

Отчёт по качеству расшифровки структуры енолазы из *Entamoeba histolytica* (PDB_ID: 3QTP)

Выполнил: Травин Дмитрий (ФББ МГУ, группа 402)

Содержание

Аннотация.....	2
Введение	2
Результаты и обсуждение.....	3
1. Общая информация о модели.....	3
2. Индикаторы качества модели в целом.....	4
3. Маргинальные остатки в структуре белка. Примеры.	9
4. Разбор примеров отдельных маргинальных остатков.....	10
5. Сравнение модели из PDB с моделью из PDB_redo.....	14
Заключение и выводы.....	16
Список литературы и web-сервисов	17

Аннотация

В приводимом ниже отчёте ведётся работа со структурой енолазы из *Entamoeba histolytica* 3QTP. Приводится общая информация о структуре (основные экспериментальные данные и параметры, использованные при улучшении модели), рассматриваются показатели качества модели в целом, выделяются маргинальные остатки на основе различных параметров, разбирается ряд отмеченных на предыдущем шаге остатков с нетипичными параметрами. Работа завершается выводами касательно качества расшифровки структуры. В работе используется ряд интернет сервисов (PDB, EDS, WhatCheck, MolProbity и др.).

Введение

Енолаза (также известная как фосфоенолпируват гидратаза) [EC 4.2.1.11] – фермент класса лиаз, принимающий участие в гликолизе и катализирующий предпоследнюю (девятую) реакцию этого метаболического пути: превращение 2-фосфоглицерата в фосфоенолпируват, сопровождающееся

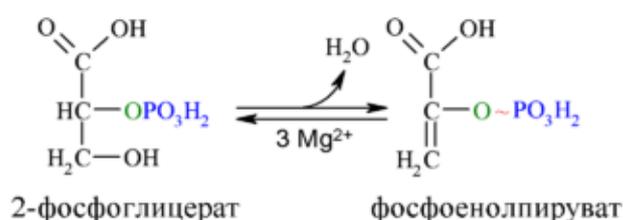


Рисунок 1. Реакция, катализируемая енолазой.

отщеплением молекулы воды (см. Рис. 1). Обратная реакция также катализируется этим ферментом и происходит, например, в процессе глюконеогенеза.

В файле с PDB_ID: 3QTP приводится структура димера енолазы из *Entamoeba histolytica* (продукт гена ENL-1). Интерес именно к этому ферменту у авторов работы [1] обусловлен тем, что ежегодно, будучи паразитом, *E. histolytica* заражает более чем 50 миллионов человек, причём примерно в 100 000 случаев заражение имеет летальный исход. На сегодняшний день стандартной терапией против амёбиаза является препарат метронидазол, к которому в последнее время начинает появляться устойчивость в природных популяциях [2], и который кроме того имеет ряд неприятных побочных эффектов. В связи с этим крайне актуален поиск новых и принципиально иных подходов к лечению заражения, и в этом плане ферменты гликолиза являются одним из важнейших кандидатов, поскольку *E. histolytica* не имеет митохондрий, а гликолиз является единственным путём получения энергии микроорганизмом.

Ещё одним важным аспектом является сравнительный анализ ферментов гликолиза среди различных паразитов (*E. histolytica*, *Toxoplasma gondii* и др.) и с таковыми у человека. Это связано с тем, что антипаразитарный препарат, мишенью для которого являются ферменты гликолиза, должен не вредить нормальному метаболизму человеческих клеток.

Результаты и обсуждение

1. Общая информация о модели

Енолаза из *E. hisolytica* была гетерологически экспрессирована в *E.coli* BL21(DE3) и закристаллизована в виде гомодимера. При этом в структуре одна из субъединиц (А) находится в так называемой открытой конформации, а вторая (В) в закрытой. Структура также содержит три уникальных низкомолекулярных лиганда: 2-фосфоглицерат, ион Mg^{2+} и ион сульфата. Кроме того, в структуре присутствуют молекулы воды.

Структура была получена Е. С. Schulz и R. Ficner из Геттингенского университета имени Георга-Августа в 2011 году. Для решения фазовой проблемы был использован метод молекулярного замещения на основе структуры 1E9I [3] – енолазы из *E.coli*.

С использованием информации с сайта EDS (Electron Density Server, [4]) были получены следующие данные об экспериментальных данных:

- Получено 75593 рефлексов, всего измерено 75593 значения структурных факторов, из них для оптимизации были использованы 71813 (95.0%), тестовый набор рефлексов составил, соответственно, 3780 штук (5.0%).
- Заявляемое авторами разрешение структуры составляет **1,90 Å** (хорошее).
- Полнота данных составляет **99,5%**.
- Диапазон разрешений структурных факторов составляет **46.34 - 1.90 Å**.

Далее приведу информацию касательно пространственных характеристик ячейки кристалла, полученную из PDB-файла (поле CRYST1):

```
CRYST1    64.420    92.620   160.580   90.00   90.00   90.00  P 21 21 21  8
```

Первые три числа в строке соответствуют длинам рёбер кристаллографической ячейки (a, b и c), за ними идут значения углов (α , β и γ), название пространственной группы симметрии $P 2_1 2_1 2_1$ и число молекул в одной ячейке, равное восьми. На рисунке 2 показаны три соседние ячейки кристалла и входящие в них димеры белков (те димеры, которые могут быть получены путём применения преобразований кристаллографической симметрии один из другого, покрашены одинаковыми цветами). Видно, что в каждой ячейке находится 4 молекулы димеров, всего 8 молекул фермента.

Некристаллографическая симметрия в пределах ячейки отсутствует. Для пространственной группы $P 2_1 2_1 2_1$ возможны следующие операторы симметрии:

$x, y, z; -x+1/2, -y, z+1/2; -x, y+1/2, -z+1/2; x+1/2, -y+1/2, -z.$

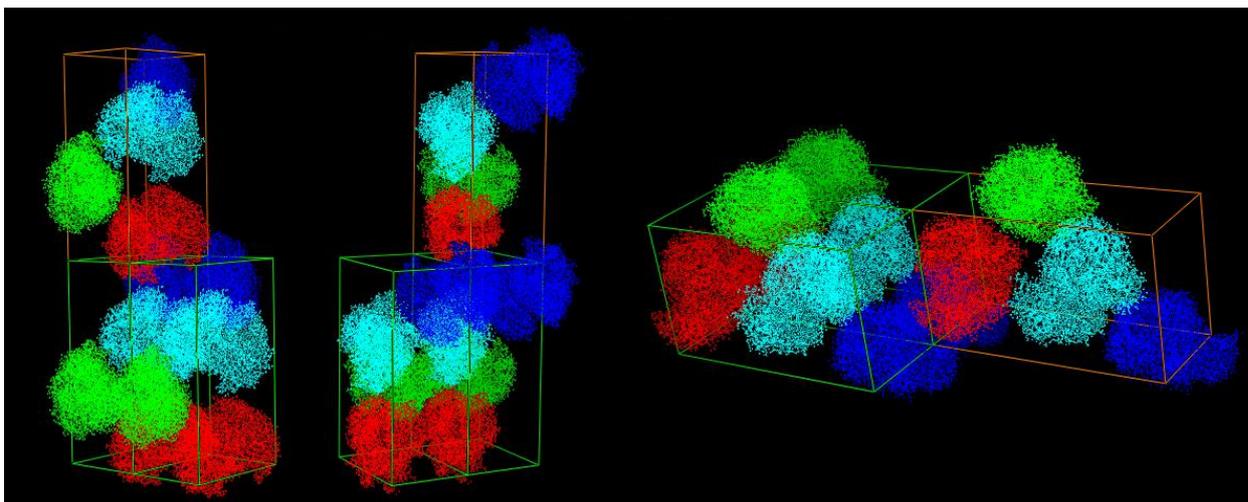
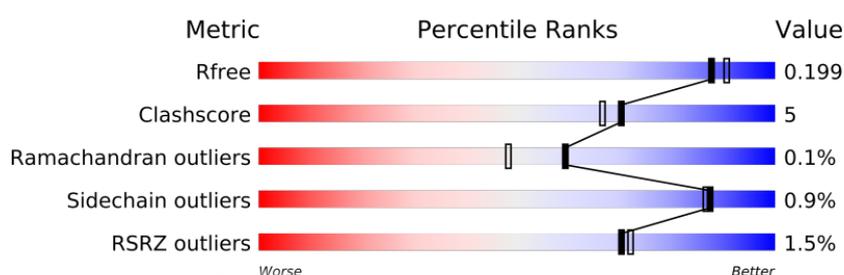


Рисунок 2. Несколько кристаллографических ячеек в структуре кристалла, показаны молекулы в составе трёх соседних ячеек (в каждой ячейке 4 димера: синий, красный, зелёный и голубой). См. комментарии в тексте.

2. Индикаторы качества модели в целом

Основные характеристики того, насколько хорошо полученная модель описывает реальные данные приведены на сайте PDB [5] в виде поля (см. Рис. 3), на котором происходит сравнение различных параметров (графа Metric) модели с таковыми у других моделей такого же разрешения (белый «ползунок») и с таковыми у всех структур, полученным методами рентгеновской кристаллографии (черный «ползунок»). Сдвиг в синее поле считается хорошим относительно большинства, а в красное, соответственно, плохим. Справа от поля (графа Value) представлено абсолютное значение параметра для рассматриваемой модели. Под визуальным представлением можно видеть таблицу, в которой указано, на основе какого числа моделей было сформировано верхнее представление (каков был объём выборки, с которой сравнивался конкретный параметр).



Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	66092	3684 (1.90-1.90)
Clashscore	79885	4465 (1.90-1.90)
Ramachandran outliers	78287	4413 (1.90-1.90)
Sidechain outliers	78261	4414 (1.90-1.90)
RSRZ outliers	66119	3686 (1.90-1.90)

Рисунок 3. Параметры модели в сравнении с таковыми для других моделей в базе данных PDB (всех и одинакового разрешения).

Теперь попробуем разобраться с основными параметрами модели по одному и понять с какой стороны они её характеризуют.

R-фактор для модели составляет **0.153** (15.3%). Этот параметр показывает то, насколько рассчитанные уже по построенной модели модули структурных факторов (амплитуды гармоник Фурье) F^{calc} отличаются от таковых полученных в эксперименте F^{obs} при одних и тех же числах (h,k,l). Таким образом, R-фактор является мерой схожести модели с экспериментальными данными и рассчитывается по формуле (1).

$$(1) \quad R = \frac{\sum_{hkl} |F_{hkl}^{calc} - F_{hkl}^{obs}|}{\sum_{hkl} F_{hkl}^{obs}} * 100\%$$

Хорошими значениями для R-фактора являются значения менее 25%. Наша модель вписывается с неплохим запасом. Но подводным камнем в данном случае является то, что при оптимизации модели мы минимизируем именно этот параметр и может возникнуть ситуация, когда происходит переоптимизация (overfitting), и, уменьшая R-фактор, мы всё дальше уходим от реальной структуры белка. Для недопущения такой ситуации служит ещё один параметр: R_{free} -фактор.

R_{free} -фактор вычисляется точно так же как и выше, но при этом в качестве рефлексов используется специальная группа рефлексов (в нашем случае 3780 штук (5.0% от всех)), не участвовавших в оптимизации модели. Параметр рассчитывается по конечной оптимизированной модели. В нашем случае R_{free} -фактор составляет **0.199**, что считается неплохим значением, при этом обращу внимание, что разница между R и R_{free} составляет менее 5%, что неплохо (переоптимизации, скорее всего, не произошло).

Ещё одним параметром, приведённым выше является **Clashscore**, который показывает среднее число неблагоприятных перекрытий атомов (более чем на 0.4Å) на 1000 атомов в структуре. В нашей структуре всего 14 745 атомов, найдено 62 участка с clash. Среднее значение составляет около **5**, оно и приведено на рисунке 3.

Ряд далее приведённых характеристик может быть получен при использовании сервиса MolProbity [6], хотя ряд из них можно почерпнуть из отчётов, доступных в PDB на странице структуры.

Хорошим показателем качества структуры является количество остатков, которые выбиваются за допустимые зоны (**Ramachandran outliers**) на карте Рамачандрана (карте значений торсионных углов остова ϕ и ψ), которая построена по полученной модели. На рисунке 4 представлена карта для всего остова (A), и для отдельных аминокислотных остатков, имеющих отличную от общей карту (B-E). В рассматриваемой модели **96.9%** остатков лежали в благоприятных (**favored**) и **99,9%** в разрешённых (**allowed**) областях карты.

Был обнаружен один остаток (0,1% общего числа), значения торсионных углов которого лежали в запрещённой области (это ARG-404, его мы рассмотрим в разделе, посвящённом маргинальным остаткам). Маргинальный аргинин отмечен красным кружком на рисунке 4А.

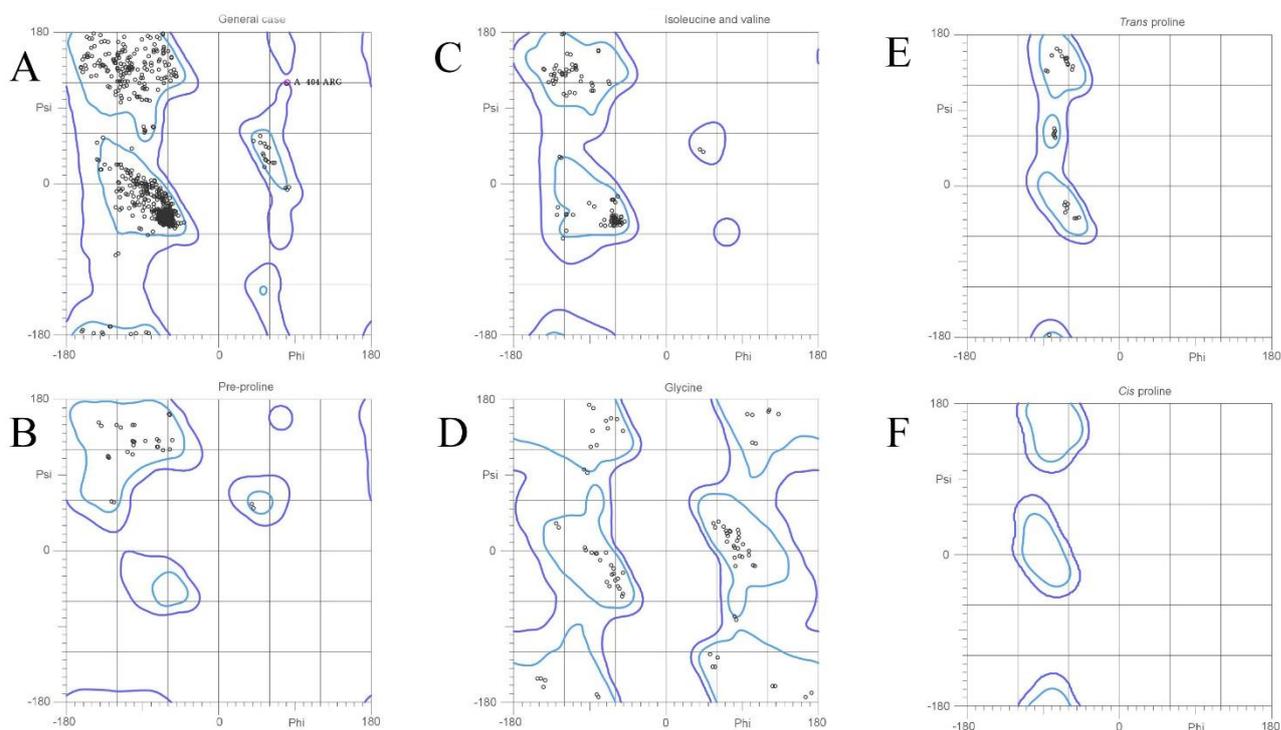


Рисунок 4. Карты Рамачандрана для модели 3QTP: А - общий случай, В - препролиновые остатки, С - изолейцины и валины, D – глицины, Е – транс-пролины и F – цис-пролины (в модели отсутствуют). На данных картах голубой контур окружает 98% всех остатков (по множеству структур) – это благоприятная область, а внешний фиолетовый контур 99.5% всех остатков – это разрешённая область.

Аналогично рассчитывается процент выбивающихся остатков по показателям торсионных углов боковых цепей (**Sidechain outliers**). Это значение в данном случае несколько больше, чем для остова и составляет **0.9%**.

Пространственный R-factor (Real Space R-factor = RSR) является ещё одной мерой соответствия модели реальности, этот параметр рассчитывается по формуле (2) и показывает, насколько реальная электронная плотность согласуется с таковой, построенной в модели.

$$(2) \quad RSR = \frac{\sum_{A \in L} |\rho_{\text{ЭКСП}} - \rho_{\text{МОДЕЛЬ}}|}{\sum_{A \in L} \rho_{\text{ЭКСП}}} [\cdot 100\%]$$

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 (3).

$$(3) \quad RSRZ = (RSR - \langle RSR \rangle) / \text{Sigma}$$

На рисунке 5 представлено распределение значений RSR по цепи А белка. Среднее значение RSRZ составило -0.47 для цепи А и -0.38 для цепи В. Распределение значений RSRZ для цепи А (для цепи В не привожу) представлено на рисунке 6. Оба рисунка (5 и 6) получены при помощи соответствующих сервисов EDS [4].

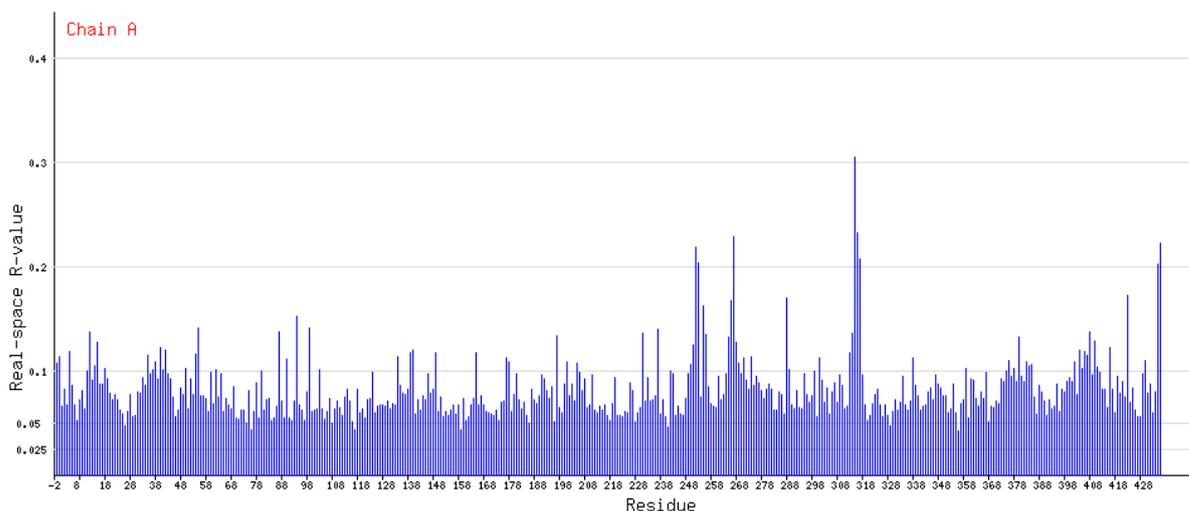


Рисунок 5. Распределение значения RSR от номера остатка в цепи А структуры 3QTP.

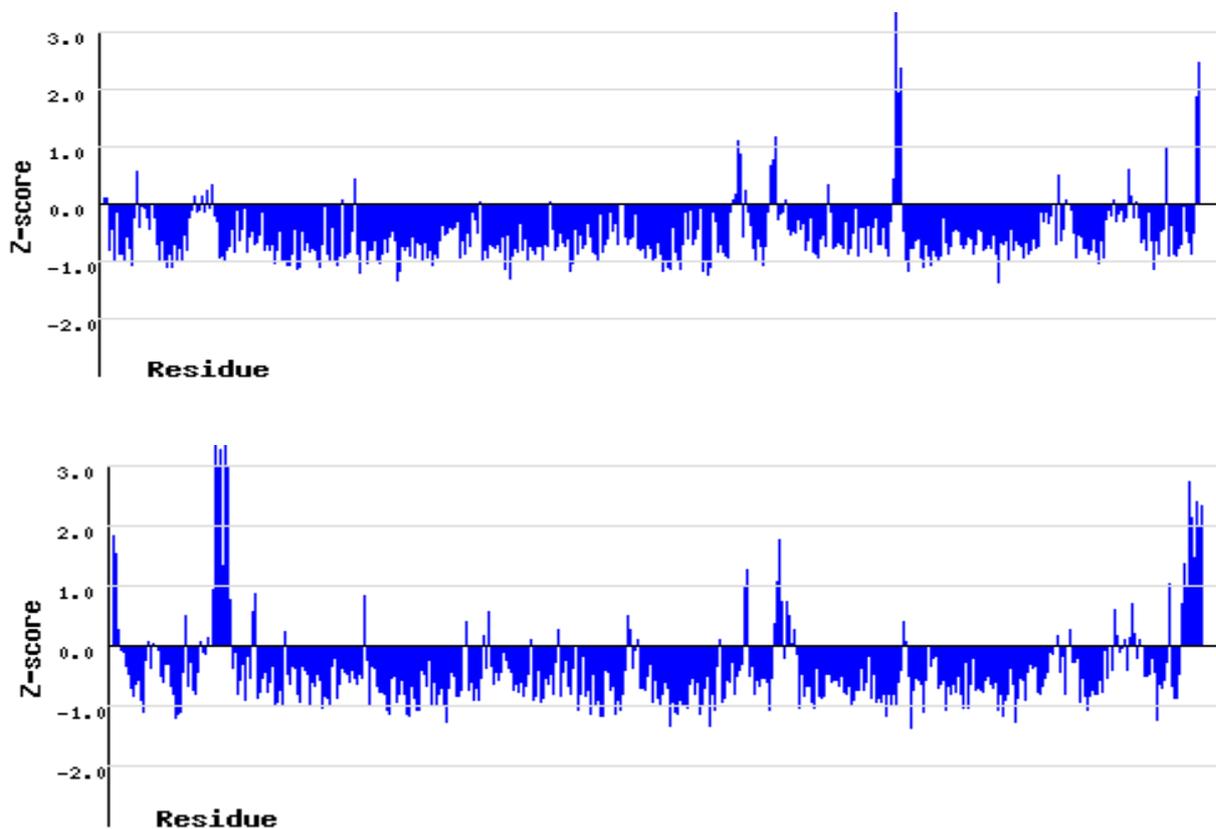


Рисунок 6. Распределение значения RSRZ от номера остатка в цепи А (сверху) и В (снизу) структуры 3QTP.

Маргинальными с точки зрения RSRZ (**RSRZ outliers**) считаются остатки, имеющие значение этого параметра больше двух, то есть отклоняющиеся от среднего качества описания электронной плотности более чем на два стандартных отклонения в «плохую» сторону. Всего таких остатков в обеих цепях белка **11 (1,25%)** при общем числе остатков равном 882. Эти остатки также будут подробнее рассмотрены в разделе о маргиналах.

Кроме очевидно выбивающихся остатков на рисунках 5 и 6 видно, что в целом электронная плотность модели неплохо соответствует экспериментальной (у большинства остатков RSR не более 10% и отрицательный RSRZ).

Комфортность окружения остатков и некоторые параметры боковых цепей оценивались при помощи web-интерфейса программы **WhatCheck** (раздел Structure Validation) [7][8]. Здесь я привожу параметр, который оценивался, выдачу WhatCheck и перевод-интерпретацию:

- 1) **Планарность** (сохраняется ли в модели плоская конфигурация химических групп с плоской геометрией): всё хорошо, выдача WhatCheck ниже:

```
All of the side chains of residues that have an intact planar group are planar within expected RMS deviations. All of the atoms that are connected to planar aromatic rings in side chains of amino-acid residues are in the plane within expected RMS deviations.
```

- 2) **Длины связей** (насколько отличаются длины связей от стандартных длин в белках): всё хорошо, отличия даже меньше чем обычно, поскольку Z-score меньше единицы, выдача WhatCheck ниже:

```
All bond lengths are in agreement with standard bond lengths using a tolerance of four sigma. RMS Z-score for bond lengths: 0.319
```

- 3) **Комфортность окружения боковых радикалов аминокислотных остатков:** было получено среднее значение параметра -0.043 (Z-score = -0.34), что соответствует хорошей структуре (пользуюсь рекомендациями по интерпретации результатов в WhatCheck), кроме того не было найдено ни одного остатка с некомфортным окружением (значение параметра менее чем -2,5).

- 4) **Распределение значений торсионных углов ω :** в норме распределение должно соответствовать гауссовскому с параметрами среднего $+178^\circ$ и стандартным отклонением $5,5^\circ$, для нашей модели выдача WhatCheck: Omega average and standard deviation= 179.700 5.174 Результат соответствует обычному.

- 5) **Конформация пролинового кольца:** обычно пролины имеют конкретный угол отклонения C- γ атома от плоскости кольца (обычно это примерно $+72$ или -90 градусов по углу ϕ). Был обнаружен пролин с не очень типичным углом:

```
516 PRO (75 ) B 100.4 envelop C-beta (108 degrees)
```

Правда авторы сами отмечают, что результатам этого теста не стоит доверять на 100%, и к ним надо относиться очень спокойно.

Маргинальные остатки в структуре. Примеры

В таблице 1 приведена информация о нескольких маргинальных остатках, выделенных нами в предыдущем разделе при рассмотрении параметров качества модели в целом. Для каждого остатка приводится трёхбуквенное сокращение, номер и цепь (А или В), а также причины, заставившие нас обратить внимание на этот остаток.

В подборе примеров маргинальных остатков нам также поможет рисунок из PDBReport, который наглядно демонстрирует отклонения (геометрические) и несоответствие электронной плотности (см. Рис. 7).

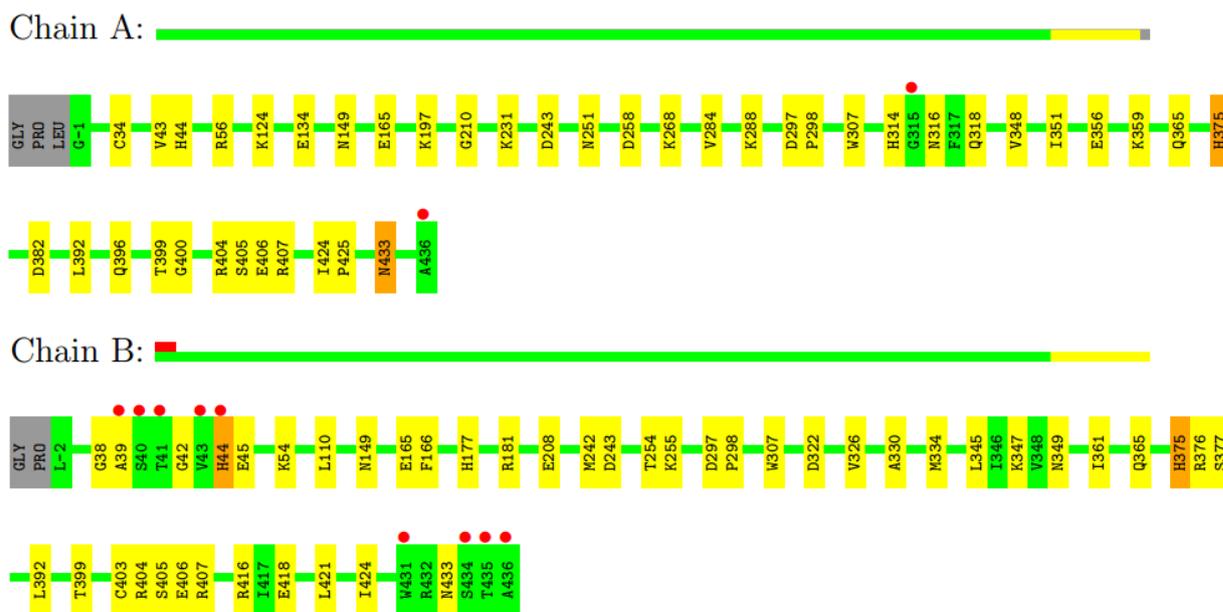


Рисунок 7. Качество отдельных остатков в структуре, маргиналы по одному геометрическому параметру жёлтые, по двум оранжевые. Если остаток плохо описывается исходной электронной плотностью ($RSRZ > 2.0$) рядом с ним ставится красная точка.

Таблица 1. Примеры маргинальных остатков в PCA модели енолазы PDB_ID: 3QTP.

№	Остаток	Цепь	Что обращает внимание на остаток?
1	ARG-404	A	Запрещённая конформация по торсионным углам остова ($\varphi=81.3$, $\psi=121.2$)
2	ALA-39	B	RSRZ outlier (RSRZ=3.5)
3	VAL-43	B	RSRZ outlier (RSRZ=3.5)
4	GLY-315	A	RSRZ outlier (RSRZ=3.0)
5	HIS-44	B	RSRZ outlier (RSRZ=2.9), перекрывание (clash) с Val-326 (цепь B) на 0,68 Å. Маргинал по торсионным углам в боковой цепи остатка (non-rotameric sidechain), вероятно, нужна инверсия боковой цепи
6	ASN-433	A	Маргинал по торсионным углам в боковой цепи остатка (non-rotameric sidechain), перекрывание (clash) с Gln-365 (цепь B) на 0,53 Å.

7	HIS-375	А и В	Маргинал по торсионным углам в боковой цепи остатка (non-rotameric sidechain), перекрывание (clash) His-375 цепи В с ARG-407 (цепь В) на 0,66 Å, вероятно, нужна инверсия боковой цепи
8	PRO-75	В	Не совсем типичный угол (100.4 градуса) в пролиновом кольце конформации конверт.

Анализ ряда маргинальных остатков в структуре

В этом разделе мы проанализируем подробно некоторые пять остатков из приведённых в таблице 1 и попытаемся сделать выводы о том, с чем связана та или иная аномалия в каждом конкретном случае.

HIS 44 (В)

Гистидин 44 расположен в “торце” неструктурированной петли на поверхности глобулы белка и смотрит в растворитель, в исходной модели обнаруживаются два его полярных контакта с молекулами воды (см. Рис. 8а). В поле нашего интереса этот остаток попал по целому ряду причин (см. Таблицу 1), в том числе потому, что плохо описывалась экспериментальная электронная плотность, и потому, что в боковой цепи были не те торсионные углы.

Отдельно заглянем в PDB, где увидим, что изотропный температурный фактор (ADP, Atomic Displacement Parameter) для атомов кольца боковой цепи: CG 51.81, ND1 38.43, CD2 53.44, CE1 43.85, NE2 56.30, больше чем в среднем по окружающим остаткам и белку в целом. Это означает, что остаток более чем в среднем подвержен тепловым колебаниям и электронная плотность сравнительно более «размазана». Некоторые проблемы с электронной плотностью наблюдаются и при увеличении уровня срезки (рис. 8б, 8в).

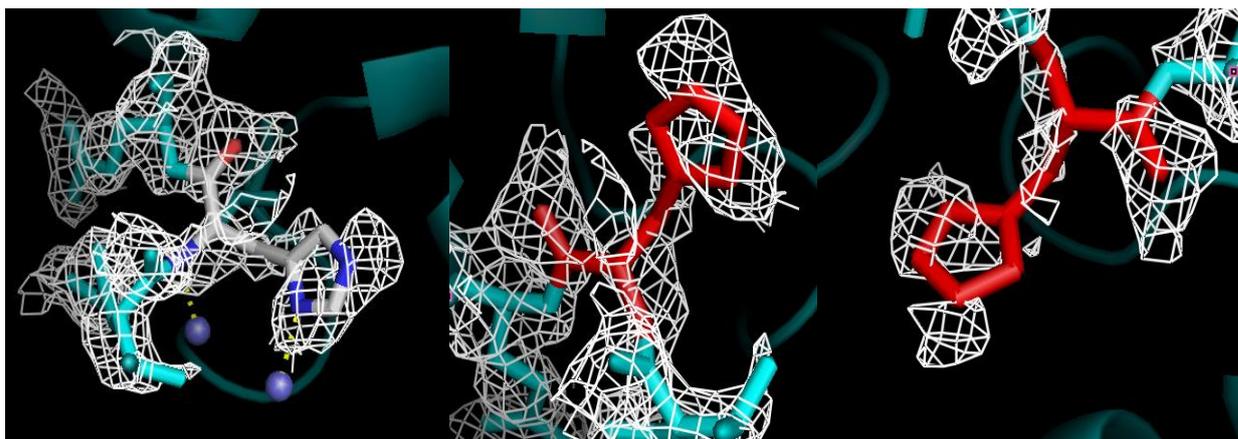


Рисунок 8. HIS 44 (В) Слева направо: **а** – взаимодействия гистидиновой боковой цепи с молекулами воды, **б и в** – электронная плотность вокруг остатка на разных уровнях срезки (**б** – 0,8, **в** – 1,0). Видно, что при увеличении уровня срезки электронная плотность на гистидине фрагментируется.

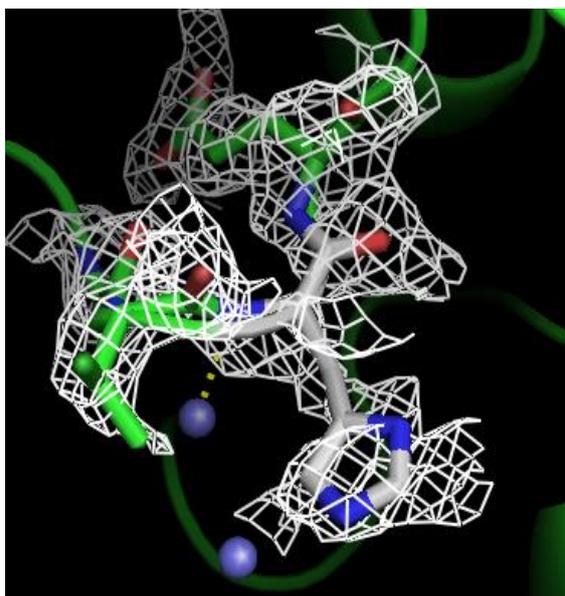


Рисунок 9. HIS 44 (B) в выдаче программы PDB_redo

Из всего сказанного про остаток гистидина можно сделать первый вывод: судя по всему, атомы боковой цепи весьма подвижны и несколько двигаются в пространстве, что не позволило записать чёткую электронную плотность вокруг боковой цепи (даже в сравнении с соседними остатками ЭП смотрится «хуже»). На рисунке 9 представлен тот же участок, но уже в выдаче программы PDB_redo [9]. Видно, что был произведён flip-переход (это можно отследить по расположению двух атомов азота в кольце по сравнению с рисунком 8). При этом теряется полярный контакт (водородная связь с молекулой воды), но вот описание электронной плотности становится более адекватным.

ARG 404 (A)

Аргинин 404 в цепи A лежит на стыке двух субъединиц белка и привлёк наше внимание тем, что был определён как маргинал на основе карты Рамачандрана для всего белка (см. Рис. 4). Попробуем разобраться, что заставляет этот остаток иметь такие неоптимальные значения торсионных углов, лежащие на самой границе допустимой области. На рисунке 10 приводится вид остатка, ЭП вокруг него на уровне срезки 1.0 и водородные связи, которые он может завязывать с окружением. Видно, что ЭП очень хорошо описана вокруг остатка.

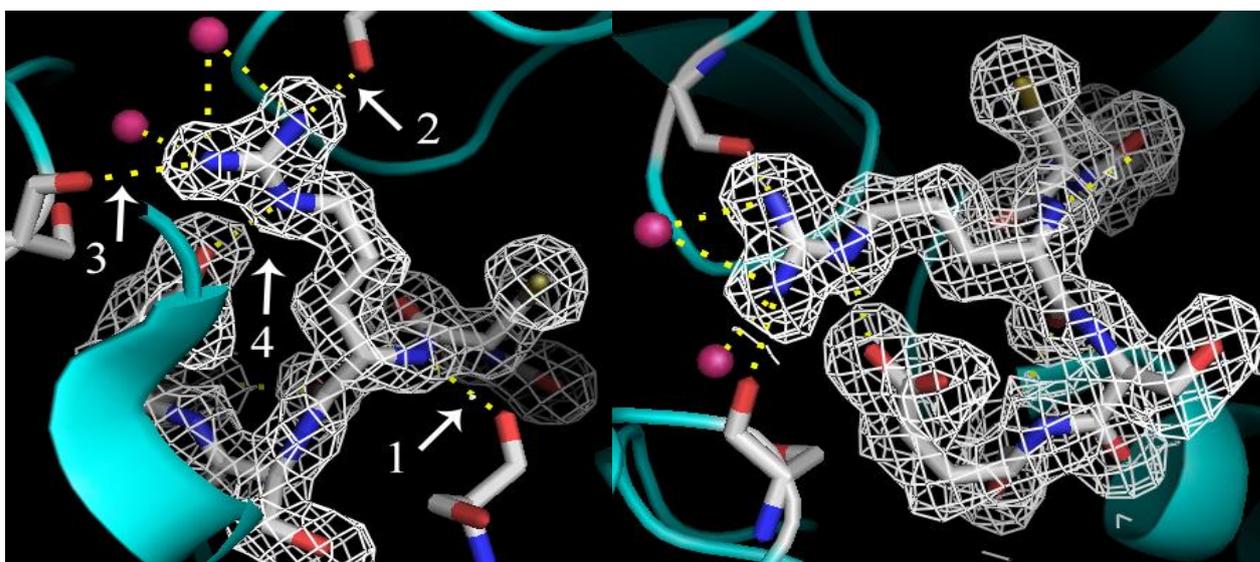


Рисунок 10. Остаток ARG-404 (A). Слева белыми стрелками показаны водородные связи с другими остатками, в том числе с из другой субъединицы (стрелка 1) и очень близки остатком глутамата (стрелка 4). Две молекулы воды, образующие водородные связи с гуанидиновым группой показаны бардовым.

Из рисунка 10 явственно следует, что аргинин участвует в формировании пяти водородных связей с другими аминокислотными остатками и 2 с молекулами воды. Вероятно, вывод по необычности торсионных углов этого остатка должен быть следующим: значения углов лежат на границе допустимой зоны, поскольку энергия, теряемая при не совсем оптимальном расположении остова может быть компенсирована всеми водородными связями, которые при такой конформации завязываются у остатка и его окружения. Другими словами, это скорее необычность, чем ошибка расшифровки.

ASN 433 (A)

Аспарагин 433 цепи А – четвёртая с конца аминокислота в цепи белка, расположена близ С-конца цепи и привлекла наше внимание тем, что была определена как маргинал по торсионным углам в боковой цепи аминокислотного остатка. Температурный фактор весьма небольшой, остаток хорошо вписан в ЭП даже на высоком значении уровня срезки (1,5 σ на рисунке 11). Почему же радикал «загнут» не оптимально? В данном случае я могу опять же объяснить тем, что боковой радикал, взаимодействуя посредством водородных связей с другими участками глобулы, изгибается, принимая не совсем оптимальные значения по углам χ_i . В выдаче PDB_redo также не происходит никаких изменений с этим остатком (см. Рис. 11б).

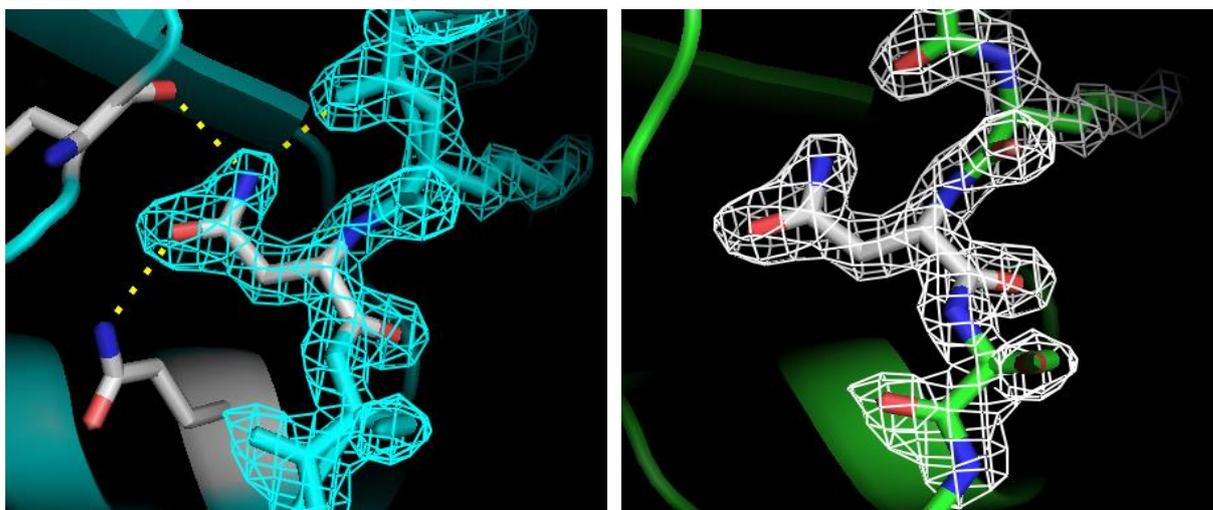


Рисунок 11. Остаток ASN-433 (A). **а** - Слева показаны водородные связи остатка с другими остатками в глобуле белка, которые фиксируют С-конец вблизи основной коровой структуры белка. **б** – тот же участок в выдаче программы оптимизации PDB_redo.

С помощью сервиса MolProbity были добавлены водороды в имеющуюся структуру, на основе полученного файла попробуем проанализировать маргиналов, определённых как имеющих перекрытие с другими атомами (обычно такие перекрытия включают конфликт именно между атомами водорода).

GLN 396 (A) и GLN 318 (A)

Эти два остатка попали в зону нашего внимания благодаря тому, что они оба находятся в первой и третьей строчках таблицы близких взаимодействий (clashes). Причём перекрытие весьма значительно больше чем пороговой значение 0,4Å.

Таблица 2. Выдержка из таблицы перекрытий атомов касательно рассматриваемых остатков.

Atom-1	Atom-2	Distance(Å)	Clash(Å)
1:A:318:GLN:HE22	1:A:396:GLN:HE22	1.07	0.92
1:A:318:GLN:NE2	1:A:396:GLN:HE22	1.86	0.69

Попробуем разобраться в чём тут дело. На рисунке 12 показаны два глутамина в двух вариантах визуализации: слева в виде сфер с атомарным радиусом, хорошо видны перекрытия водородов и кислородов, а справа в стандартной визуализации *sticks*, показаны длины водородных связей.

Вывод здесь напрашивается следующий: в этом месте структуры два глутамина создают мощный мостик, основанный на симметричном полярном взаимодействии, в связи со сближающим действием двух параллельных водородных связей происходит более близкое сближение амидных групп на концах радикалов, что и приводит к перекрытию атомов, детектируемому программами.

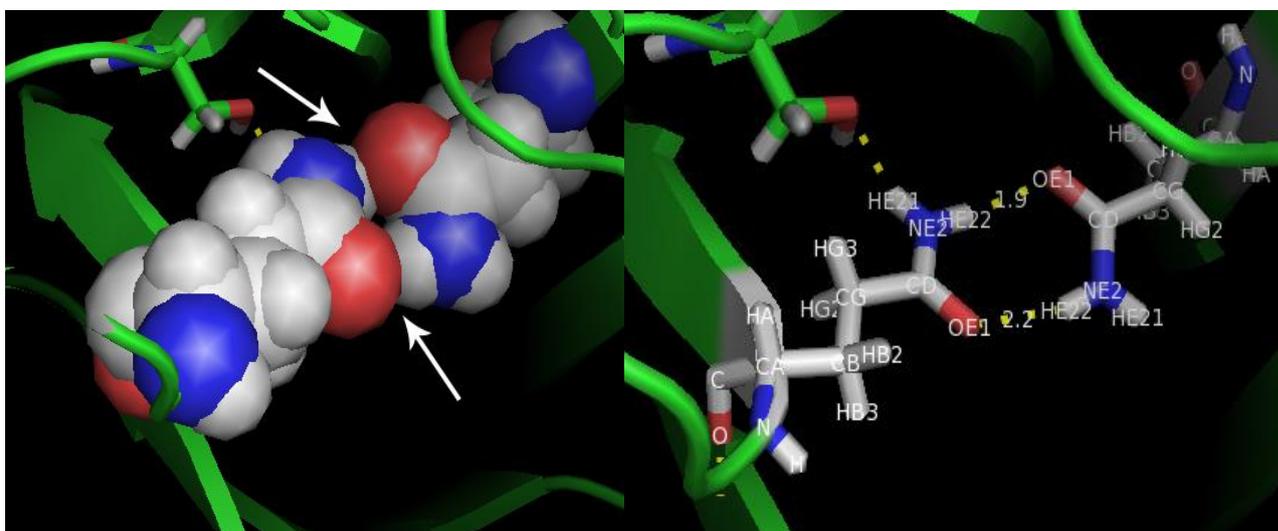


Рисунок 12. Остатки GLN 396 (A) и GLN 318 (A). **а** – (слева) белыми стрелками обозначены две зоны контакта-перекрытия атомов на модели, где атомы представлены своими реальными радиусами. **б** – тот же участок, показаны водородные связи и их длины, подписаны названия атомов.

GLY 315 (A)

Этот глицин является первым аминокислотным остатком после участка стабильной вторичной структуры (α -спирали) и обеспечивает, судя по всему, поворот цепи после спирали. В наше поле зрения он попал как RSRZ outlier (RSRZ=3.0).

На рисунке 13 показан этот участок с электронной плотностью на разных уровнях срезки (слева 1.5σ , справа 1σ). Видно, что глицин хуже вписывается в ЭП при высоком уровне срезки, нежели соседние с ним остатки. Сложно сказать, с чем связано плохое совпадение модели и электронной плотности в этом случае, очевидных ошибок расшифровки тут очевидно нет.

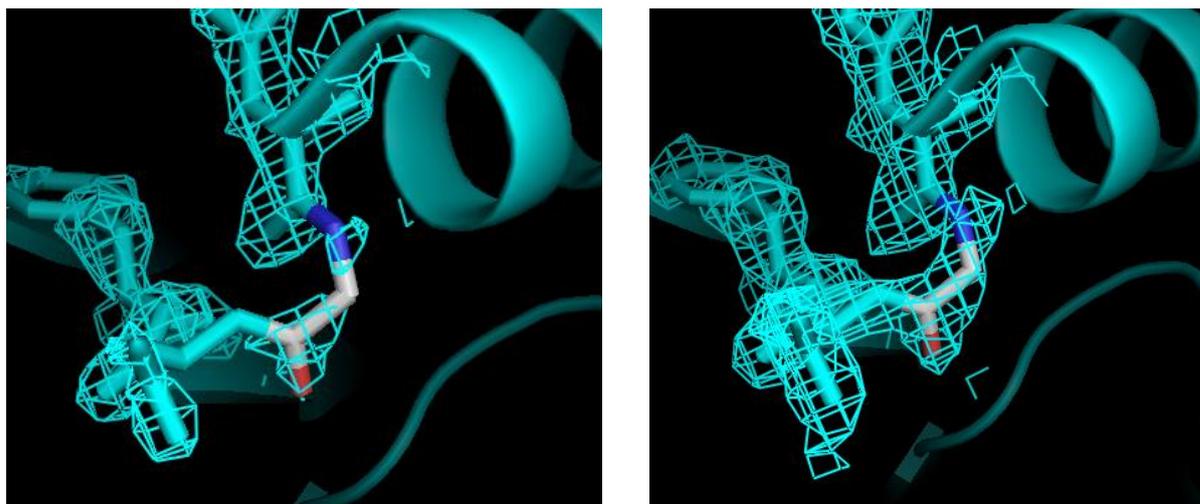


Рисунок 13. Остаток GLY 315 (A) и электронная плотность вокруг модели при разном уровне срезки: слева 1.5σ , справа 1σ .

Сравнение модели из PDB с моделью из PDB_redo

PDB_redo представляет из себя сервис оптимизации, который позволяет улучшить модель с привлечением ряда параметров, которые используются в процессе улучшения. Авторская модель была загружена в программу, скачан файл, содержащий новую модель после оптимизации. Программа сама считает основные параметры до и после оптимизации, а также запускает валидацию WhatCheck по основным параметрам. Все это разом вместе с результатами доступно по ссылке: [9]. Примеры изменений в результате оптимизации уже приводились в предыдущем разделе (см. стр. 11). Непосредственно в отчёте разберу основные параметры качества структуры и проведу визуальное сравнение модели до и после оптимизации методом совмещения.

R-values etc.

	From PDB header	Calculated from data	After conservative optimisation	After full optimisation
R	0.1530	0.1510	0.1430	0.1442
R-free	0.2020	0.1927	0.1809	0.1817

Выше приведены данные из выдачи программы касательно R-фактора и R-free в исходной модели и после двух вариантов оптимизации. Видно, что в результате обработки модели происходит уменьшение обоих параметров, а в добавок уменьшается и разница между ними, что свидетельствует о реальной оптимизации модели, а не о подгонке многих параметров.

На рисунке 14 показано пространственное совмещение двух моделей в двух видах визуализации. Видно, что модели в целом сходны и повторяют одна другую, а отличия заключаются в мелких поворотах боковых (пример см. на рис. 15).

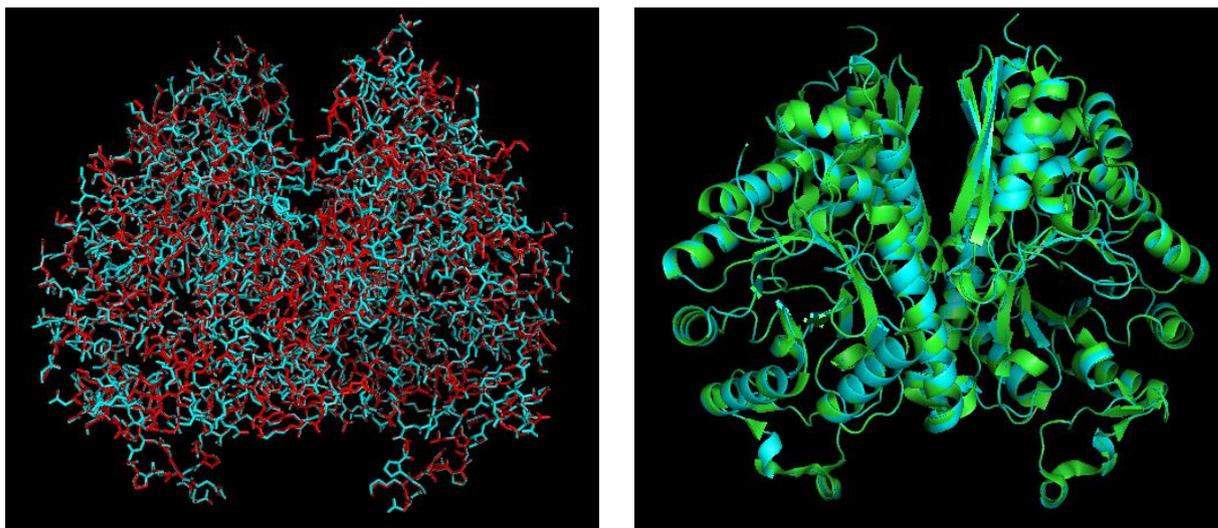


Рисунок 14. Пространственное совмещение структур до и после полной оптимизации программой PDB_redo.

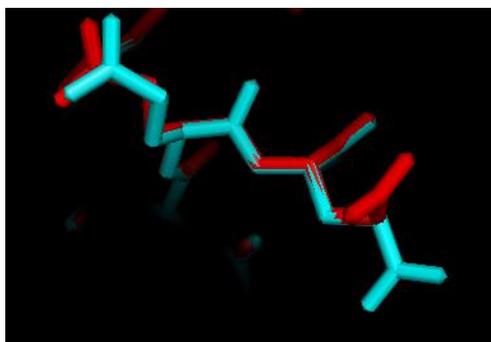


Рисунок 15. Пример бокового радикала, который был повернут программой PDB-redo (красная структура – до оптимизации, голубая – после).

Заключение о качестве расшифровки структуры

Исходя из вышеприведённой информации, могу отметить, что качество расшифровки структуры енолазы можно считать хорошим. Основные параметры качества структуры соответствуют принятым стандартам, многие из них имеют более хорошие значения по сравнению с другими структурами такого разрешения в базе данных PDB, маргинальных остатков не очень много, а те, которые были нами рассмотрены, стали маргиналами скорее в силу особенностей структуры, нежели из-за очевидных ошибок в расшифровке структуры. Повторная оптимизация модели (PDB_redo), хотя и улучшила ряд параметров, но принципиальных изменений в неё не внесла. Итак, хорошая качественная структура.

Список использованной литературы и web-сервисов

- [1] E. C. Schulz, M. Tietzel, A. Tovy, S. Ankri, and R. Ficner, "Structure analysis of *Entamoeba histolytica* enolase.," *Acta Crystallogr. D. Biol. Crystallogr.*, vol. 67, no. Pt 7, pp. 619–27, Jul. 2011.
- [2] W. A. Petri, "Therapy of intestinal protozoa.," *Trends Parasitol.*, vol. 19, no. 11, pp. 523–6, Nov. 2003.
- [3] "RSCB PDB: 1E9I." [Online]. Available: <http://www.rcsb.org/pdb/explore/explore.do?structureId=1E9I>.
- [4] "EDS: PDB entry 3qtp." [Online]. Available: <http://eds.bmc.uu.se/cgi-bin/eds/uusfs?pdbCode=3qtp>.
- [5] "RSCB PDB: 3QTP." [Online]. Available: <http://www.rcsb.org/pdb/explore/explore.do?structureId=3QTP>.
- [6] "MolProbity." [Online]. Available: <http://molprobity.biochem.duke.edu/>.
- [7] "WhatCheck: Structure validation." [Online]. Available: <http://swift.cmbi.ru.nl/servers/html/index.html>.
- [8] R. W. Hoofst, G. Vriend, C. Sander, and E. E. Abola, "Errors in protein structures.," *Nature*, vol. 381, no. 6580, p. 272, May 1996.
- [9] "Выдача PDB_redo." [Online]. Available: http://www.cmbi.ru.nl/pdb_redo/qt/3qtp/index.html.