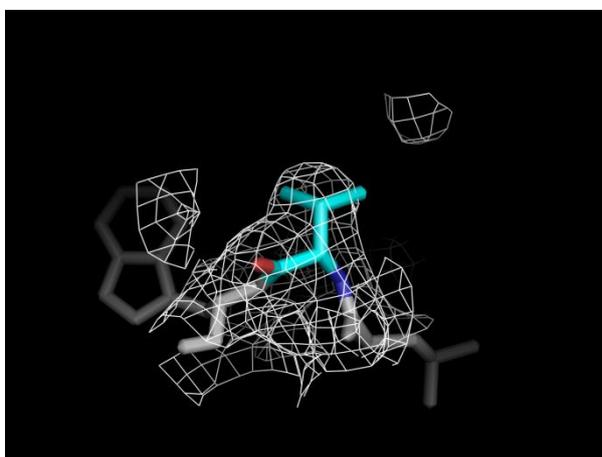


Рисунок 7. Остаток VAL`177 и электронная плотность вокруг него на уровне подрезки = 0.5 σ

В данном случае наиболее вероятно, что произошла ошибка расшифровки.

ALA`43

Остаток ALA`43 имеет плохое соответствие с экспериментальной электронной плотности (RSRZ = 3.5 > 2). По рисунку 8 видно, что на уровне подрезки = 2 σ даже основная часть а.о. не описывается электронной плотностью достаточно хорошо. На уровне 1 σ БОльшая

часть лейцина попадает в область электронной плотности, однако кислород пептидной группы – нет.

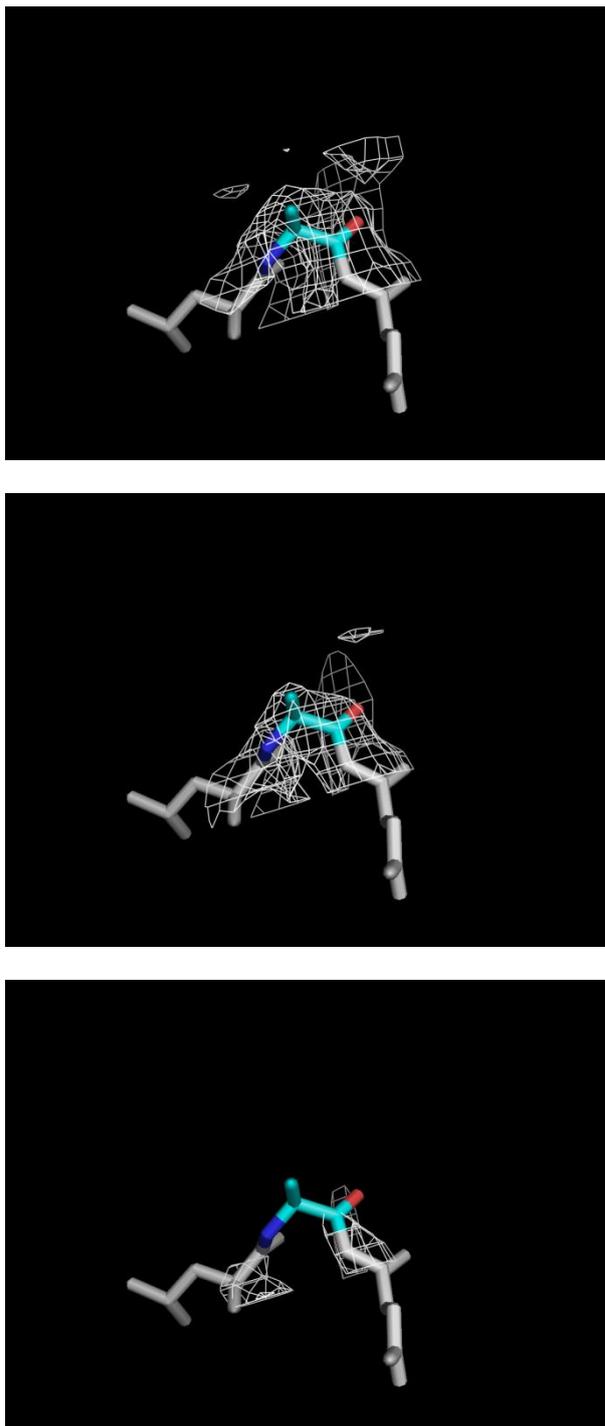


Рисунок 8. Остаток LEU⁷⁴ и электронная плотность вокруг него на уровне подрезки = 0.5 σ (сверху); 1 σ (посередине); 2 σ (внизу)

Чем обусловлено высокое значение RSRZ в данном случае я не могу однозначно сказать. Такое могло бы иметь место, если бы остаток располагался в области без жесткой структуры. Однако, LEU⁷⁴ входит в альфа-спираль и нельзя говорить о его сильной подвижности.

LYS`55

Значения $RSRZ > 2$ для LYS`55 свидетельствуют о том, что остаток плохо вписан в электронную плотность по сравнению с другими структурами с тем же разрешением. На рисунке 9 видно, что на уровне подрезки $= 0.5 \sigma$ боковая цепь лизина не совпадает со своей электронной плотностью. Тем не менее, остов а.о. электронной плотностью описывается достаточно хорошо даже на уровне 2σ , по крайней мере не хуже соседних остатков.

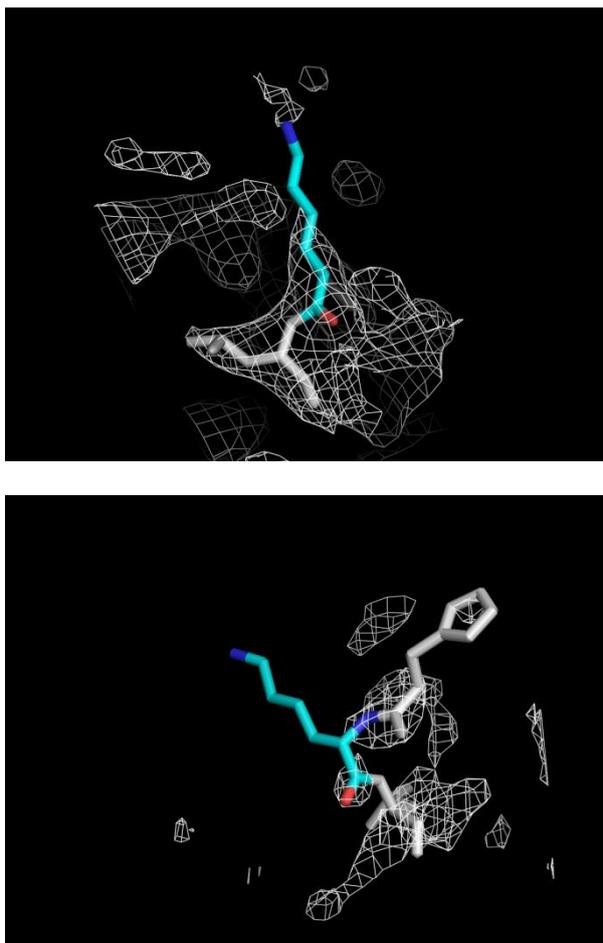


Рисунок 9. Остаток LYS`55 и электронная плотность вокруг него на уровне подрезки $= 0.5 \sigma$ (сверху); 2σ (снизу)

Боковая подвижная цепь LYS`55 направлена наружу белка и не стеснена какими-либо стерическими факторами. Поэтому, вероятно, ошибка расшифровки обусловлена множеством вариантов расположения цепи.

Сравнение модели из PDB с моделью из PDB_redo

База данных PDB_redo содержит оптимизированные версии существующих структур из PDB с картами электронной плотности, описанием внесенных в модель

изменений и оценку качества модели. В таблице №4 приведены индикаторы качества моделей PDB и PDB-REDO. Видно, что значения всех факторов улучшились.

Таблица №4. Индикаторы качества моделей PDB и PDB- REDO		
Параметр	PDB	PDB- REDO
R-фактор	0.2263	0.2091
R_free	0.2655	0.2436
RMSZ длин связей	0.658	0.378
RMSZ углов связей	0.647	0.578

В модель были внесены следующие изменения: изменены 2 ротамера, удалено 8 молекул воды. 16 остатков в итоге стали соответствовать электронной плотности лучше, а 1 – хуже.

Было интересно посмотреть, как изменилось количество маргинальных остатков в улучшенной структуре. Для этого я снова воспользовалась сервером MolProbity (но уже загружая новый файл) и получила следующее: маргинальных остатков из запрещенной области карты Рамачандрана больше нет, число остатков с неблагоприятной конформацией боковых цепей сократилось с 5 до 4, но уменьшилось число предпочитаемых ротамеров (соответственно, стало больше «допустимых», но не лучших).

Таким образом, это усовершенствование заметно сказалось на многих индикаторах качества, однако ухудшавшаяся ситуация с ротамерами не позволяет однозначно считать модель лучше, как мне кажется.

Заключение

По результатам оценки качества расшифровки структуры белка альдегид декарбонилазы, с PDB ID 4quw, полученной методом рентгеноструктурного анализа, можно сказать, что по всем рассмотренным параметрам качество структуры является высоким.

Список литературы

[1] Jia C, Li M, Li J, et al. Structural insights into the catalytic mechanism of aldehyde-deformylating oxygenases. *Protein Cell*. 2014;6(1):55-67. doi: 10.1007/s13238-014-0108-2

[2] Jia, C.J.; Li, M.; Chang, W.R. Crystal structure of the apo form of cyanobacterial aldehyde-deformylating oxygenase. *Protein Cell*. 2014;6 p.55-67

[3] <http://eds.bmc.uu.se/eds/>

[4] <http://molprobit.biochem.duke.edu/>

[5] <https://www.rcsb.org/>