

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

Отчёт по качеству расшифровки структуры дезоксиуридин-5'-трифосфат-
нуклеотидгидролазы из *Escherichia coli* (PDB ID: 1DUP) методом
рентгеноструктурного анализа

Анна Желтова,
4 курс ФББ МГУ

Москва, 2017 г.

Аннотация

В отчёте рассмотрена общая информация о ферменте и модели, представлены значения индикаторов качества модели, проведён анализ маргинальных аминокислотных остатков и сравнение модели из PDB с моделью из PDB_redo.

Введение

В данном отчете оценивается качество PCA расшифровки структуры дезоксиуридин-5'-трифосфат-нуклеотидгидролазы (дУТФаза) *Escherichia coli* (DEOXYURIDINE 5'-TRIPHOSPHATE NUCLEOTIDO HYDROLASE (D-UTPASE)).

Идентификатор в БД PDBе - [1DUP](#).

Фермент дУТФаза катализирует гидролиз дУТФ (рис.1.) и поддерживает низкую внутриклеточную концентрацию дУТФ. дУТФаза, превращает дУТФ, являющийся субстратом ДНК полимеразы, в дУМФ, который уже не является субстратом ДНК полимеразы. дУМФ используется для синтеза тимидиловых нуклеотидов, поскольку превращается сначала в дТМФ, а потом – в дТТФ. Таким образом, это приводит к тому, что урацил не может быть включен в ДНК.

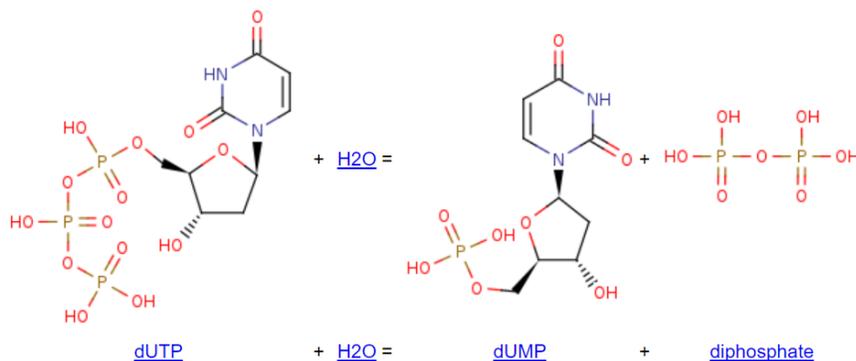


Рис.1. Реакция, катализируемая дУТФазой. БД BRENDA [1]

dUTPase из *Escherichia coli* строго специфична для ее субстрата dUTP за счет активного центра, различающего нуклеотиды по отношению к сахарному фрагменту, а также пиримидинового основания (рис.2.).

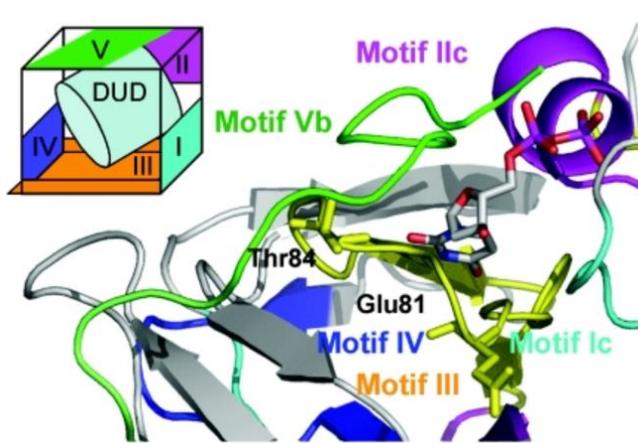


Рис.2. Активный сайт с dUDP (DUD) и схематическая диаграмма активного сайта (вставка). dUDP используется для моделирования фермента, потому что это неконкурентный ингибитор, который локализуется на активном сайте. [2]

48 структур были решены для этого класса ферментов:

1DUC, 1DUD, 1DUN, 1DUP, 1DUT, 1EU5, 1EUW, 1F7D, 1F7K, 1F7N, 1F7O, 1F7P, 1F7Q, 1F7R, 1MQ7, 1OGH, 1OGK, 1OGL, 1PKH, 1PKJ, 1PKK, 1RN8, 1RNJ, 1SEH, 1SIX, 1SJN, 1SLH, 1SM8, 1SMC, 1SNF, 1SYL, 1VYQ, 1W2Y, 2BSY, 2BT1, 2CJE, 2D4L, 2D4M, 2D4N, 2HQU, 2HR6, 2HRM, 2OKB, 2OKD, 2OKE, 2OLO, 2OL1, 2PY4.

Для дУТФазы E.coli известны следующие структуры дУТФазы:

1. 1DUD [3]
2. 1DUP [4]
3. 1EU5 [5]
4. 1EUW [5]
5. 1RN8 [6]
6. 1RNJ [6]
7. 1SEH [6]
8. 1SYL [6]
9. 2HR6 [7]

10. [2HRM](#) [7]

Рассматриваемая кристаллическая трехмерная структура 1DUP была получена группой ученых в составе: Cedergren-Zeppezauer E.S., Larsson G., Nyman P.O., Dauter Z., Wilson K.S. Статья «Crystal structure of a dUTPase» была опубликована в Nature [4].

В статье рассматривают трехмерную структуру DUTPase E. coli, определенную методом рентгеновской кристаллографии с разрешением 1,9 Å. Фермент представляет собой симметричный тример, 152 аминокислотных остатка в субъединице, первые 136 видны в кристаллической структуре. Третичная структура напоминает "jelly-roll" складку и не показывает «классический» нуклеотидсвязывающий домен. В четвертичной структуре существует сложное взаимодействие между субъединицами, которое может иметь важное значение для катализа.

Результаты и обсуждение

Общая информация о модели

Общая информация о модели представлена в таблице 1.

Таблица 1. Общая информация о модели 1DUP

Состав комплекса	Цепь белка 136 аминокислот, 123 молекулы воды
Год	1995
Авторы	Z.Dauter,K.S.Wilson,G.Larsson,P.O.Nyman,E.Cedergren
Метод решения фазовой проблемы	К сожалению, не удалось найти ни в записи, ни в статье (в статье имеется упоминание «Details of the crystallography work will be published elsewhere», однако этой статьи мной не было найдено).
Число измеренных рефлексов	13597
Разрешение	1.90 Å
Полнота набора структурных факторов	99.4%
Диапазон разрешений структурных факторов	1.9Å - 62.27Å
Параметры кристаллографической ячейки	Длина (Å): a: 86.61 Å b: 86.61 Å c: 62.27 Å Угол (°): α: 90° β: 90° γ: 120°
Наличие некристаллографических симметрий в асимметрической ячейке	Не обнаружены

Значения индикаторов качества

Индикаторы качества модели служат показателем того, насколько модель соответствует действительности. С помощью сервиса MolProbity [8] были установлены следующие индикаторы качества, представленные в таблице 2.

Таблица 2. Индикаторы качества модели 1DUP ★

	Значение	Процент/ Перцентиль
R-factor	0.150	
R_free	Not available*	
RSR	0.056	5.6%
ClashScore (число недопустимых наложений атомов на 1000)	6.4	93 rd percentile* (N=773, 1.90Å ± 0.25Å) **
MolProbity score (интегральная оценка структуры по данным MolProbity [8]: combine clashscore, rotamer, and Ramachandran evaluations into a single score)	2.08	62 nd percentile* (N=12147, 1.90Å ± 0.25Å) ***
Ramachandran Plot statistics ▲		
Most favoured regions	105	93.8%
Additional allowed regions	5	4.5%
Disallowed regions	2	1.8%
Ramachandran outliers	0	0%
Poor rotamers (число остатков с маргинальными по отклонению от ротамеров боковыми цепями)	6	5.36% 🟢
Ramachandran favored (число и процент остатков в предпочитаемой области)	130	97.1%
Non-glycine and non-proline residues	112	
End-residues (excl. Gly and Pro)	2	
Glycine residues	13	
Proline residues	19	
G-Factor 🟡		
Main-chain covalent forces:		
Main-chain bond lengths	0.04	
Main-chain bond angles	-1.50 🌞	
Average score	-0.86 🌞🌞	
Sidechain outliers		
Analysed	112	
Rotameric	104	93%

Outliers	8	7%
Protein Geometry		
Cis Prolines	0	0% ●
C β deviations >0.25Å (число C β -атомов с неприемлемым отклонением от ожидаемого положения)	9	7.32% ●
Bad bonds	7 / 1047	0.67% ★
Bad angles	59 / 1422	4.15% ★★

★ - Полноту наличия индикаторов качества модели можно считать неполной, так как не было найдено никаких иных индикаторов качества при том, что поиск был проведен в различных источниках.

*- данных не было найдено ни в записи, ни с помощью сервиса MolProbity[8].

** - Перцентиль данной структуры по отношению к структурам примерно такого же разрешения: 93% структур ClashScore хуже (больше), чем у данной. Что говорит об очень хорошем качестве модели.

*** - Перцентиль данной структуры по отношению к структурам примерно такого же разрешения: 62% структур MolProbity score хуже (больше), чем у данной. Хотя значение и «не отличное», качество модели нормальное.

▲ - Основываясь на анализе 118 структур с разрешением не менее 2.0 ангстрем и R-факторе не более 20.0, ожидается, что модель хорошего качества будет иметь более 90% в Most favoured regions. Для нашей модели данный показатель = 93.8%, что свидетельствует о хорошем качестве модели.

◆ - Goal: <0.3%, из чего можно сделать вывод, что число остатков с маргинальными по отклонению от ротаметров боковыми цепями превышает предполагаемое хорошее значение 0.3% почти в 18 раз.

⬡ - Соотношения длин и связей основной цепи сравниваются с идеальными значениями Engh & Huber (1991), полученными из данных о малых молекулах. Поэтому структуры, уточненные с использованием разных ограничений, могут иметь явно большие отклонения от нормы.

☀ - Значения ниже -1,0 - очень необычные. Однако принимая во внимание предыдущее примечание, делать выводы на данном этапе рано.

☀️☀️ - Значения ниже -0,5 * - необычные. Но делать выводы на данном этапе еще рано.

● - Меньше ожидаемого значения 5%.

● - Goal: 0, что говорит о плохом качестве модели.

★ - Goal: 0%, что свидетельствует о том, что качество модели недостаточно хорошее.

★★ - Goal: <0.1%, что свидетельствует о том, что качество модели недостаточно хорошее.

Отсутствие маргиналов говорит о хорошем качестве модели, однако, учитывая наличие неполного набора индикаторов, отклонение некоторых значений от нормали и плохие значения для Poor rotamers, C β deviations >0.25Å, Bad bonds и Bad angles, данная гипотеза нуждается в проверке.

Попробуем проверить правильность гипотезы, используя EDS [9]. Найдем Real Space R-factor (RSR). Маргиналы детектируются по соответствию боковой цепи экспериментальной электронной плотности. Рассматривается сумма по узлам пространственной решетки в окружении группы атомов. Для того, чтобы признать качество модели хорошим, значение RSR должно быть меньше 20%.

Найденное значение RSR = 0.056, что составляет 5.6 %, что значительно меньше 20% и свидетельствует о хорошем качестве модели.

Рассмотрим гистограмму значений RSR для всех остатков в структуре 1DUP (рис.3).

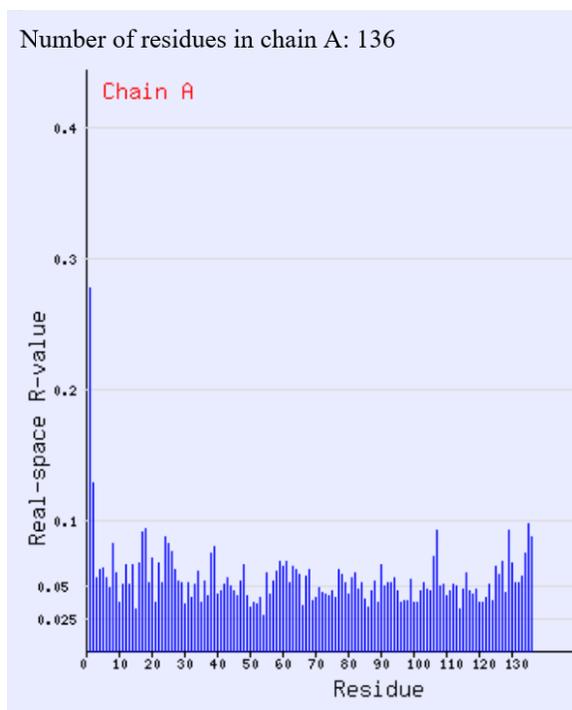


Рис.3. Гистограмма значений RSR для всех остатков в структуре 1DUP [10]. По оси Y отложены значения RSR, по оси X - номера остатков. Значения, которые превышают 0.2 считаются маргинальными.

Рассмотрим относительный R-фактор - Z-score. Z-score вычисляется относительно всех остатков такого же типа из структур с таким же разрешением. Значения $Z > 2$ свидетельствуют о том, что остаток плохо вписан в электронную плотность по сравнению с другими структурами с тем же разрешением. Что свидетельствует о том, что данный остаток скорее всего является маргиналом. Если RSR плохой, а Z-score – хороший, то значит координаты атомов расшифрованы плохо, но не хуже, чем в других подобных структурах. Значения Z-score для приведены на рисунке 4.

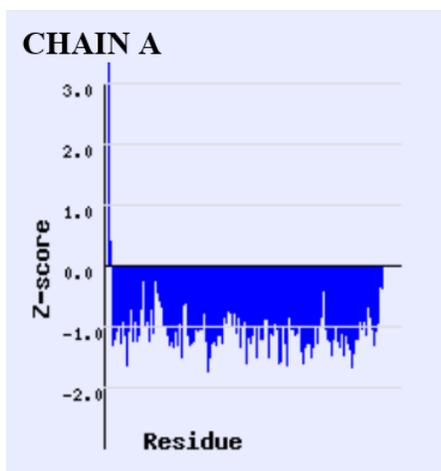


Рис.4. Значения Z-score [11]. По оси Y отложены значения Z-score, по оси X отложены номера остатков. Остатки, значение для которых превышает 2.0 считаются маргинальными.

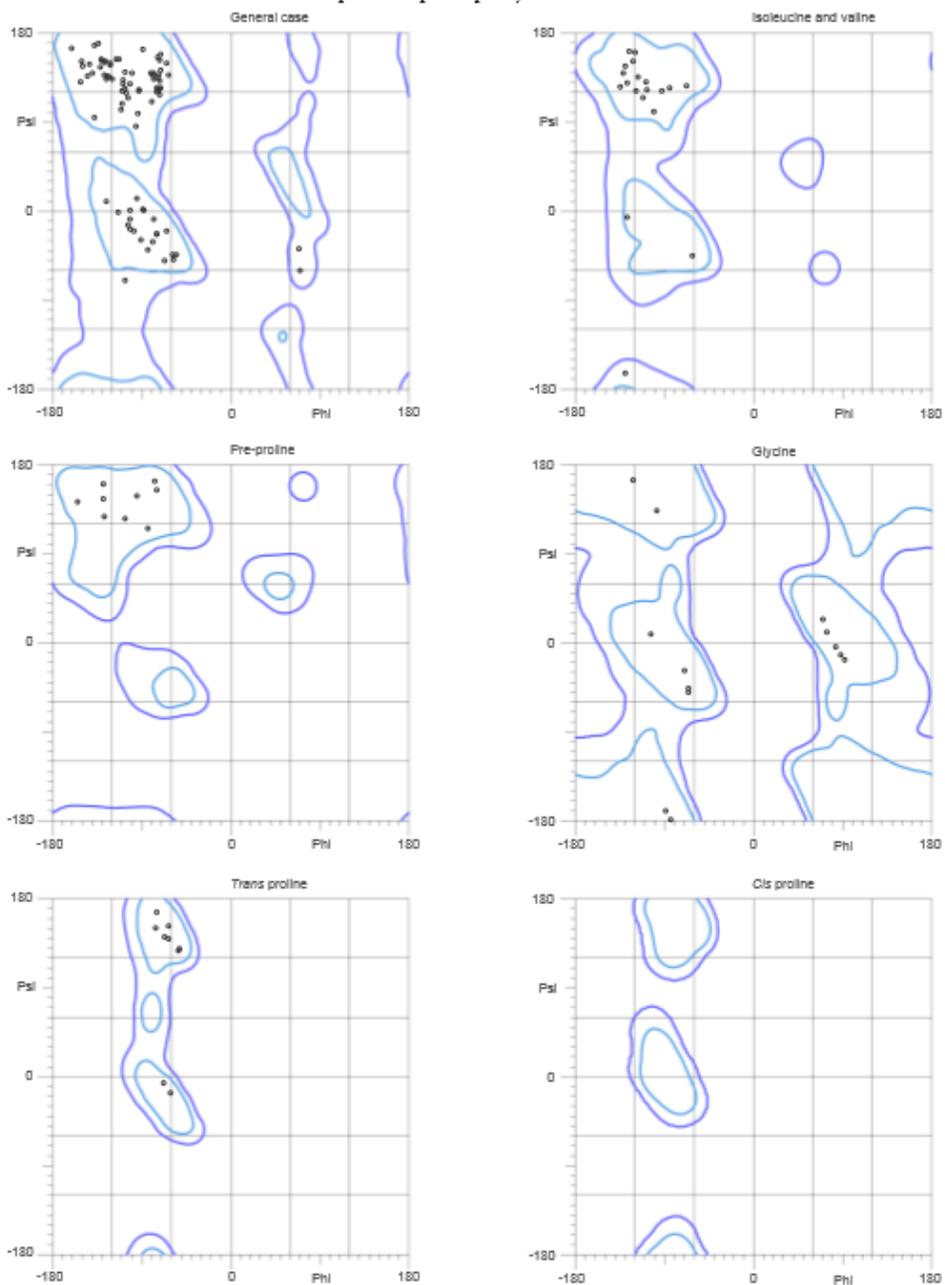
Итак, согласно данным по Z-score и RSR, был найден один маргинальный остаток 1 MET, Chain A.

Также определение наличия маргинальных остатков возможно с помощью карты Рамачандрана. На горизонтальной оси карты откладывается угол Phi, а на вертикальной - угол Psi. Карты Рамачандрана можно получить как с помощью EDS [9], так и с помощью MolProbity [8], а также PDBsum (PROCHECK analyses) [12].

На рисунке 5 представлена карта Рамачандрана, полученная с помощью MolProbity [8]. Маргинальных остатков не было обнаружено. Обнаружено три наименее предпочтительных остатка, которые не являются маргиналами.

MolProbity Ramachandran analysis

pdb1dupFH.pdb, model 1



97.0% (130/134) of all residues were in favored (98%) regions.

100.0% (134/134) of all residues were in allowed (>99.8%) regions.

There were no outliers.

Рис.5. Карта Рамачандрана, полученная с помощью MolProbity [8]

На рисунке 6 представлена карта Рамачандрана, полученная с помощью EDS[9]. Было найдено три маргинальных остатка, представленных в таблице 3.

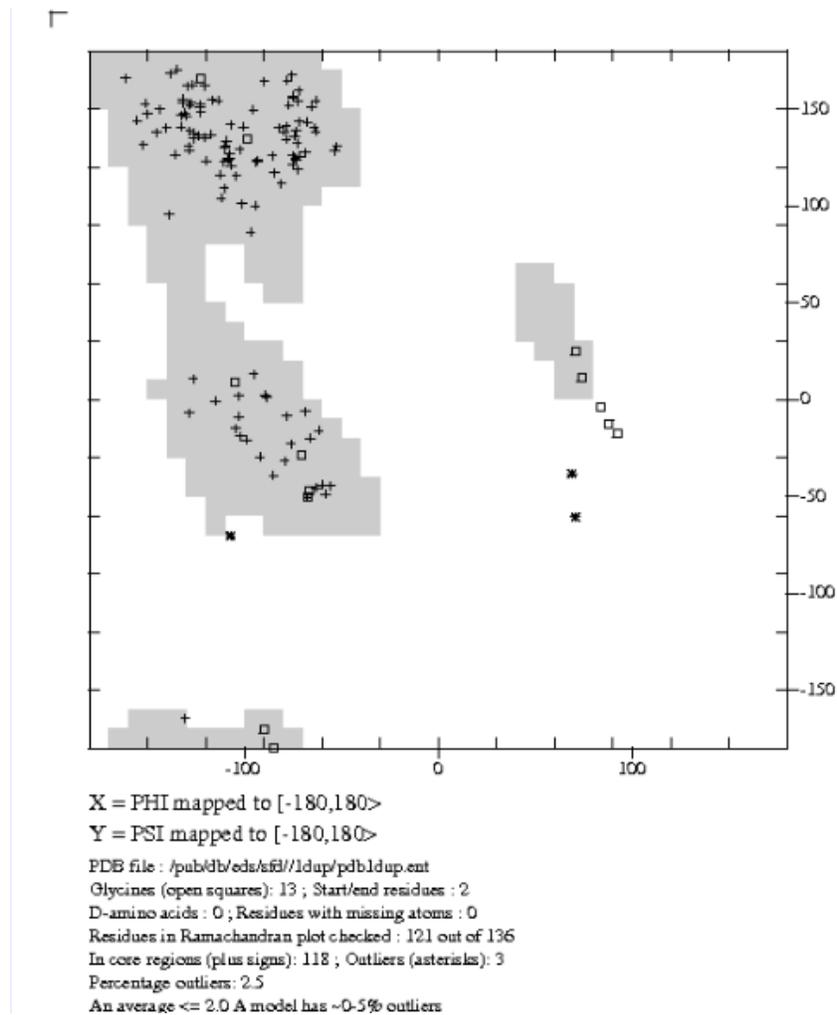


Рис.6. Карта Рамачандрана, полученная с помощью EDS[9].

Таблица 3. Маргинальные остатки, найденные с помощью EDS[9].

Residue	Type	Phi Angle	Psi Angle
A16	LYS	-107.0	-70.7
A60	ALA	69.1	-38.4
A85	LEU	70.9	-60.8

На рисунке 7 представлена карта Рамачандрана, полученная с помощью PDBsum (PROCHECK analyses) [12]. Было найдено два маргинальных остатка, представленных в таблице 4.

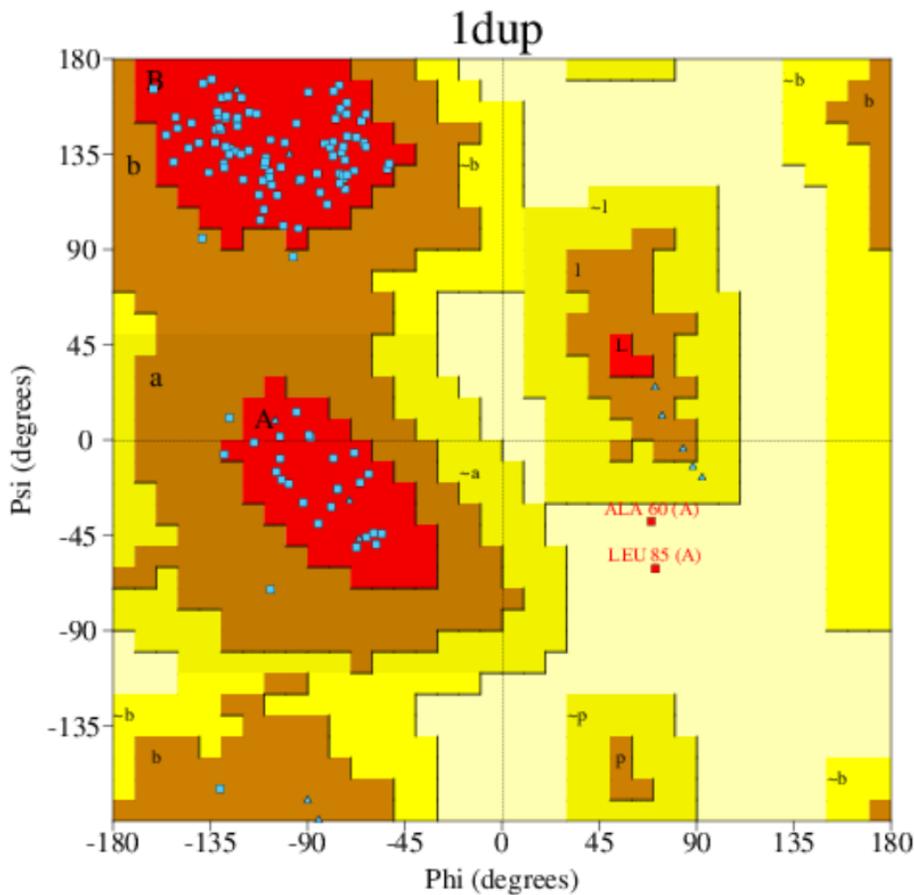


Рис.7. Карта Рамачандрана, полученная с помощью PDBsum (PROCHECK analyses) [12]

Таблица 4. Маргинальные остатки, найденные с помощью PDBsum (PROCHECK analyses).

Residue	Type	Phi Angle	Psi Angle
A60	ALA	69.1	-38.4
A85	LEU	70.9	-60.8

Как следует из данных по Z-score и RSR, а также карт Рамачандрана, полученных различными методами, были найдены различные маргинальные остатки. Так данные по Z-score и RSR позволили найти только один

маргинальный остаток - 1 MET, Chain A. Карта Рамачандрана, полученная с помощью MolProbity [8] не выявила маргинальных остатков, а карты Рамачандрана, полученные с помощью EDS[9] и PDBsum (PROCHECK analyses) [12] дали схожие результаты, однако EDS[9] выявил на один остаток больше. Разное число маргинальных остатков, выявленных различными методами, возможно может быть связано с различными пороговыми значениями при определении маргиналов и использованием разных сервисов.

Список маргинальных остатков

Таблица 5 содержит список маргинальных остатков с указанием показателя, по которому они отобраны.

Таблица 5. Список найденных маргинальных остатков с указанием показателя, по которому они отобраны.

Residue	Type	Phi Angle	Psi Angle	Indicator by which they are selected
A16	LYS	-107.0	-70.7	Находятся в запрещенной области карты Рамачандрана
A60	ALA	69.1	-38.4	
A85	LEU	70.9	-60.8	
A1	MET			Плохое значение RSR факторов
A84	ASN			Clashscore >0.4, = 0.69Å
A123	VAL			Clashscore >0.4, = 0.64Å

Анализ выбранных маргинальных остатков

Проанализируем выбранные маргинальные остатки. Вычислим функцию электронной плотности и рассмотрим остатки.

Используемые для визуализации команды:

```
load pdb1dup.ent # загружаем структуру из PDB файла
```

```
load 1dup.csr4 # загружаем карту электронной плотности
```

```
hide (resn HOH)
```

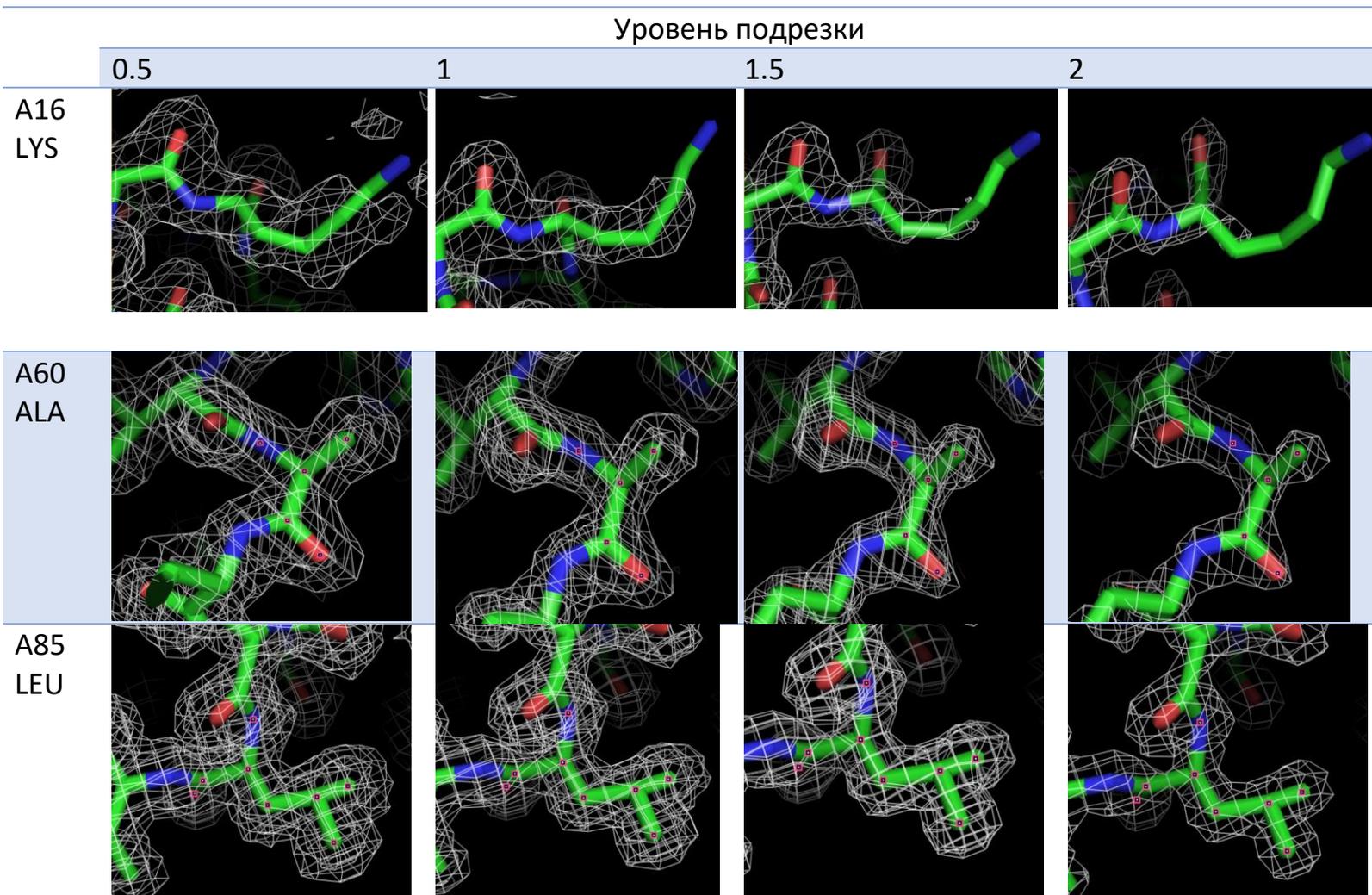
```
show sticks
```

isomesh new05_map, 1dup, 0.5, visible, 2.5, carve=2.5 # Используем несколько уровней подрезки 0.5, 1, 1.5, 2.

zoom resi 16 and chain A # указываем номер остатка и цепь

В таблице 6 представлены изображения электронной плотности, соответствующие различным уровням подрезки для выбранных остатков, которые являются маргинальными согласно ранее полученным данным.

Таблица 6. Изображения электронной плотности, соответствующие различным уровням подрезки для выбранных остатков



Визуализировав предположительно маргинальные остатки, можно сделать следующие выводы:

- 1) LYS A16, найденный сервисом EDS[9] можно считать маргинальным остатком, так как трудно идентифицировать не только расположение данного остатка в белке, но и тип остатка. Аналогичные выводы можно сделать для маргинального остатка MET A1, вычисленного по данным Z-score и RSR.
- 2) А вот маргинальные остатки ALA A60 и LEU A85, вычисленные по картам Рамачандрана с помощью PDBsum (PROCHECK analyses) [12] и EDS[9], вероятно таковыми не являются, так как их расположение и тип можно идентифицировать на различных уровнях подрезки.
- 3) Что касается маргиналов ASN A84 и VAL A123, можно отметить, что стерических затруднений не наблюдается, расстояния между атомами больше, чем их Ван-дер-Ваальсовы радиусы, что говорит о том, что это не маргинальные остатки и вероятно ошибка в вычислениях сервиса.
- 4) Таким образом, наиболее точные, на мой взгляд, данные об маргинальных остатках могут быть найдены только с использованием комбинации сервисов. Наиболее «близкими к истине» оказались данные Z-score и RSR, а так же MolProbity [8], который, хоть и не нашел ни одного маргинального остатка по карте Рамачандрана, сделал «меньший процент ошибок».

Сравнение модели из PDB с моделью из PDB_redo [13]

Выполнили сравнение модели 1DUP из PDB с оптимизированной моделью из PDB_REDO. Полученные результаты приведены в таблице 7.

Таблица 7. Значения некоторых параметров для оценки качества двух моделей [13]. Индикаторы, значение которых считается «хорошим», выделены зеленым цветом, а индикаторы, значение которых считается «плохим» – красным.

Validation metrics from PDB-REDO		
	PDB	PDB-REDO
Crystallographic refinement		
R	0.1431	0.1252
R-free	0.1706	0.1592
	R-free was considered biased, the estimated unbiased R-free was used.	
Bond length RMS Z-score	0.977	0.722
Bond angle RMS Z-score	1.839	0.835
Model quality (raw scores percentiles)		
Ramachandran plot appearance	0.769 91	0.903 93
Rotamer normality	-2.820 25	0.624 95
Coarse packing	-2.442 N/A	-2.499 N/A
Fine packing	-2.220 7	-2.514 4
Bump severity	0.005 90	0.000 99
Hydrogen bond satisfaction	0.948 97	0.966 100

На рисунке 8 представлено качество модели по сравнению с моделями, сходными по разрешению.

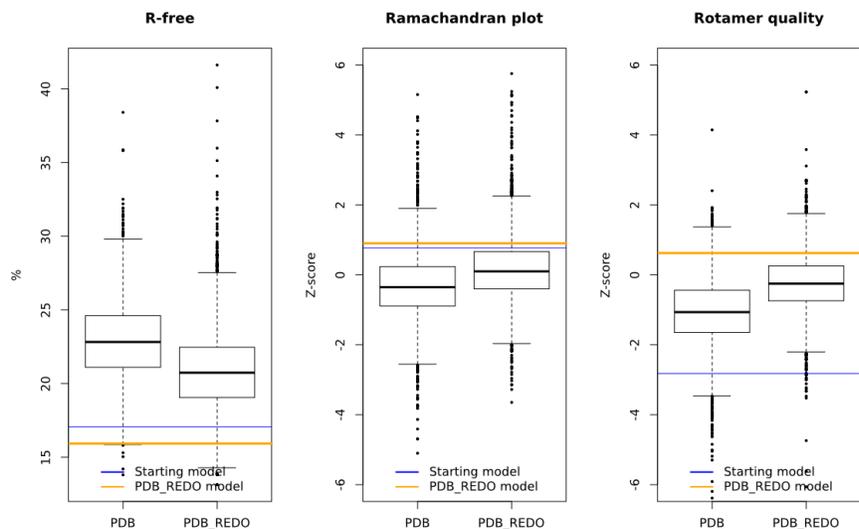


Рис.8. Качество модели по сравнению с моделями, сходными по разрешению

Можно сделать следующие выводы:

- 1) Значения R-фактора и R-free улучшились. Это свидетельствует о том, что новая модель лучше соответствует экспериментальным данным.
- 2) В целом, значения параметров улучшились.

Заключение

Рассмотренная модель 1DUP обладает от среднего до хорошего качеством разрешения (в зависимости от того, по какому индикатору качества оценивать). Встречаются как очень хорошие значения показателей, так и плохие. При этом, хотелось бы отметить, что при выполнении задания по электронной плотности в практикуме №1, я сделал вывод, что разрешения данной модели не достаточно для того, чтобы можно было легко восстановить точное расположение атомов, однако это возможно сделать. Как показало более подробное изучение структуры, восстановление точного расположения атомов действительно возможно, число маргинальных остатков минимально, что свидетельствует в пользу хорошего качества модели. PDB_REDO позволил улучшить показатели модели, что является несомненным плюсом.

Список используемой литературы

- [1] <http://www.brenda-enzymes.org/enzyme.php?ecno=3.6.1.23>
- [2] Zhang Y. et al., 2005, DOI: [10.1128/JVI.79.15.9945-9953.2005](https://doi.org/10.1128/JVI.79.15.9945-9953.2005)
- [3] G.Larsson et al., 1996, PubMed id: [8646539](https://pubmed.ncbi.nlm.nih.gov/8646539/)
- [4] E.S.Cedergren-Zeppezauer et al., 1992, PubMed id: [1311056](https://pubmed.ncbi.nlm.nih.gov/1311056/)
- [5] A.González et al., 2001,
PubMed id: [11375495](https://pubmed.ncbi.nlm.nih.gov/11375495/) DOI: [10.1107/S0907444901004255](https://doi.org/10.1107/S0907444901004255)
- [6] O.Barabás et al., 2004, PubMed id: [15208312](https://pubmed.ncbi.nlm.nih.gov/15208312/) DOI: [10.1074/jbc.M406135200](https://doi.org/10.1074/jbc.M406135200)
- [7] J.Kovári et al., 2008, PubMed id: [17932923](https://pubmed.ncbi.nlm.nih.gov/17932923/) DOI: [10.1002/prot.21757](https://doi.org/10.1002/prot.21757)
- [8] <http://molprobity.biochem.duke.edu/>
- [9] <http://eds.bmc.uu.se/cgi-bin/eds/uusfs?pdbCode=1dup>
- [10] <http://eds.bmc.uu.se/cgi-bin/eds/realr?pdbCode=1dup&type=1&NBINS=50&SPACE=2>
- [11] <http://eds.bmc.uu.se/cgi-bin/eds/zscore?1dup>
- [12] https://www.ebi.ac.uk/thornton-srv/databases/cgi-bin/pdbsum/GetPage.pl?pdbcode=1dup&template=procheck_summary.html
- [13] <https://pdb-redo.eu/db/1dup>