

# Отчет по качеству расшифровки структуры 2IEA

Выполнила: Желудкевич Анна  
группа 402, ФББ

## Оглавление

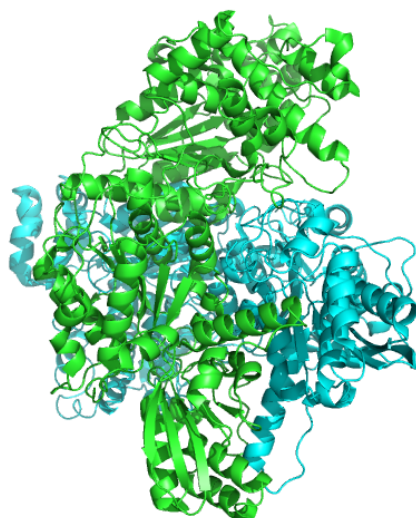
Оглавление.....	1
Аннотация.....	1
Введение.....	1
Результаты и обсуждения.....	2
Общая информация о модели.....	2
Значение индикаторов качества модели в целом.....	2
Примеры маргинальных остатков.....	7
Анализ некоторых маргинальных остатков.....	8
PDB и PDB_redo.....	10
Заключение.....	11
Список литературы и ссылки на источники.....	11

## Аннотация

Данный отчет представляет из себя работу со структурой 2IEA — пируват дегидрогеназы *E.coli*. В тексте приводится общая информация о модели, а также анализ качества расшифровки данной структуры на основе некоторых показателей (R-фактор, R-free, карта Рамачандрана).

## Введение

Пируват дегидрогеназа (EC 1.2.4.1) — фермент класса оксидоредуктаз, входящий в состав пируватдегидрогеназного комплекса, катализирующего превращение трехуглеродного продукта гликолиза - пирувата - в ацетил-кофермент А, который затем вовлекается в цикл трикарбоновых кислот. Ниже представлена пространственная структура пируват дегидрогеназы из файла PDB 2IEA.



Как утверждают авторы статьи [1], данный фермент представлен тремя различными видами:  $\alpha 2$ -гомодимером,  $\alpha 2\beta 2$ -гетеротетрамером у низших организмов и таковым у млекопитающих. Для последних двух типов структуры уже описаны [2], [3], когда полученная авторами

Рисунок 1: Пространственная структура гомодимера пируват дегидрогеназы

структура является первой среди всего типа  $\alpha$ 2-гомомеров данного фермента.

Более того, сравнительный белковый анализ BLAST показал схожесть субъединиц данного фермента с субъединицами из грам-отрицательных бактерий (*Vibrio cholerae*, *Haemophilus influenzae*, *Azotobacter vinelandii*, *Neisseria meningitidis*), а также актиномицетов (*Mycobacterium leprae*, *Mycobacterium tuberculosis*). Таким образом, данная структура является первым представителем семейства схожих ферментов, необходимых в процессе метаболизма сахаров у многих патогенных бактерий.

## Результаты и обсуждения

### Общая информация о модели

Фермент закристаллизован в виде гомодимера, связанного с двумя молекулами тиаминпирофосфата (по одному на каждую субъединицу), двумя ионами магния (по одному на каждую субъединицу), также присутствуют 682 молекулы воды. Авторы работы: Arjunan, P, Nemeria, N, Brunskill, A, Furrey, W из лаборатории биокристаллографии Медицинского Центра Ветеранов, Питтсбург, Пенсильвания. Структура получена в 2002 году. Для решения фазовой проблемы был использован метод многоволнового аномального рассеяния (белок с селенометионинами при трех разных длинах волн). Характеристики, приведенный ниже, можно посмотреть на сайте EDS [4](Electron Density Server) или же в файле структурных факторов с сервера RCSB PDB [5].

Количество рефлексов — 149133. Разрешение очень высокое: 1.80 Å , диапазон разрешений структурных факторов составляет 40.72 - 1.80 Å, полнота данных 87,9%.

Далее рассмотрим характеристики ячейки кристалла:

CRYST1 81.690 141.600 82.460 90.00 102.40 90.00 P 1 21 1 4

Первые три числа в строке соответствуют длинам ребер кристаллографической ячейки (a, b и c), следующие три числа — значения углов ( $\alpha$ ,  $\beta$  и  $\gamma$ ), название группы симметрии (P 1 2<sub>1</sub> 1) и число молекул в ячейке (4). Разрешены следующие операторы симметрии:

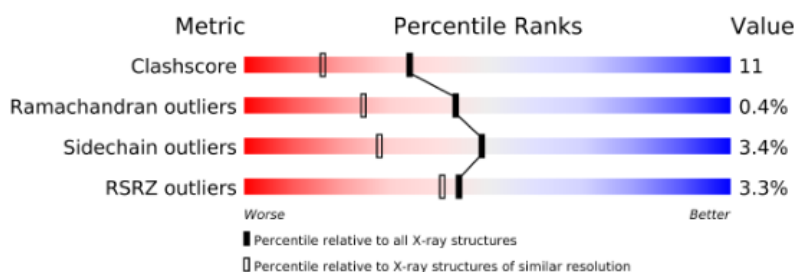
X,Y,Z -X,Y+1/2,-Z

Кристаллографическая ассиметрическая единица содержит один димер.

### Значение индикаторов качества модели в целом.

На сайте PDB имеются основные характеристики того, насколько хорошо модель описывает реальные данные. Данная информация приведена в таблице, где для различных

параметров проводится сравнение модели с таковыми параметрами моделей такого же разрешения (белый бегунок) и с таковыми и у всех структур, полученных методом РСА (черный бегунок). Смещение в красное поле говорит о худшем качестве по сравнению с большинством моелей, в синее — о лучшем. Ниже приводится таблица, где указано на основании какой выборки проводилось сравнение моделей.



Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
Clashscore	79885	1470 (1.86-1.86)
Ramachandran outliers	78287	1451 (1.86-1.86)
Sidechain outliers	78261	1451 (1.86-1.86)
RSRZ outliers	66119	1269 (1.86-1.86)

Рисунок 2: Сравнение параметров данной модели с таковыми других моделей из базы данных PDB

**R-фактор** для модели составляет 0.186 (18,6%). Этот параметр показывает то, насколько рассчитанные уже по построенной модели модули структурных факторов (амплитуды гармоник Фурье)  $F_{hkl}^{calc}$  отличаются от таковых полученных в эксперименте  $F_{hkl}^{obs}$  при одних и тех же числах (h,k,l). Таким образом, R-фактор является мерой схожести модели с экспериментальными данными. Формула для вычисления R-фактора приведена ниже.

$$R = \frac{\sum_{hkl} |F_{hkl}^{calc} - F_{hkl}^{obs}|}{\sum_{hkl} F_{hkl}^{obs}} * 100\%$$

Хорошими значениями для R-фактора являются значения менее 25%. Наша модель попадает под это значение.

**Rfree-фактор** вычисляется точно так же как и выше, но при этом в качестве рефлексов используется специальная группа рефлексов, не участвовавших в оптимизации модели. Параметр рассчитывается по конечной оптимизированной модели. В нашем случае Rfree-фактор составляет 0.203, что считается неплохим значением.

При оптимизации модели R-фактор минимизируется. Но может возникнуть такая ситуация, что при сильной минимизации произойдет «переоптимизация», где мы уйдем от правильной структуры, для этого как раз и рассчитывается R-free. Их разница небольшая, так что можно сказать, что переоптимизации модели не произошло.

Ещё одним параметром, приведённым выше является Clashscore, который показывает

среднее число неблагоприятных перекрываний атомов (более чем на  $0.4\text{\AA}$ ) на 1000 атомов в структуре. В нашей структуре всего 25808 атомов (включая водороды), найдено 275 клэшей, среднее значение, соответственно, равно 11.

Еще одним хорошим показателем качества структуры является количество остатков, которые выбиваются за допустимые зоны (Ramachandran outliers) на карте Рамачандрана (карте значений торсионных углов остова  $\phi$  и  $\psi$ ), которая построена по полученной модели. На рисунке 3 представлена карта для всего остова, и для отдельных аминокислотных остатков, имеющих отличную от общей карту. В рассматриваемой модели 95.1% остатков лежали в благоприятных (favored) и 99,6% в разрешённых (allowed) областях карты.

У трех остатков значения торсионных углов лежат в запрещённой области (это ASN-172, MET-396, TYR-394). На рисунке 3 они представлены красными кружками.

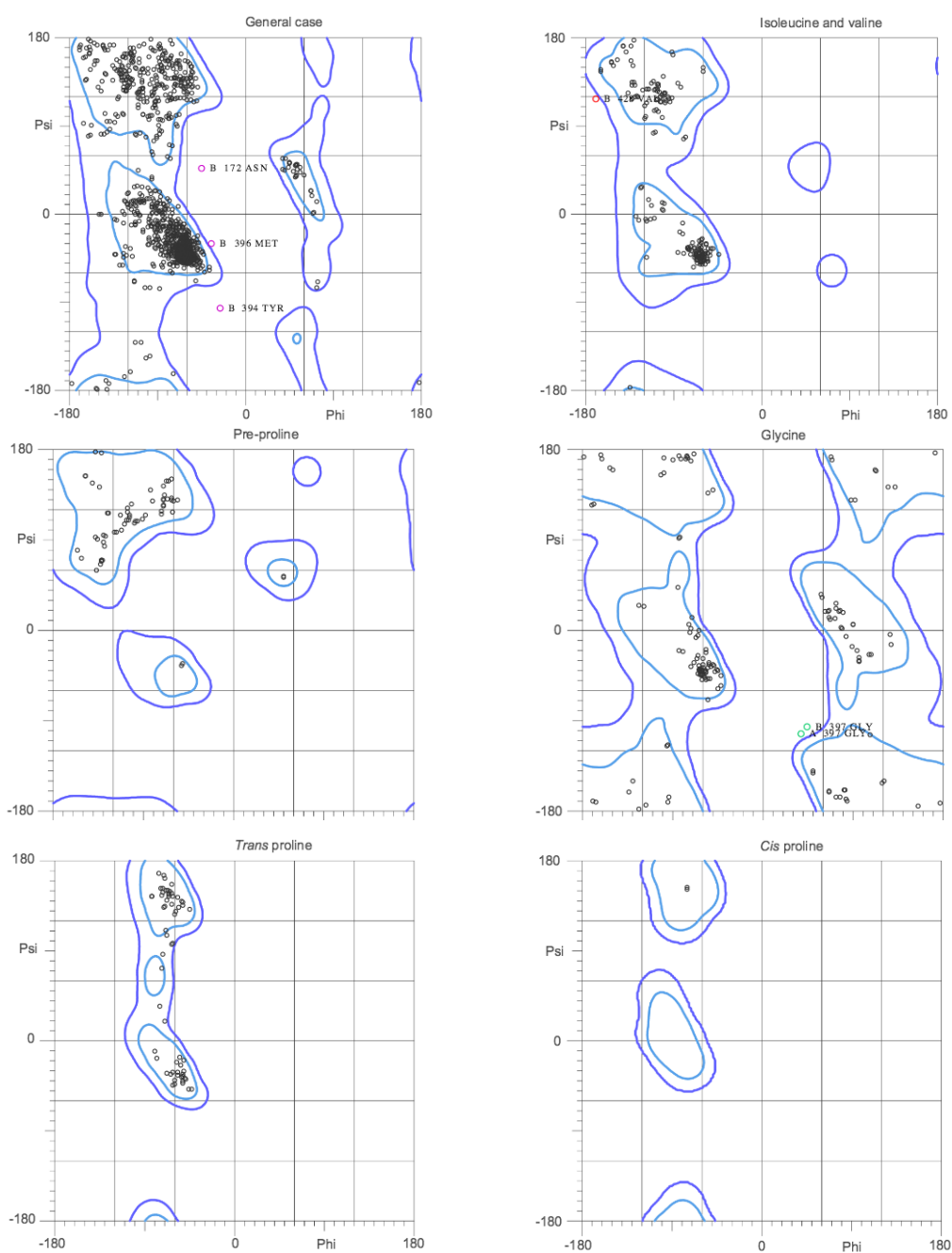


Рисунок 3: Карты Рамачандрана

Аналогично рассчитывается процент выбивающихся остатков по показателям торсионных углов боковых цепей (Sidechain outliers), значение в нашем случае равно 3%.

**Пространственный R-factor (Real Space R-factor = RSR)** является ещё одной мерой соответствия модели реальности, этот параметр рассчитывается по формуле, приведенной ниже, и показывает, насколько реальная электронная плотность согласуется с таковой, построенной в модели.

$$RSR = \frac{\sum_{A \in L} |\rho_{\text{эксп}} - \rho_{\text{модель}}|}{\sum_{A \in L} \rho_{\text{эксп}}} [\cdot 100\%]$$

На основе RSR, рассчитанных для каждого из остатков в цепи белка, можно понять какие остатки выбиваются из общего фона по параметру RSRZ – по сути это Z-score, рассчитанный по RSR с использованием формулы:

$$RSRZ = (RSR - \langle RSR \rangle) / \text{Sigma}$$

На рисунках 4 и 5 представлено распределение значений RSR по цепи А и В белка. Среднее значение RSRZ составило 0.128 для цепи А и 0.150 для цепи В. Распределения значений RSRZ для цепей А и В представлены на рисунках 6 и 7.

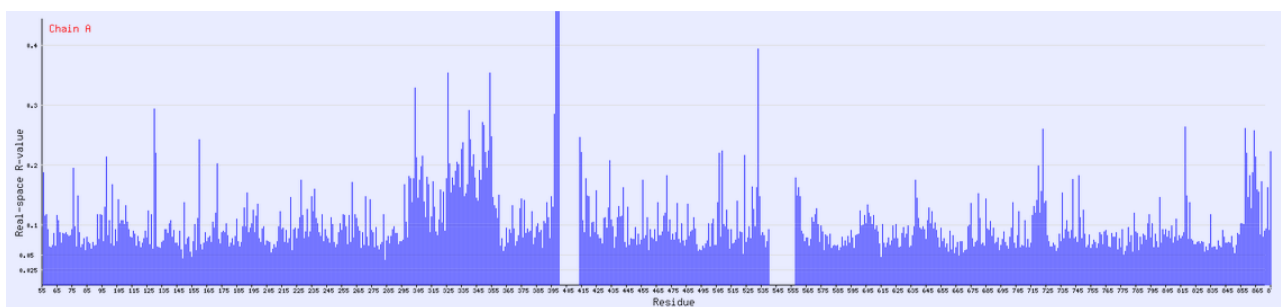


Рисунок 4: Распределение значения RSR от номера остатка в цепи А структуры 2IEA

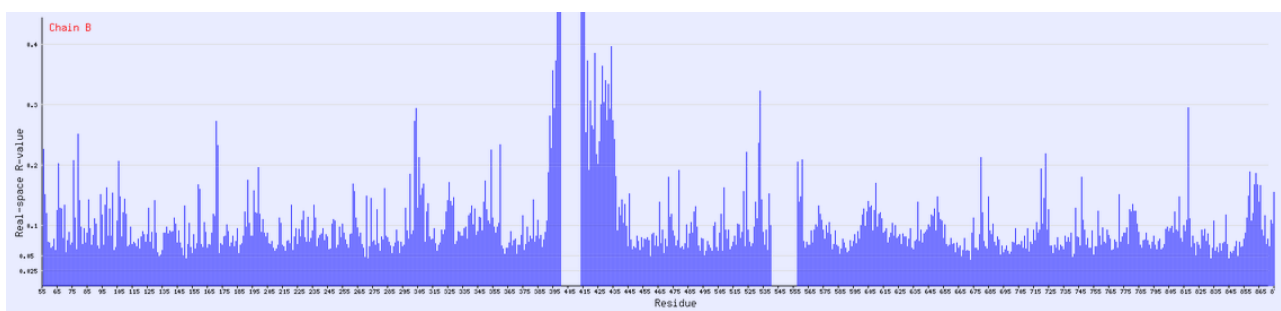


Рисунок 5: Распределение значения RSR от номера остатка в цепи В структуры 2IEA

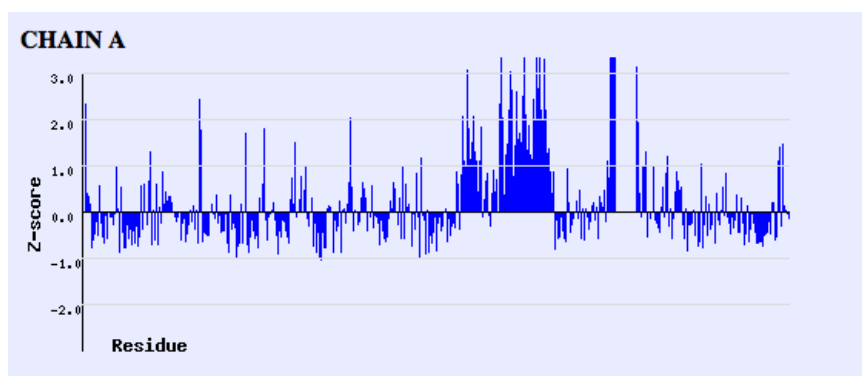


Рисунок 6: Распределение значения RSRZ от номера остатка в цепи A структуры 2IEA

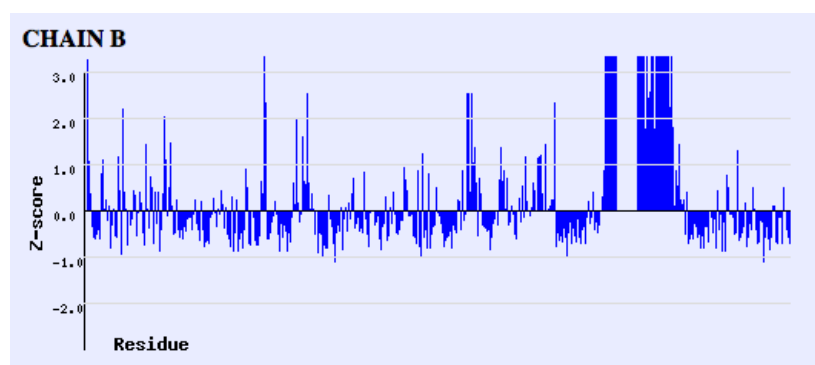


Рисунок 7: Распределение значения RSRZ от номера остатка в цепи B структуры 2IEA

Кроме очевидно выбивающихся остатков на рисунках 4 и 5 видно, что в целом у большинства остатков RSR не более 10%, но по графикам RSRZ у достаточно большого количества остатков z-score больше двух, что говорит о том, что остаток плохо вписан в электронную плотность.

Комфортность окружения остатков и некоторые параметры боковых цепей оценивались при помощи web-интерфейса программы WhatCheck [6](раздел Structure Validation).

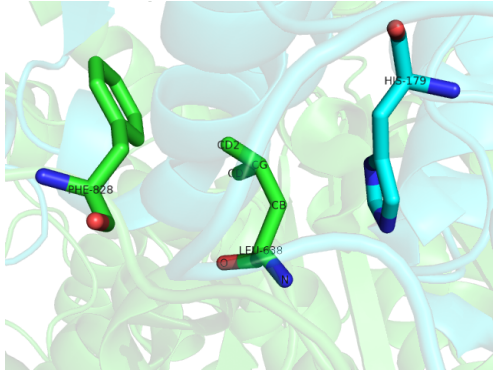
- Планарность: по результату работы WhatCheck выявлено, что все остатки, которые должны иметь плоскую конфигурацию, такую и имеют.
- Комфортность окружения боковых радикалов аминокислотных остатков: среднее значение, которое было получено -0,244 (Z-score -1,62). По комментария WhatCheck, это является хорошим результатом, среднее значение не меньше 2,5, а Z-score не меньше -2, что соответствует хорошему окружению.
- Конформация пролинового кольца: у а.к. пролин отклонение C-γ атома от плоскости



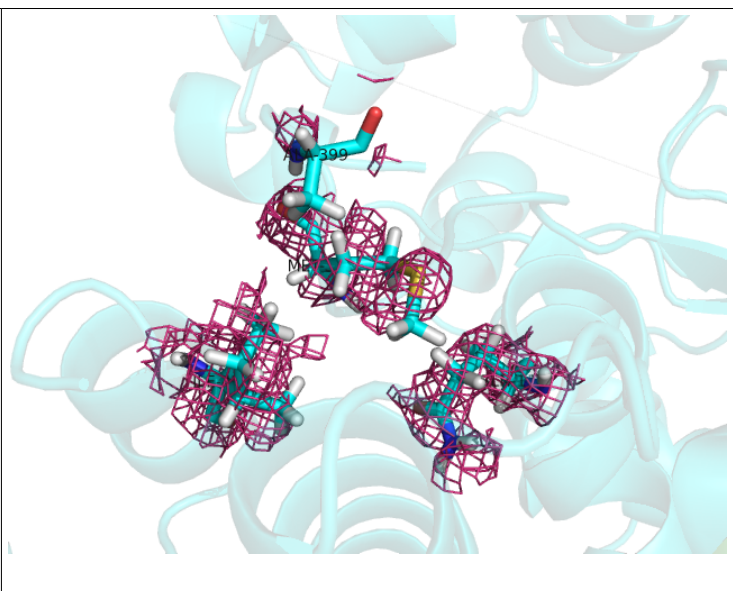
1	LEU-638	A,B	Отклонение угла CA-CB-CG на превышенное значение, перекрывание с PHE-828 (0,83A)
2	MET-396	B	Перекрытие с ALA-399 (clash 0,94A), маргинал по карте Рамачандрана (неразрешенная зона)
3	GLY-397	A,B	Маргинал по карте Рамачандрана
4	VAL-685	A	Маргинал по торсионным углам в боковой цепи остатка (non-rotameric sidechain)
5	PRO-198	B	Неправильная конформация пролинового кольца
6	ARG-533	B	RSRZ outlier (3,2)
7	ALA-399	B	RSRZ outlier (18,2)
8	GLU-76	A	Маргинал по торсионным углам в боковой цепи остатка (non-rotameric sidechain)

### Анализ некоторых маргинальных остатков

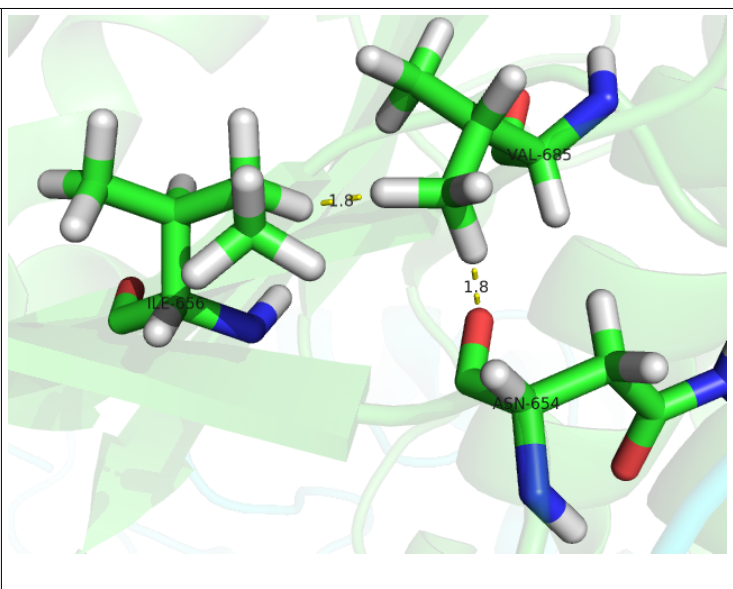
Возьмем на рассмотрение несколько остатков (638 лейцин, 396 метионин, 685 валин, 944 пролин и 399 аланин).

Остаток	Причина маргинальности	Изображение остатка
LEU-628 (A)	Отклонение угла CA-CB-CG на превышенное значение; перекрывание с PHE-828 (a) и HIS-179 (b); маргинал по торсионным углам в боковой цепи остатка (non-rotameric sidechain).	
<p>Как видно из рисунка, остаток лейцина «заключен» между остатками фенилаланина-828 и гистидина-179, который расположен на соседней цепи. При добавлении молекул воды становится заметно, что все три остатка образуют водородные связи с ними. Но мне все же не совсем ясно, почему остаток лейцина принимает такую форму.</p>		

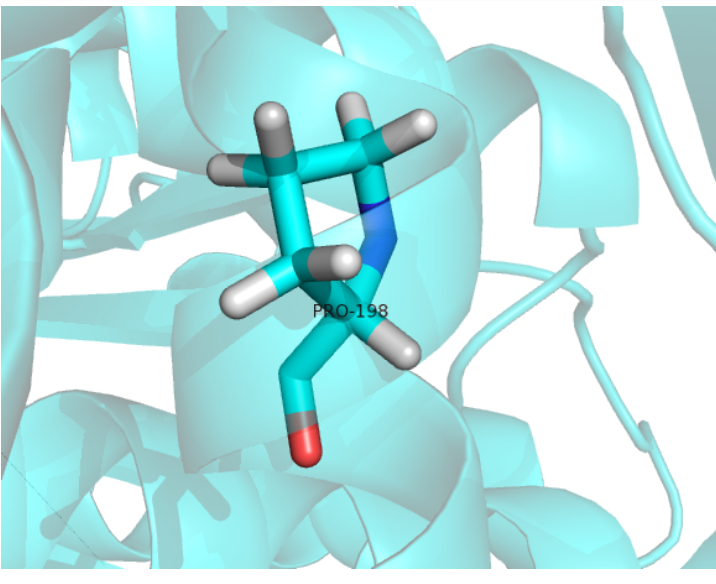


<p>MET-396 (B)</p>	<p>Перекрытие с ALA-399, GLN-111 и ILE-420; маргинал по торсионным углам в боковой цепи остатка (non-rotameric sidechain), маргинал по RSRZ.</p>	
--------------------	--	--

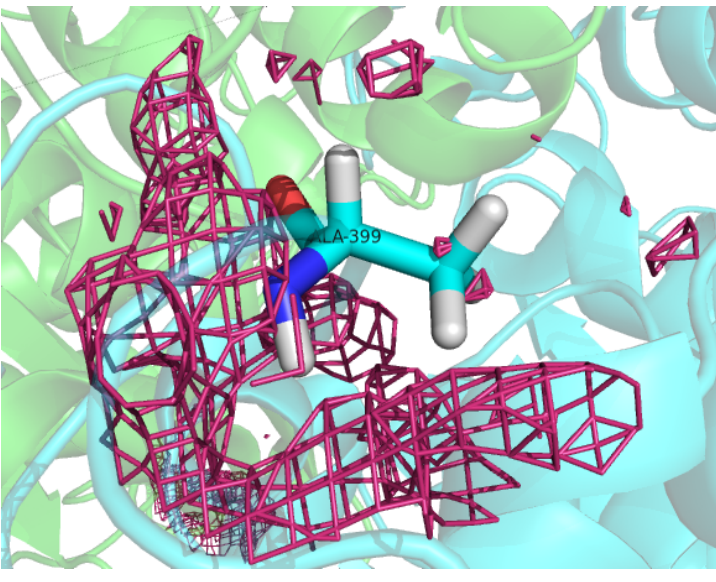
Остатки очень плохо вписываются в электронную плотность, плюс ко всему, данные три остатка действительно перекрываются с метионином. Можно предположить, что поскольку эти аминокислоты находятся на поверхности глобулы, то они весьма подвижны, что приводит к тому, что сложно закристаллизовать структуру.

<p>VAL-685 (A)</p>	<p>Перекрытие с ASN-654 и ILE-656; маргинал по торсионным углам в боковой цепи остатка (non-rotameric sidechain).</p>	
--------------------	---	---

При взгляде на рисунок, видно, что три остатка перекрываются атомами водорода, чего не должно быть. Вероятно, ошибка расшифровки.

PRO-198 (B)	Неправильная конформация пролинового кольца.	
-------------	--	--

Пролин действительно находится в неправильной конформации, более того, он еще и входит в состав альфа-спирали, которая и «изгибает» его в то, что получилось. Выглядит это необычно.

ALA-399 (B)	Перекрывание MET-396; маргинал по RSRZ	
-------------	--	---

Остаток плохо вписан в электронную плотность. Возможная причина указана выше.

## PDB и PDB\_redo

PDB\_redo [7] является сервером для оптимизации модели, используя разные параметры.

После использования данного сервера, улучшилось значение R-фактора, для некоторых улучшилось вписание остатков электронную плотность, улучшены конформация скелета, вид карты Рамачандрана (меньше остатков в сравнительно неблагоприятных областях), оптимизированы длины и углы связей, уменьшено число столкновений (clashes) атомов .

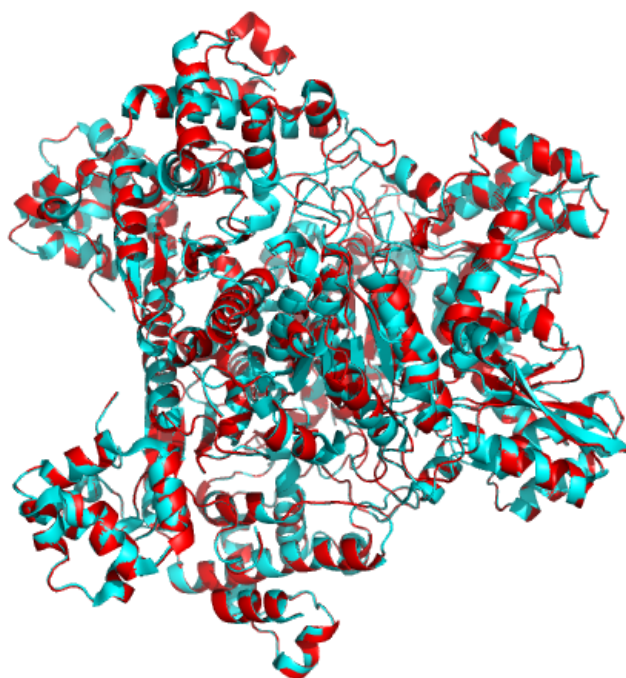


Рисунок 9: Наложение двух структур

Выше представлено изображение совмещения оригинальной структуры (голубая) и полностью оптимизированной (красная). Различий в структуре не очень много.

### Заключение

Исходя из вышеприведённой информации, могу отметить, что качество расшифровки структуры можно считать средним. Основные параметры качества структуры соответствуют принятым стандартам, но при рассмотрении маргинальных остатков возникает подозрение, что структура в некоторых местах расшифрована неверно. Повторная оптимизация модели (PDB\_redo), хотя и улучшила ряд параметров, но принципиальных изменений в неё не внесла.

### Список литературы и ссылки на источники

- [1] P. Arjunan, N. Nemeria, A. Brunskill, K. Chandrasekhar, M. Sax, Y. Yan, F. Jordan, J. R. Guest, and W. Furey, "Structure of the pyruvate dehydrogenase multienzyme complex E1 component from *Escherichia coli* at 1.85 Å resolution," *Biochemistry*, vol. 41, no. 16, pp. 5213–5221, 2002.
- [2] A. Evarsson, K. Seger, S. Turley, J. R. Sokatch, and W. G. Hol, "Crystal structure of 2-oxoisovalerate and dehydrogenase and the architecture of 2-oxo acid dehydrogenase multienzyme complexes.," *Nat Struct Biol*, vol. 6, no. 8, pp. 785–792, 1999.
- [3] A. Evarsson, J. L. Chuang, R. M. Wynn, S. Turley, D. T. Chuang, and W. G. Hol, "Crystal structure of human branched-chain alpha-ketoacid dehydrogenase and the molecular basis of multienzyme complex deficiency in maple syrup urine disease.," *Structure*, vol. 8, no. 3, pp. 277–91, 2000.

- [4] “<http://eds.bmc.uu.se/cgi-bin/eds/uusfs?pdbCode=2iea>.” .
- [5] “<http://www.rcsb.org/pdb/explore/explore.do?structureId=2IEA>.” .
- [6] “<http://swift.cmbi.ru.nl/servers/html/index.html>.” .
- [7] “[http://www.cmbi.ru.nl/pdb\\_redo/ie/2iea/index.html](http://www.cmbi.ru.nl/pdb_redo/ie/2iea/index.html).” .