

МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ИМЕНИ М. В. ЛОМОНОСОВА

ФАКУЛЬТЕТ БИОИНЖЕНЕРИИ И БИОИНФОРМАТИКИ

Радкевич Эмир, 401 группа

Анализ качества структуры UvrD *Escherichia coli*
PDB ID - 3LFU

Москва, 2018г

Содержание

1	Аннотация	2
2	Введение	2
3	Результаты	3
3.1	Основные показатели качества структуры	3
3.1.1	R-фактор и R-free	3
3.1.2	Clashscore	5
3.1.3	Карта Рамачандрана	5
3.1.4	Ротамеры	6
3.1.5	RSRZ	6
3.2	Анализ маргинальных остатков	8
3.2.1	Gln546	8
3.2.2	Asn164 и Pro165	10
3.2.3	Leu6	10
3.2.4	Val411 и Phe407	11
3.2.5	Pro414	12
3.3	Улучшение параметров модели при помощи сервиса PDB- REDO	13
4	Выводы	16

1 Аннотация

Был проведен анализ качества структуры UvrD *Escherichia coli* с PDB ID 3LFU. Были найдены потенциальные маргинальные остатки при помощи анализа структуры различными критериями. После аннотации маргинальных остатков было проведено улучшение параметров модели при помощи сервиса PDB-REDO [3].

2 Введение

Escherichia coli UvrD представляет собой ДНК-геликазу и одноцепочечную (оц) ДНК-транслоказу, которая функционирует при репарации ДНК, репликации плазмид и в качестве анти-рекомбиназы путем удаления белка RecA из оцДНК [2]. UvrD связывает и гидролизует АТФ, чтобы размотать двухцепочечную ДНК и "скользить" вдоль оцДНК по направлению от 3'-конца к 5'-концу. Хотя мономер UvrD способен перемещаться вдоль оцДНК быстро, ДНК-геликазная активность *in vitro* требует, как минимум димера UvrD.

Предыдущие кристаллические структуры UvrD, связанные с одноцепочечными дуплексами ДНК, показывают, что его 2В субдомен существует в "закрытом" состоянии и взаимодействует с дуплексом ДНК.

Для того, чтобы изучить ротационные конформационные изменения, происходящие с 2В субдоменом, были сконструированы двойные цистеиновые мутанты UvrD и помечены флуорофорами таким образом, что вращение субдомена приводило к флуоресцентному резонансному переносу энергии (FRET).

Таким образом была получена структура с PDB ID: 3LFU, в которой субдомен 2В находится в "открытом" состоянии, которое отличается от UvrD в "закрытом" состоянии поворотом 2В субдомена на 160°[2].

3 Результаты

Структура 3LFU была загружена на сайт PDB в 2010 году и опубликована в 2011. При этом, в 2017 году было добавлено описание к полю refinement. Авторы модели: Korolev S, Waksman G, Lohman TM. Разрешение структуры составляет 1.8 Å.

Параметры кристаллографической ячейки модели: длины направляющих векторов кристалла (в Å) – 194.543, 58.739, 69.418; углы между ними (в °) – 90.00, 90.00, 90.00; кристаллографическая группа – P 21 21 2. Количество цепей в элементарной ячейке – 4.

В ходе рентгеноструктурного анализа было измерено 67971 рефлексов, при этом полнота набора структурных факторов составила 91 %, а диапазон разрешений – 1.80-29.54 Å.

Фазовая проблема была решена методом молекулярного замещения. Асимметрическая единица модели содержит: 1 молекулу белка (5039 атомов), 4 молекулы сульфат-ионов (20 атомов) и 721 молекулу воды. Некристаллографических симметрий в асимметрической ячейке нет.

В таблице 1 содержится краткая информация, суммирующая все вышесказанное.

3.1 Основные показатели качества структуры

Основными показателями качества модели являются: R-free, или критерий перекрывания Ван-дер-Ваальсовых радиусов атомов (Clashscore), доля маргинальных остатков по карте Рамачандрана (Ramachandran outliers), доля маргинальных остатков по положению боковой цепи (Sidechain outliers) и доля маргинальных остатков по RSR (RSRZ outliers).

На рисунке 1 показаны основные показатели качества модели.

3.1.1 R-фактор и R-free

R-фактор характеризует то, насколько модель соответствует полученным экспериментальным данным. Таким образом, чем ниже значение R-фактора, тем лучше модель соотносится с полученной структурой.

Хороший R-фактор имеет порог сверху, равный 0.25. В нашем случае, значение этого показателя равно 0.199, что говорит о том, что модель хорошо описывает экспериментальные данные.

R-free – показатель, значение которого считается так же, как и значение R-фактора, только по меньшему количеству рефлексов (5 % от общего количества = 3393 рефлекса). Введение данного показателя помогает предотвратить подгонку модели под хороший R-фактор. Порогом

сверху для хорошего R-free считается 0.25. В нашем случае его значение равно 0.248 (рисунок 1), однако даже при таком неплохом R-free, видно, что модель переоптимизирована, так как отклонение R-free от R-фактора составляет 24.6 % при пороге в 10 %.

Таблица 1: **Краткая информация о структуре 3LFU.**

Параметр структуры	Описание параметра
Дата загрузки на сайт PDB	2010
Дата публикации структуры	2011
Авторы модели	Korolev S, Waksman G, Lohman TM
Разрешение структуры (в Å)	1.8
Длины направляющих векторов (в Å)	194.543, 58.739, 69.418
Углы между направляющими векторами (в °)	90.00, 90.00, 90.00
Кристаллографическая группа	P 21 21 2
Количество цепей в элементарной ячейке	4
Количество измеренных рефлексов	67971
Полнота набора структурных факторов (в %)	91
Диапазон разрешений (в Å)	1.80-29.54
Метод решения фазовой проблемы	Молекулярное замещение
Содержание асимметрической единицы	5039 атомов белка, 4 молекулы сульфат-ионов и 721 молекула воды
Наличие некристаллографических симметрий	-

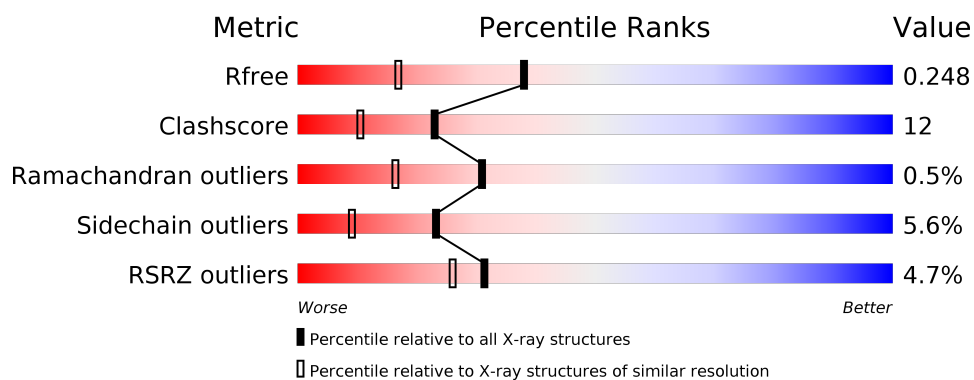


Рис. 1: Основные показатели качества структуры.

3.1.2 Clashscore

Clashscore характеризует число слишком близких контактов между парами атомов, которые не связаны между собой химической связью. При этом возникают стерические перекрытия Ван-дер-Ваальсовых радиусов (≥ 0.4 ангстрема). Значение clashscore считается как количество таких контактов на 1000 атомов. В нашем случае значение all-atom clashscore равно 12. Значение all-atom clashscore в MolProbity [5] – 12.5, что соответствует 60-ому перцентилю среди структур похожего разрешения, загруженных в PDB на момент 2004 года. Количество маргинальных остатков (слишком близко расположенных остатков) равно 123.

Таким образом, можно с уверенностью сказать, что наша структура имеет довольно большое количество слишком близких контактов между парами атомов. По этому критерию, можно заключить, что наша модель не очень хорошая.

3.1.3 Карта Рамачандрана

Карта Рамачандрана – способ визуализации торсионных углов (ϕ и ψ) аминокислотных остатков в белковой структуре. На карте можно выделить три энергетические области, которые соответствуют разрешенному, запрещенному и предпочитаемому состояниям.

На рисунке 2 представлена карта Рамачандрана для всех аминокислот, глицина и транс-пролина. Розовым кружком отмечены маргинальные остатки (остатки, расположенные в запрещенной области; найдено 3 остатка), они занимают 0.48 % от общего количества остатков (порог хорошей структуры = 0.05 %). Кроме маргинальных остатков и остатков,

которые расположены в предпочитаемых областях, есть остатки, которые располагаются в разрешенной области; их доля – 1.93 % (12 остатков). Таким образом, исходя из результатов, представленных на карте Рамачандрана, можно заключить, что модель не очень хорошая.

3.1.4 Ротамеры

Ротамеры – боковые цепи, находящиеся в предпочитаемых для данного остатка конформациях, которые возможно описать значениями торсионных углов. Этот подход нахождения маргинальных остатков отличается от карты Рамачандрана тем, что в последней используются торсионные углы полипептидного остова.

Количество маргинальных остатков, согласно wwPDB, - 29 (5.6 % при пороге в 0.3 %). Такое значительное превышение порога свидетельствует о том, что подобранная модель плохо соотносится с “реальностью”.

3.1.5 RSRZ

RSR характеризует насколько модель группы остатков/отдельного остатка соответствует экспериментально полученной электронной плотности. Так как RSR можно считать для одного остатка, значение критерия используется для нахождения маргинальных остатков.

Значение RSR по всем остаткам равно, что говорит о потенциальном значительном несоответствии модели и экспериментально полученной электронной плотности.

Существует z-score, который является относительной оценкой RSR, то есть показывает, на сколько среднеквадратичных отклонений RSR группы остатков/одного остатка отличается от среднего значения RSR по иным моделям похожего разрешения.

В нашем случае было найдено 30 маргинальных остатков ($z\text{-score} > 2$), что соответствует 4.7 % от общего числа остатков в белке. Таким образом, можно заключить, что, основываясь на описанном критерии, наша модель плохая: большое количество остатков вписываются плохо в экспериментально полученную электронную плотность.

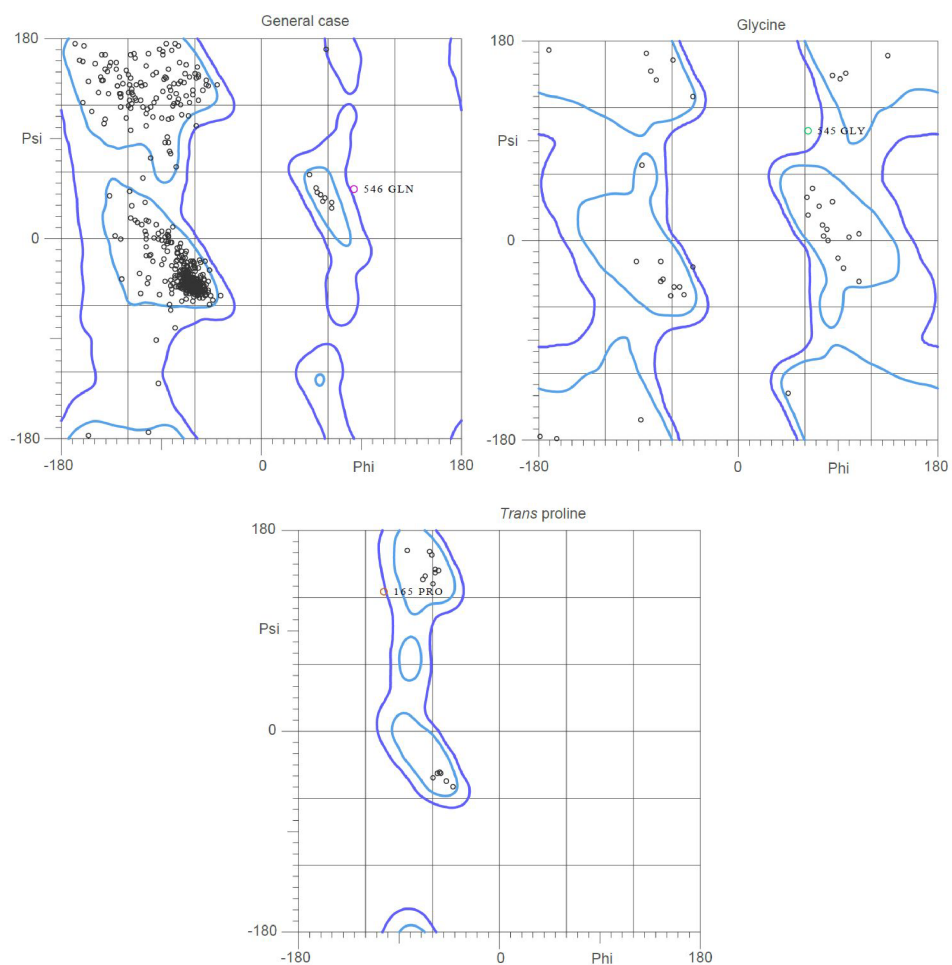


Рис. 2: Карта Рамачандрана для аминокислотных остатков структуры 3LFU.

Приведены изображения для трех случаев: все аминокислоты (general case), глицин (glycine), транс-пролин (trans proline). Три изображения (изолейцин и валин, пре-пролин, цис-пролин) не приведены, так как сервис MolProbity не нашел маргинальные остатки среди этих аминокислот. Три аминокислотных остатка Gln546, Gly545 и Pro165 находятся в запрещенных областях

3.2 Анализ маргинальных остатков

В таблице 2 приведен список маргинальных остатков, которые были найдены при помощи различных показателей качества структуры.

Таблица 2: **Потенциальные маргинальные остатки, выявленные различными показателями качества структуры.**

Остаток	Критерий маргинальности
Gly545, Gln546, Pro165	Карта Рамачандрана
Asn164, Gly590, Gln546	RSRZ
Phe407O – Val411HG23	Clashscore
Leu6, Arg42, Leu123	Ротамеры
Pro414	Окружение

3.2.1 Gln546

Согласно карте Рамачандрана (рисунок 2) остаток Gln546 – маргинальный остаток. При визуализации остатка выяснилось, что в его структуре отсутствует 5 атомов (CB, CG, CD, OE1 и OE2) (рисунок 3; соседний остаток Gly545 приведен для сравнения схожести структур остатков). В pdb файле есть поле remark, в котором выписаны атомы, которые по каким-то причинам не удалось внести в конечную структуру (рисунок 4). Видно, что там перечислены атомы рассматриваемого маргинального остатка.

Таким образом, можно сказать, что этот остаток является ошибочно вписанным, так как согласно структуре совершенно непонятно, почему на 546 позиции стоит глутамин, а не глицин.

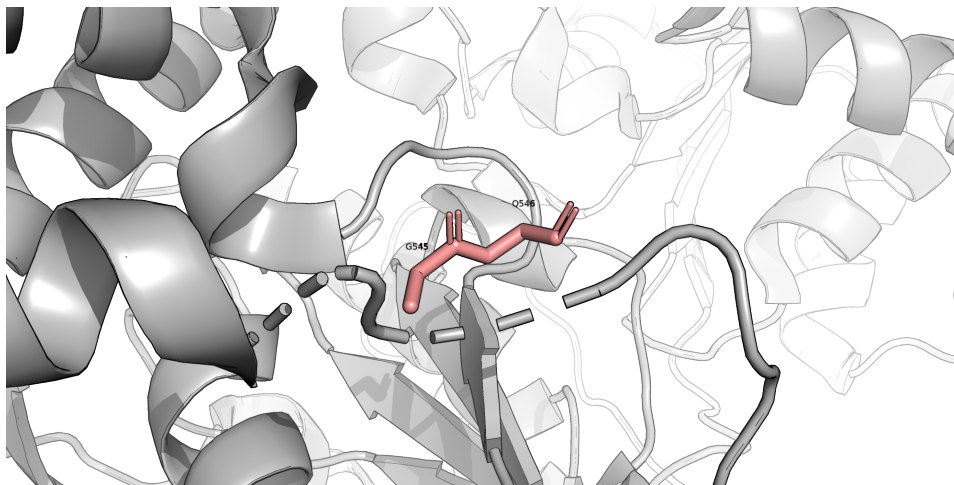


Рис. 3: Gln546 - потенциальный маргинальный остаток согласно карте Рамачандрана.

У остатка Gln546 отсутствует 5 атомов - CB, CG, CD, OE1 и OE2. По количеству атомов рассматриваемый "глутамин" сопоставим с глицином. Для сравнения приведен соседний остаток - Gly545.

```

REMARK 470 MISSING ATOM
REMARK 470 THE FOLLOWING RESIDUES HAVE MISSING ATOMS (M=MODEL NUMBER;
REMARK 470 RES=RESIDUE NAME; C=CHAIN IDENTIFIER; SSEQ=SEQUENCE NUMBER;
REMARK 470 I=INSERTION CODE):
REMARK 470  M RES CSSEQI  ATOMS
REMARK 470   VAL A   3      CG1  CG2
REMARK 470   HIS A 157      CG  ND1  CD2  CE1  NE2
REMARK 470   HIS A 158      CG  ND1  CD2  CE1  NE2
REMARK 470   ILE A 159      CB  CG1  CG2  CD1
REMARK 470   ASN A 164      CB  CG   OD1  ND2
REMARK 470   PRO A 165      CG   CD
REMARK 470   GLU A 167      CG  CD   OE1  OE2
REMARK 470   GLN A 168      CG  CD   OE1  NE2
REMARK 470   SER A 520      OG
REMARK 470   GLU A 526      CG  CD   OE1  OE2
REMARK 470   GLU A 544      CB  CG  CD   OE1  OE2
REMARK 470   GLN A 546      CB  CG  CD   OE1  NE2
REMARK 470   ALA A 547      CB
REMARK 470   ASP A 548      CG   OD1  OD2

```

Рис. 4: Аннотация "отсутствующие атомы в остатках" в поле remark в pdb файле для структуры 3LFU.

3.2.2 Asn164 и Pro165

Выдача RSRZ (таблица 2) указывает на то, что Asn164 является маргинальным остатком. При визуализации (рисунок 5) я решил показать также соседний остаток – Pro165 – так как он является маргинальным согласно карте Рамачандрана (рисунок 2). Выяснилось, что в обоих остатках (аспарагин и пролин) отсутствуют несколько атомов. В pdb файле содержится информация об этих атомах (рисунок 4).

Из всего вышесказанного можно заключить, что эти остатки являются ошибочно вписанными.

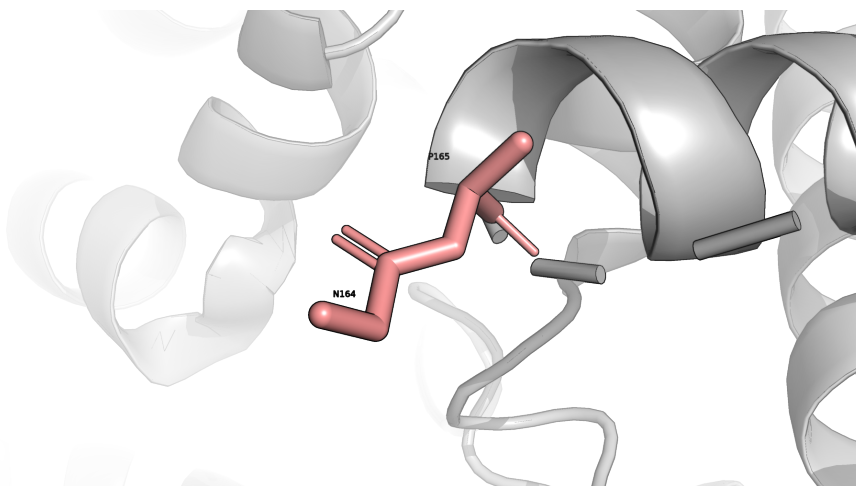


Рис. 5: Потенциальные маргинальные остатки согласно критерию RZRZ - Asn164 и Pro165.

У обоих остатков отсутствуют несколько атомов (атомы перечислены на рисунке 4).

3.2.3 Leu6

Маргинальным остатком согласно критерию “Ротамеры” является Leu6 (таблица 2). При измерении торсионных углов χ_1 и χ_2 [1] (рисунок 6) выяснилось, что они имеют значения, нехарактерные для большинства лейцинов (рисунок 7). При этом, при уровне подрезки 2 видно, что остаток вписан в электронную плотность довольно хорошо, что может говорить о том, что Leu6 является маргинальным остатком.

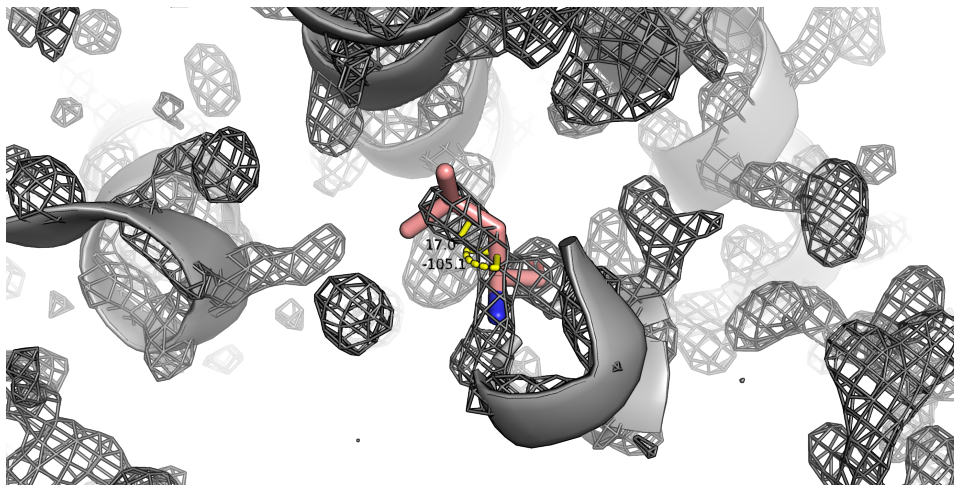


Рис. 6: Leu6 - потенциальный маргинальный остаток, выявленный критерием "Ротамеры".

Желтыми пунктирными линиями показаны торсионные углы для χ_1 и χ_2 . Значения измеренных углов не совпадают с наиболее часто встречающимися в структурах (рисунок 7). Серой сеткой проиллюстрирована электронная плотность на уровне подрезки 2.

Розовым цветом отмечены атомы углерода, синим - атом азота и красным - атом кислорода.

3.2.4 Val411 и Phe407

Согласно выдаче clashscore два остатка – Val411 и Phe407 расположены слишком близко друг к другу (таблица 2). При визуализации (рисунок 8) в глаза сразу бросается близкое расположение атома N Val411 и атома O Phe407. В принципе, это ожидаемо, что атом N может заряжаться положительно, а атом O – отрицательно, при этом атомы сближаются. Clashscore говорит о пересечении Ван-дер-Ваальсовых радиусов атома HG23 Val411 и атома O Phe407. Так как остатки могли сблизиться благодаря образованию зарядов на атомах N Val411 и O Phe407, неудивительно, что Ван-дер-Ваальсовы радиусы HG23 Val411 и O Phe407 могли пересечься слишком сильно. Возможно, авторы не разнесли эти остатки дальше друг от друга, так как оба остатка находятся в альфа-спирали, поэтому разнесение этих остатков привело бы к дестабилизации вторичной структуры (хотя это объяснение верно лишь только если быть уверенным, что в обсуждаемые аминокислоты являются не ошибочно вписанными в структуру).

Из вышесказанного, можно предположить, что Val411 и Phe407 являются ошибочно вписанными остатками.

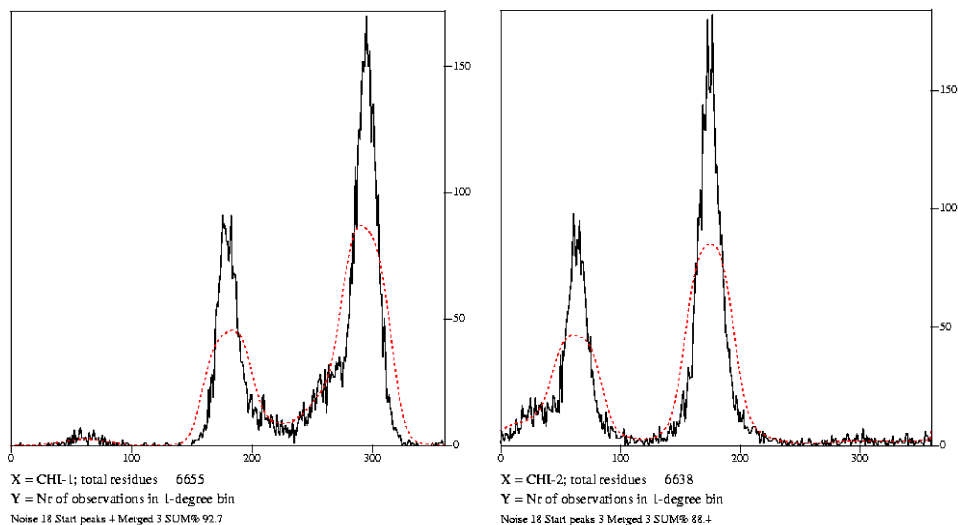


Рис. 7: Наиболее характерные значения торсионных углов χ_1 и χ_2 для лейцина.

Картинки построены при помощи сервиса для подсчета торсионных углов [4].

3.2.5 Pro414

Было найдено 2 цис-пролина в структуре, что, безусловно, очень много, учитывая тот факт, что пролинов в нашей структуре всего 18 штук. После визуализации стало понятно, что Pro414 по-видимому и правда находится в цис-конформации (рисунок 9). Соседние остатки (413 и 415) приведены для того, чтобы удостовериться в том, что пролин находится в цис-конформации. СА атом Pro414 и СА атом Thr413 находятся "по одну сторону то есть Pro414 находится в цис-конформации. При этом, при уровне подрезки 2, Pro414 хорошо вписывается в электронную плотность.

Таким образом, по-видимому Pro414 нельзя считать маргинальным остатком.

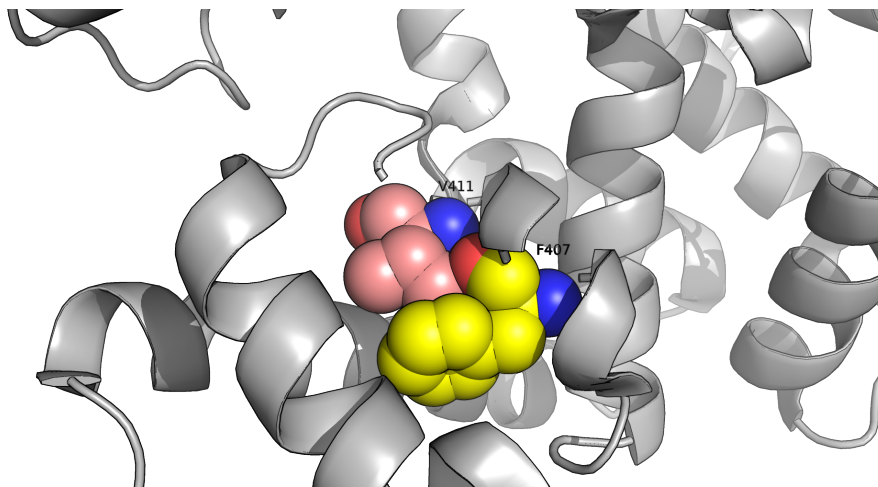


Рис. 8: Расположение Val411 и Phe407; атомы остатков показаны в виде сфер.

Два остатка - Val411 и Phe407 - расположены слишком близко друг к другу (согласно критерию clashscore Ван-дер-Ваальсовы радиусы атомов HG23 Val411 и O Phe407 пересекаются на 1.17 Å).

Розовым цветом отмечены атомы углерода остатка Val411, желтым - атомы углерода остатка Phe407, красным - атом кислорода остатка Phe407.

3.3 Улучшение параметров модели при помощи сервиса PDB-REDO

Сервис PDB-REDO [3] улучшил параметры модели (рисунок 10). Видно, что обновленная модель (PDB-REDO на рисунке) по сравнению с моделями структур с похожим разрешением показывает неплохие результаты, чего нельзя о модели до оптимизации (PDB на рисунке).

Кроме того, в таблице 3 приведены некоторые переменные, произошедшие со структурой после оптимизации.

Исходя из всего вышесказанного, можно с уверенностью сказать, что сервис PDB-REDO явно улучшил параметры модели, а также саму структуру белка UvrD с PDB ID 3LFU.

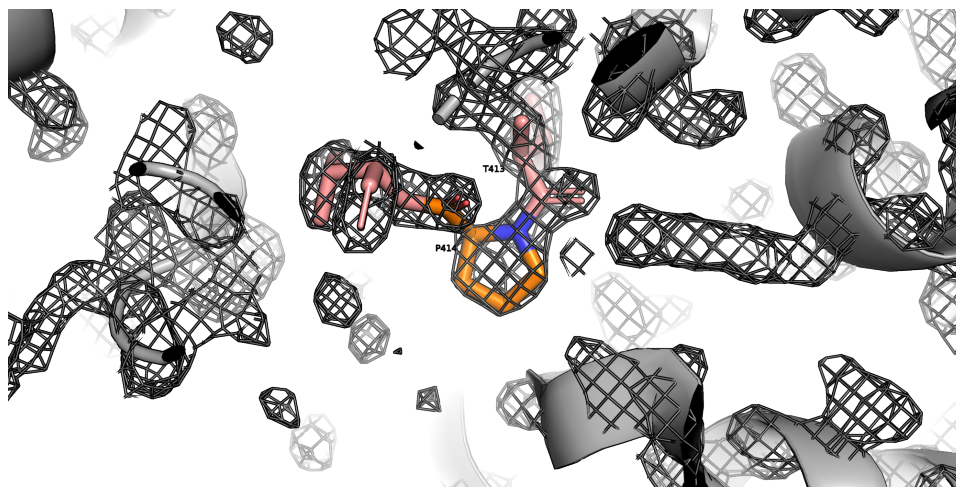


Рис. 9: Pro414 в цис-конформации в структуре 3LFU. Оранжевым выделены атомы углерода остатка Pro414, красным выделен атом кислорода остатка Pro414, синим выделен атом азота остатка Pro414 розовым - атомы остатков Thr413 и Thr415.

Таблица 3: Параметры модели, подвергшиеся изменениям PDB-REDO.

Описание изменения	Количество изменений
Изменены ротамеры	27
Повернуты боковые цепи	13
Удалены молекулы воды	114
Повернуты пептиды	1
Остатки точнее вписаны в электронную плотность	2

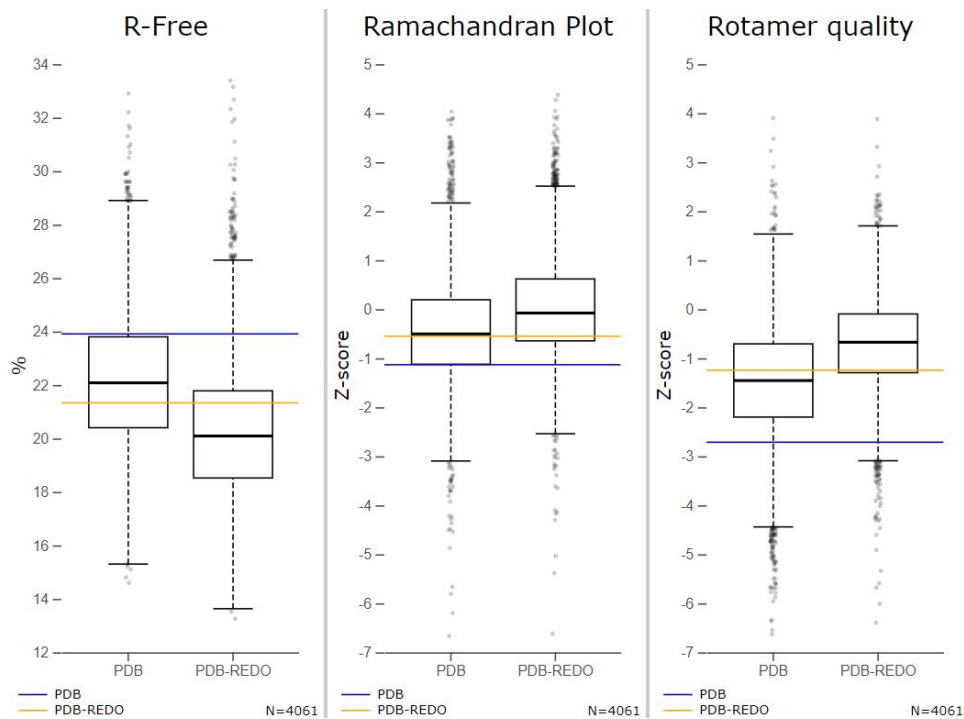


Рис. 10: Сравнение параметров модели структуры 3LFU до и после улучшения (RDB-REDO).

Для сравнения были взяты 4061 структура с похожим разрешением; по этой группе структур были построены три box-plot для трех различных параметров - R-free, Ramachandran plot и Rotamer quality.

Сервис PDB-REDO выдает улучшенное значение параметров. Если во всех трех случаях до улучшения значения параметров модели не попадали в "ящик" (25-75 перцентили, или интерквартильный "размах") - синяя линия, то после - попадали - желтая линия. При этом, box-plot строится отдельно для структур до и после улучшения.

4 Выводы

Практически все проверяемые остатки оказались маргинальными, часто из-за того, что в структуре отсутствовали атомы или остатки (рисунок 4). Некоторые значения торсионных углов отличались от наиболее часто встречаемых для исследуемых остатков (рисунок 7). Кроме того, наша структура оказалась переоптимизированной. При этом, крайне важный показатель качества структуры – доля зарегистрированных рефлексов – оказался довольно низким (91%). К тому же, сервис PDB-REDO выдал структуру, которая была улучшена по всем параметрам (таблица 3), что может говорить, как о хорошей работе алгоритма, так и о не очень хорошей модели.

В любом случае, можно с уверенностью сказать, что предлагаемая модель структуры, несмотря на хорошее разрешение в 1.8 Å, достаточно плохая, по сравнению со структурами похожего разрешения.

Список литературы

- [1] D. Caballero, W. W. Smith, C. S. O’Hern, and L. Regan. Equilibrium transitions between side-chain conformations in leucine and isoleucine. 83(8):1488–1499.
- [2] H. Jia, S. Korolev, A. Niedziela-Majka, N. K. Maluf, G. H. Gauss, S. Myong, T. Ha, G. Waksman, and T. M. Lohman. Rotations of the 2b sub-domain of *e. coli* UvrD helicase/translocase coupled to nucleotide and DNA binding. 411(3):633–648.
- [3] R. P. Joosten, J. Salzemann, V. Bloch, H. Stockinger, A.-C. Berglund, C. Blanchet, E. Bongcam-Rudloff, C. Combet, A. L. Da Costa, G. Deleage, M. Diarena, R. Fabbretti, G. Fettahi, V. Flegel, A. Gisel, V. Kasam, T. Kervinen, E. Korpelainen, K. Mattila, M. Pagni, M. Reichstadt, V. Breton, I. J. Tickle, and G. Vriend. PDB_redo: automated re-refinement of x-ray structure models in the PDB. 42:376–384.
- [4] G. J. Kleywegt and T. A. Jones. Databases in protein crystallography. 54:1119–1131.
- [5] R. Rodriguez, G. Chinae, N. Lopez, T. Pons, and G. Vriend. Homology modeling, model and software evaluation: three related resources. 14(6):523–528.