

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

Отчёт по качеству расшифровки структуры 2-дегидро-3-деоксифосфооктонатаьдолазы из  
*Burkholderia pseudomallei* (PDB ID: 3UND) методом  
рентгеноструктурного анализа

Самборская Маргарита,  
4 курс ФББ МГУ

Москва, 2017 г.

## Аннотация

Целью отчета было рассмотрение информации о ферменте и его функции, качестве расшифровки его структуры, проведение анализа маргинальных аминокислотных остатков и сравнение модели PDB с моделью из PDB\_redo.

## Введение

В данном отчете оценивается качество PCA расшифровки структуры 2-дегидро-3-деоксифосфооктонат альдолазы из *Burkholderia pseudomallei* (PDB ID: 3UND). Структура этого белка была расшифрована группой ученых Vaugh, L., Gallagher, L.A., Patrapuvich, R в статье Combining functional and structural genomics to sample the essential *Burkholderia* struotome в Plos One в 2013 году. Этот белок имеет ключевую роль в метаболизме представителей рода *Burkholderia*, болезнетворных грам-отрицательных бактерий, вызывающих инфекции в раковых пациентах и пациентах страдающих фиброзом.

Фермент представляет собой симметричный тетрамер, 285 аминокислотных остатков в субъединице.

## Общая информация о модели

- метод решения фазовой проблемы - X-RAY DIFFRACTION
- число измеренных рефлексов (см. в файле структурных факторов) - 71084
- разрешение 2.1 Å, полнота набора структурных факторов – 97.3, диапазон разрешений структурных факторов – 2.1 Å – 50 Å
- параметры кристаллографической ячейки Длина (Å): a: 46.84 Å b: 146.24 Å c: 178.94 Å  
Угол (°):  $\alpha$ : 90°  $\beta$ : 90°  $\gamma$ : 90°
- наличие некристаллографических симметрий в асимметрической ячейке - не обнаружены

## Значение индикаторов качества

- R-фактор – 0.167
- R\_free - 0.192
- Карта Рамачандрана представлена в дальнейшем

Также краткая аннотация валидации структуры представлена на схеме 1.

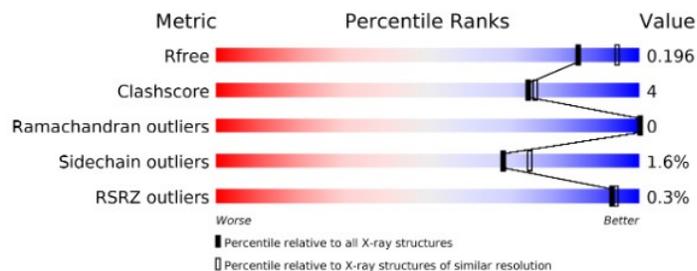


Схема 1. Краткая аннотация вализации 3UND

### Нахождение маргинальных остатков при помощи EDS

Попробуем найти маргинальные остатки, используя EDS [1].

EDS предоставляет возможность рассмотреть большой спектр критериев качества модели.

1) Real Space R-factor (RSR) - пространственный R-фактор для всех остатков и молекул растворителя

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 = **0.102**, то есть 10,2 %, что меньше 20% и свидетельствует о приемлемом качестве модели.

Рассмотрим гистограмму значений RSR для всех остатков в четырех цепях структуры.



Рис.1. Гистограмма значений RSR для всех остатков всех четырех цепей в структуре 3UND. По оси Y отложены значения RSR, по оси X - номера остатков. Значения, которые превышают 0.2, можно считать маргинальными.

Кроме критерия RSR маргинальные остатки также могут быть найдены с помощью Z score [2].

$$Z\text{-score} = (RSR - \langle RSR \rangle) / \sigma$$

Z-score вычисляется относительно всех остатков такого же типа из структур с таким же разрешением. Значения  $Z > 2$  свидетельствуют о том, что остаток скорее всего является маргинальным по сравнению с другими структурами с тем же разрешением.

То есть RSR это характеристика абсолютная, а Z score – относительная (относительно других структур с тем же разрешением).

Далее представлены параметры Z score – средний Z score и sigma для четырех цепей.

Chain	Nres	<Z>	Sigma	%<-1	%<0	%>1	%>2
A	283	-0.54	0.73	13.07	88.69	1.41	0.00
B	284	-0.21	0.59	2.46	71.13	2.46	0.35
C	282	-0.64	0.81	16.67	93.62	1.42	0.71
D	284	-0.31	0.56	2.11	77.82	1.06	0.00
<b>All</b>	<b>1133</b>	<b>-0.42</b>	<b>0.68</b>	<b>8.56</b>	<b>82.79</b>	<b>1.59</b>	<b>0.26</b>

Рис. 2 Параметры Z score для всех остатков четырех цепей в структуре 3UND.

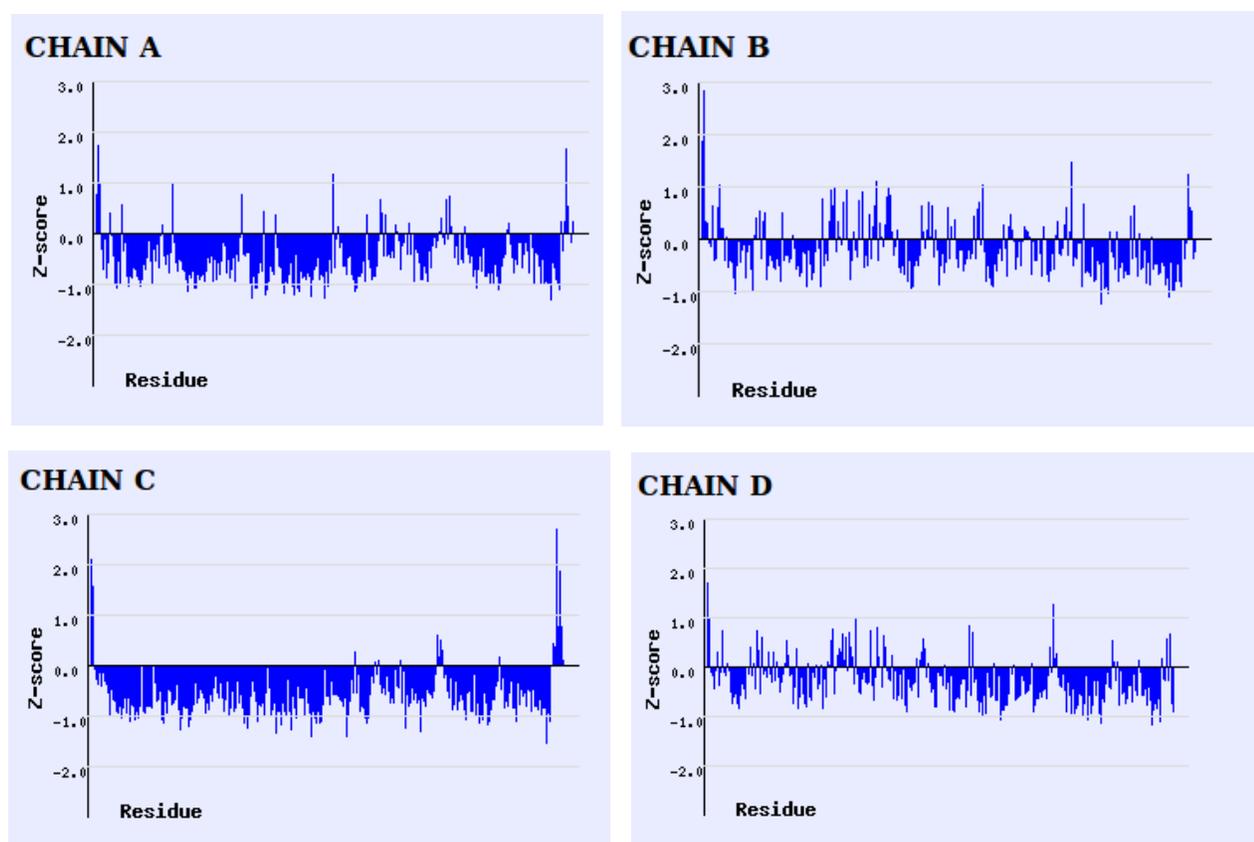


Рис.3 Значения Z-score. По оси Y отложены значения Z-score, по оси X отложены номера остатков. Остатки, значение для которых превышает 2.0 можно считать маргинальными.

Специально для удобства пользователей в EDS существует функция, позволяющая лучше отфильтровать маргинальные остатки – significant regions [3].

Она выделяет остатки с  $RSR > (<RSR> + 2.00 * \sigma)$

В данном случае мне показалось более целесообразным использовать significant regions как критерий отбора маргинальных участков. Было выделено три участка – Gly с цепи B и Ser и Ala с цепи C. Данные остатки также являются маргиналами согласно критериям Z score и RSR.

Также возможно нахождение маргинальных остатков посредством построения карты Рамачандрана сервисом EDS [4], MolProbity [5] или PDBsum [6].

На рисунке 6 представлена карта Рамачандрана, полученная с помощью MolProbity. Маргинальных остатков не было обнаружено. Обнаружено восемь остатков, не находящихся в запрещенной области, но и не находящихся в предпочтительной области.

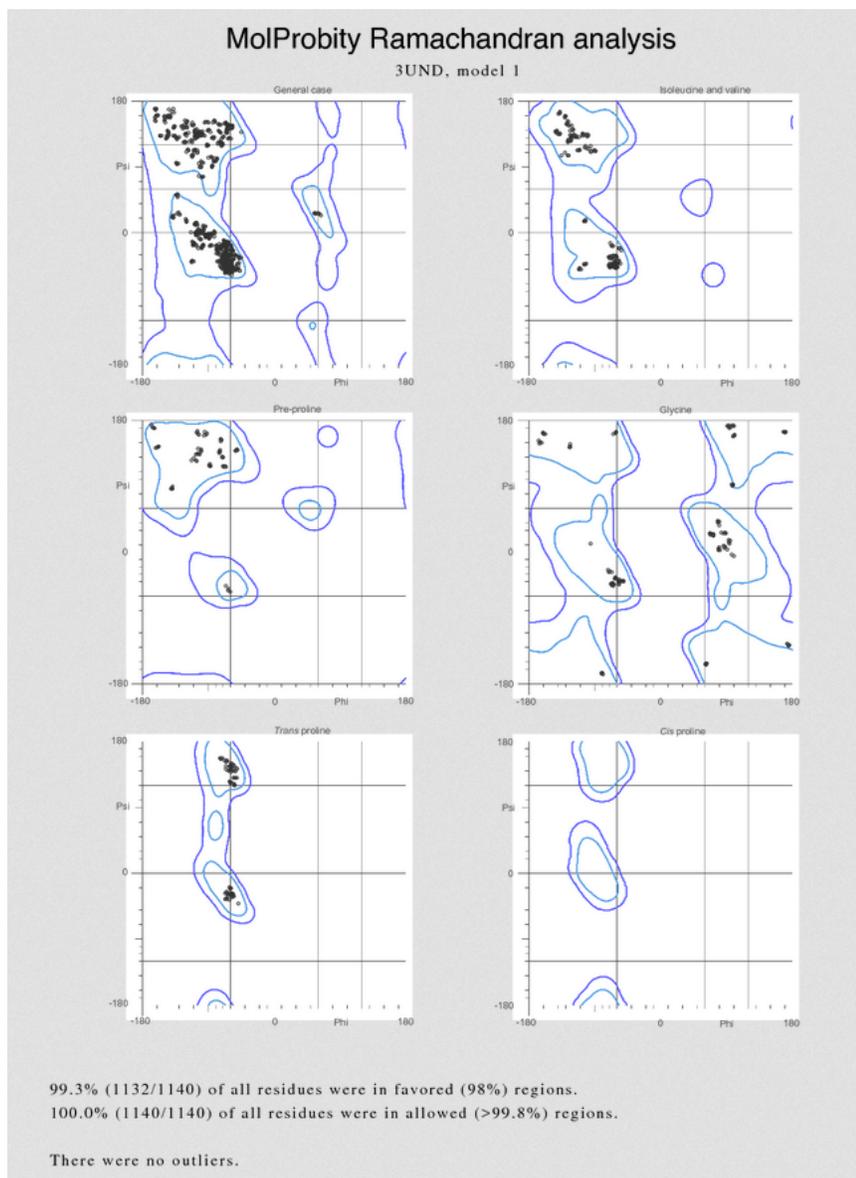
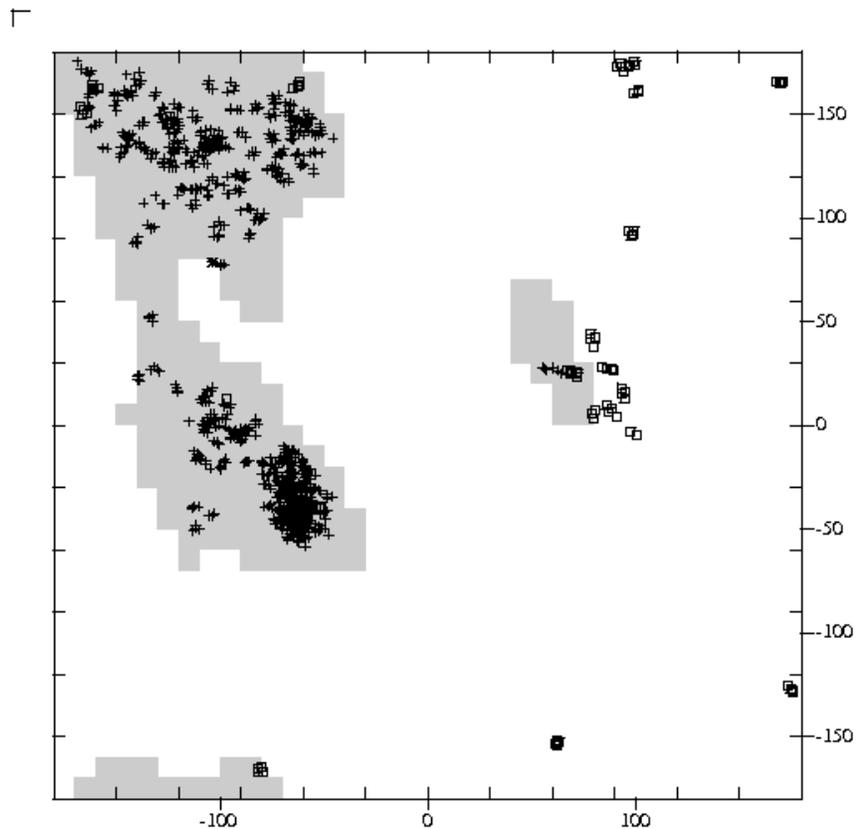


Рис.4 Карта Рамачандрана, полученная с помощью MolProbity.

На рисунке 7 представлена карта Рамачандрана, полученная с помощью EDS. Был найден один маргинальный остаток Arg C 160.

Данный остаток был занесен в список рассматриваемых маргиналов.



X = PHI mapped to [-180,180]  
 Y = PSI mapped to [-180,180]  
 PDB file : /pub/db/eds/sfd/3und/pdb3und.ent  
 Glycines (open squares): 106 ; Start/end residues : 8  
 D-amino acids : 0 ; Residues with missing atoms : 0  
 Residues in Ramachandran plot checked : 1019 out of 1133  
 In core regions (plus signs): 1018 ; Outliers (asterisks): 1  
 Percentage outliers: 0.1  
 An average  $\leq 2.0 \text{ \AA}$  model has ~0-5% outliers  
 See: Kleywegt, G.J. and Jones, T.A. (1996). Structure 4, 1395-1400.

Рис.5 Карта Рамачандрана, полученная с помощью EDS.

На рисунке 8 представлена карта Рамачандрана, полученная с помощью PDBsum (PROCHECK analyses). Маргинальных остатков не было найдено. Были найдены 4 остатка, не находящиеся в предпочтительной области (most favoured region), а находящиеся в разрешенной области (additional allowed region).

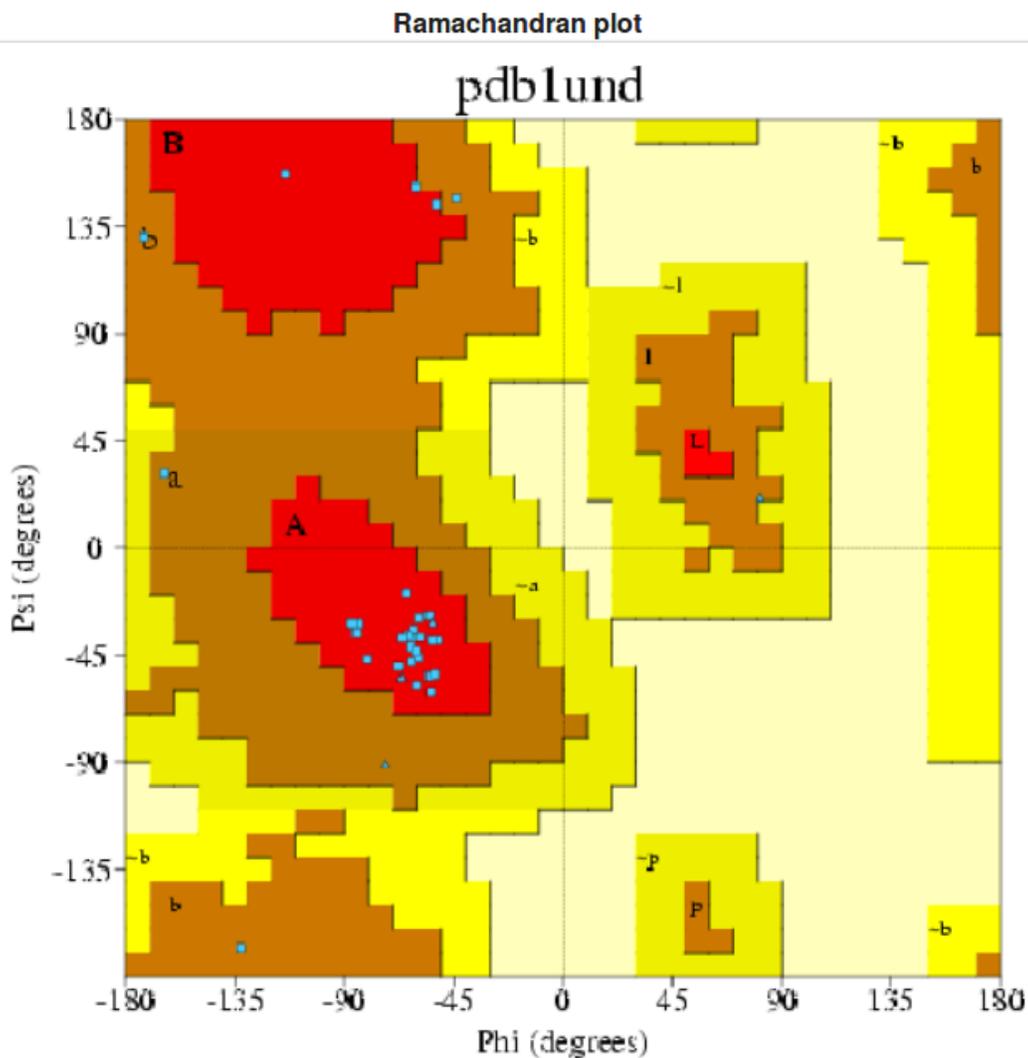


Рис.6 Карта Рамачандрана, полученная с помощью PDBsum (PROCHECK analyses)

С помощью Z-score и RSR, а также карт Рамачандрана, полученных различными методами, были найдены различные маргинальные остатки.

С помощью significant regions было найдено три маргинальных остатка, один был найден с помощью карты Рамачандрана, полученной с помощью EDS.

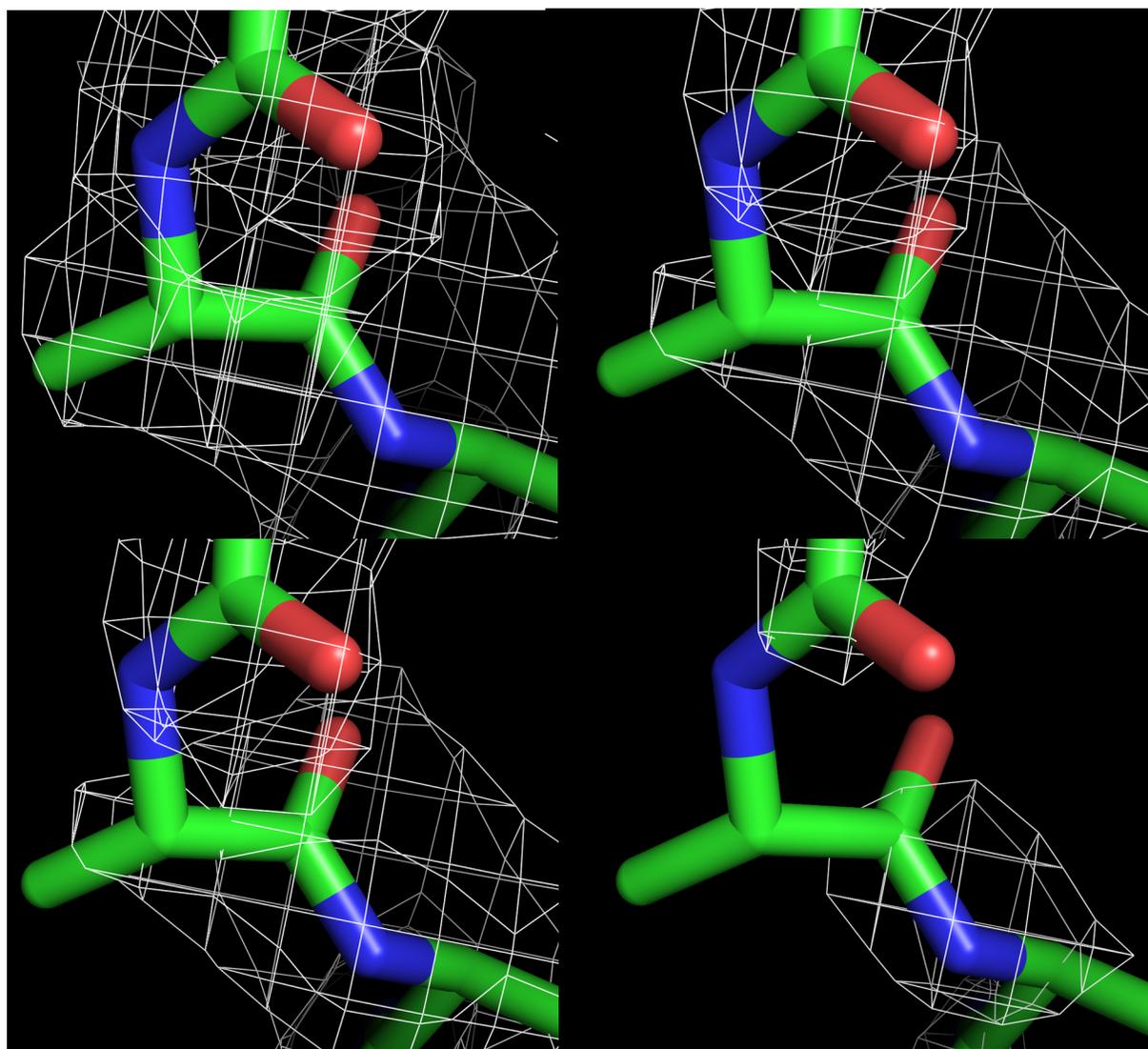
Стоит отметить, что в данном случае (на данной молекуле) с помощью критериев RSR и Z score с использованием порогов 0.2 и 2.0 возможно выделить очень большое количество маргинальных остатков, в то время как similarity score и карты Рамачандрана находят меньшее количество маргинальных остатков.

### Список маргинальных остатков

Номер/остаток/цепь	RSR	Phi	Psi	Метод выбора
277/Ala/C	0.253			Significant region
0/Ser/C	0.238			Significant region
160/Arg/C		103.6	78.6	
279/Glu/C	0.288			RSR
139/Gln/A	0.218			RSR

### **Анализ выбранных маргинальных остатков**

При анализе маргинальных остатков была рассмотрена функция электронной плотности вокруг них. Было использовано несколько уровней подрезки 0.5, 1, 1.5, 2 для каждого из выбранных остатков.



*Рис. 7 Электронная плотность вокруг Ala 277 на уровнях подрезки 0.5, 1, 1.5, 2 (слева направо, сверху вниз). Этот остаток, идентифицированный методом significant region, можно считать маргинальным, так как трудно идентифицировать не только расположение данного остатка в белке, но и тип остатка.*

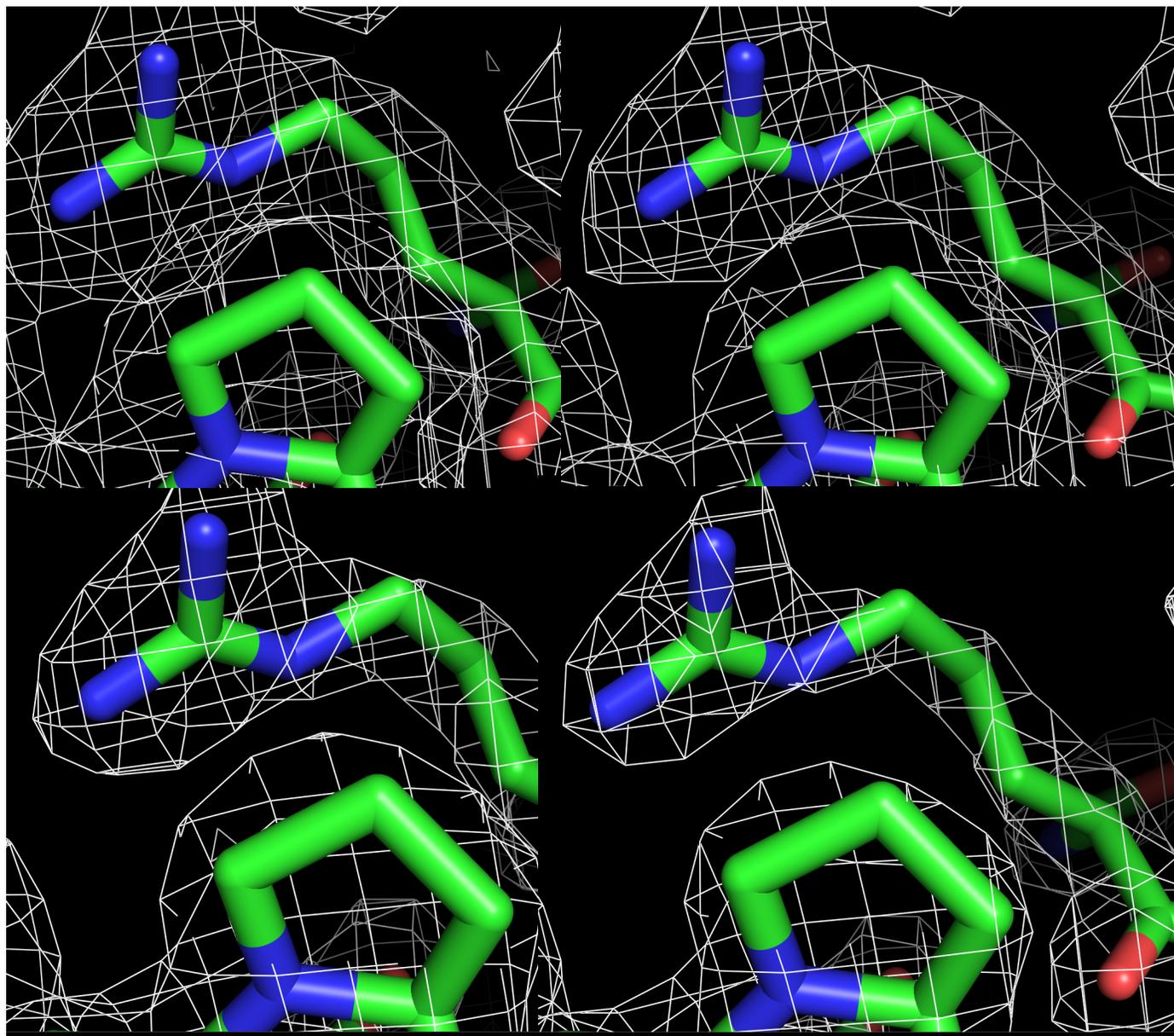
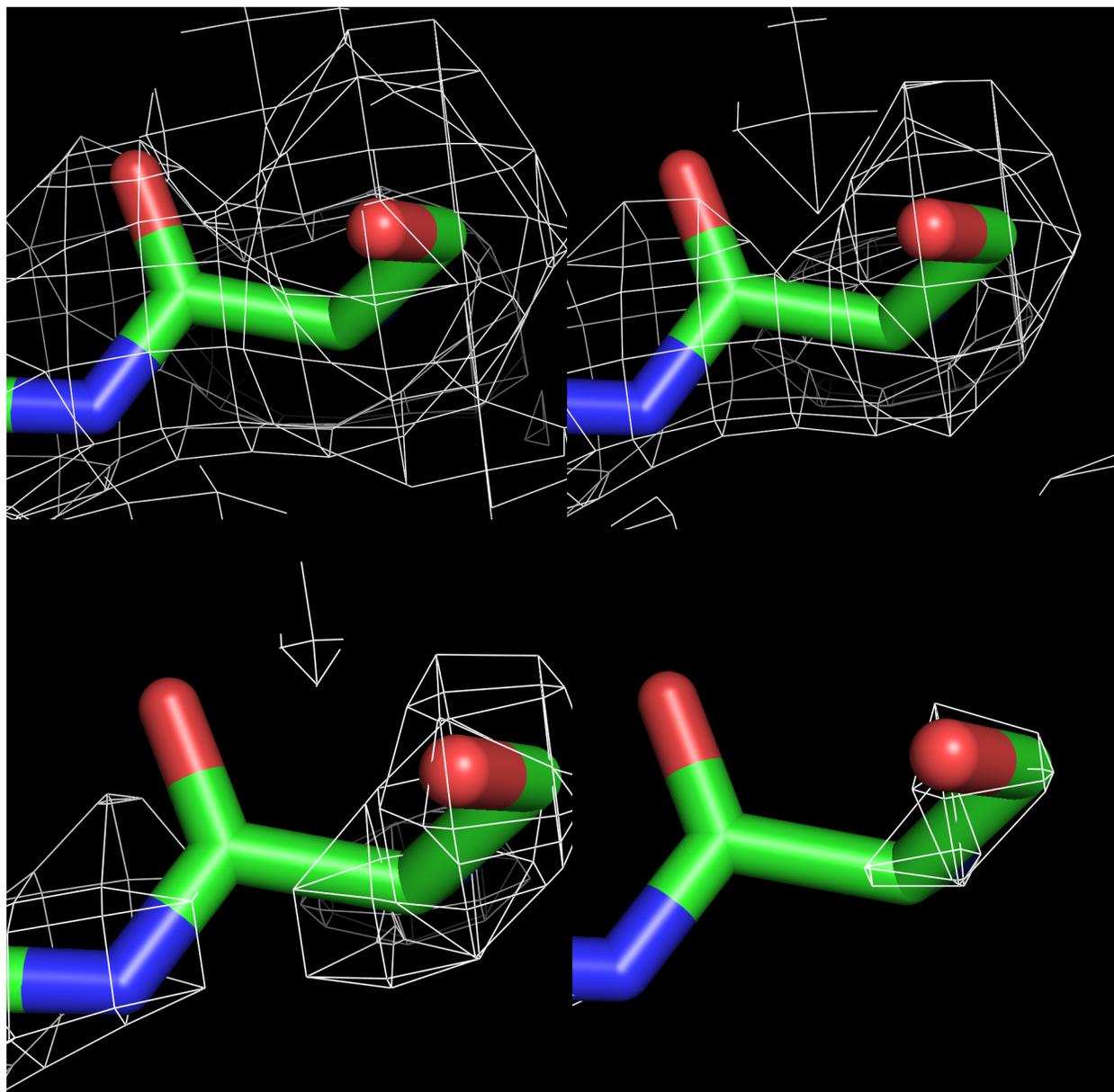
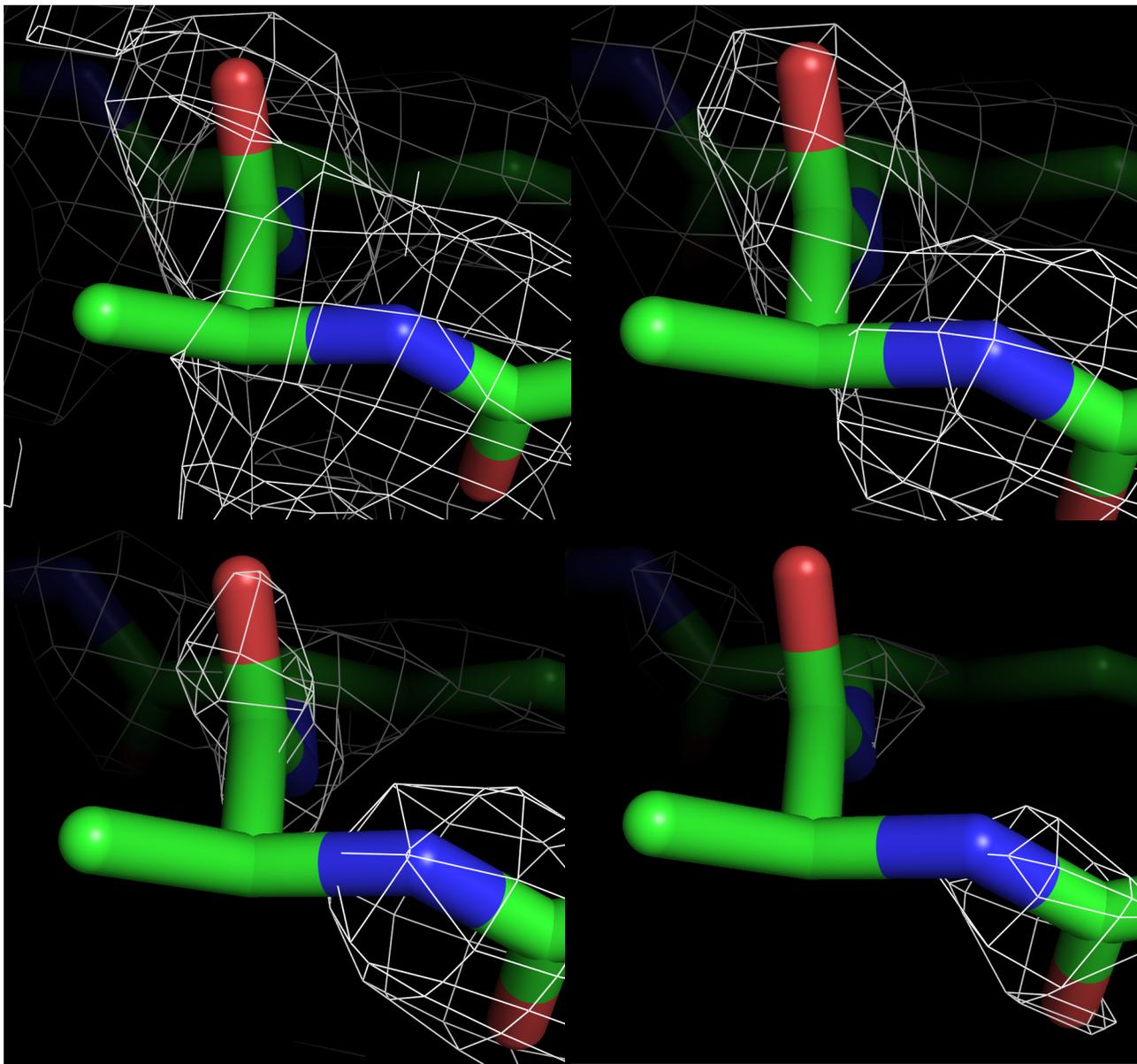


Рис. 8 Электронная плотность вокруг Ser 0 на уровнях подрезки 0.5, 1, 1.5, 2 (слева направо, сверху вниз). Этот остаток, идентифицированный методом *significant region*, вероятно маргинальным не считается, так как его расположение и тип можно идентифицировать на различных уровнях подрезки.



*Рис. 9 Электронная плотность вокруг Arg 160 на уровнях подрезки 0.5, 1, 1.5, 2 (слева направо, сверху вниз). Этот остаток, идентифицированный методом карты рамачандрана, можно считать маргинальным, так как трудно идентифицировать не только расположение данного остатка в белке, но и тип остатка.*



*Рис. 10 Электронная плотность вокруг Glu 279 на уровнях подрезки 0.5, 1, 1.5, 2 (слева направо, сверху вниз). Этот остаток, идентифицированный методом RSR, можно считать маргинальным, так как трудно идентифицировать не только расположение данного остатка в белке, но и тип остатка.*

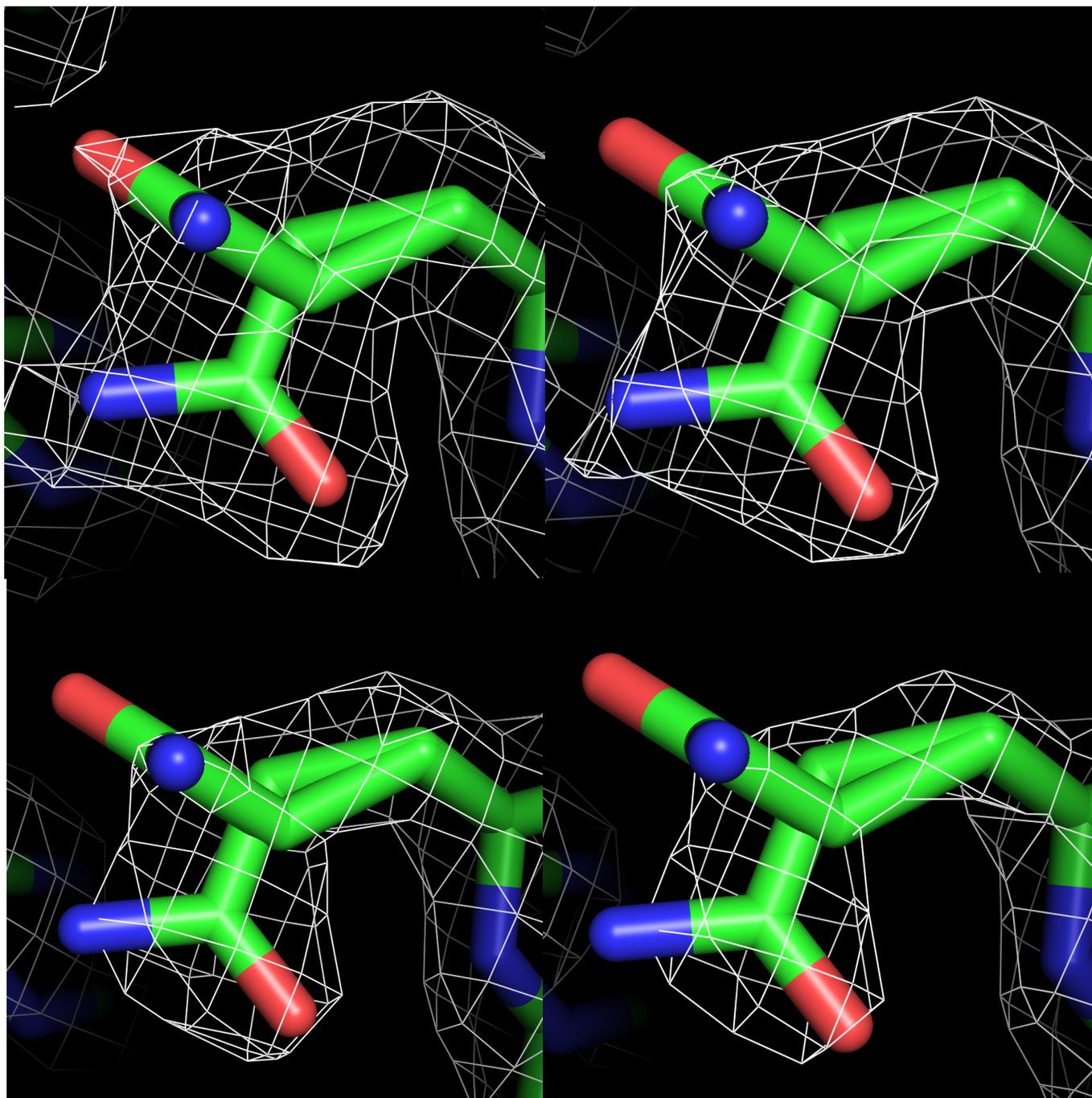


Рис. 11 Электронная плотность вокруг 139 Gln на уровнях подрезки 0.5, 1, 1.5, 2 (слева направо, сверху вниз). Этот остаток, идентифицированный методом RSR, вероятно маргинальным не считается, так как его расположение и тип можно идентифицировать на различных уровнях подрезки.

Наибольший процент точных предсказаний показал метод карты Рамачандрана, однако, трудно сделать какое-либо заключение на столь малой выборке.

### Сравнение модели из PDB с моделью из PDB\_redo

PDB-redo это база данных для белков, качество pdb структур которых было “улучшено” различными методами. Она позволяет скачивать оптимизированные структуры белков (Re-refined and rebuilt или просто Re-refined), существующих в базе данных PDB и предоставляет большое количество валидационных данных.

С помощью сервиса PDB\_redo version 7.00 было произведено сравнение модели 1UND из PDB с оптимизированной моделью [7]. Полученные валидационные метрики представлены в таблице 2.

Таблица 2. Значения валидационных параметров для оценки качества до и после. Положительный результат отмечен зеленым, отрицательный – красным.

Validation metrics from PDB-REDO		
	PDB	PDB-REDO
<b>Crystallographic refinement</b>		
<i>R</i>	0.1637	0.1615
<i>R-free</i>	0.1908	0.1832
<i>Bond length RMS Z-score</i>	0.718	0.577
<i>Bond angle RMS Z-score</i>	0.760	0.779
<b>Model quality (raw scores   percentiles)</b>		
<i>Ramachandran plot appearance</i>	49	53
<i>Rotamer normality</i>	76	74
<i>Coarse packing</i>	N/A	N/A
<i>Fine packing</i>	8	12
<i>Bump severity</i>	68	90
<i>Hydrogen bond satisfaction</i>	49	46
<b>Model quality (raw scores   percentiles)</b>		
<i>Ramachandran plot appearance</i>	-0.836	-0.699
<i>Rotamer normality</i>	-0.521	-0.584
<i>Coarse packing</i>	-0.535	-0.636
<i>Fine packing</i>	-2.147	-1.968
<i>Bump severity</i>	0.015	0.005
<i>Hydrogen bond satisfaction</i>	0.899	0.896

В целом R и R free уменьшились, плотность упаковки стала выше. Также было проведено сравнение качества модели по сравнению с моделями, сходными по разрешению. Стоит отметить, что несмотря на то, что по всем трем показателям (R free, Ramachandran plot и Rotamer quality) после PDB\_REDO стала показывать лучшие значения, необходимо также обратить внимание на дельту для нашей молекулы и для общей выборки. По всем трем графикам четко видно, что дельта для нашей структуры сильно уступает разнице между “до” и “после” для общей выборки.

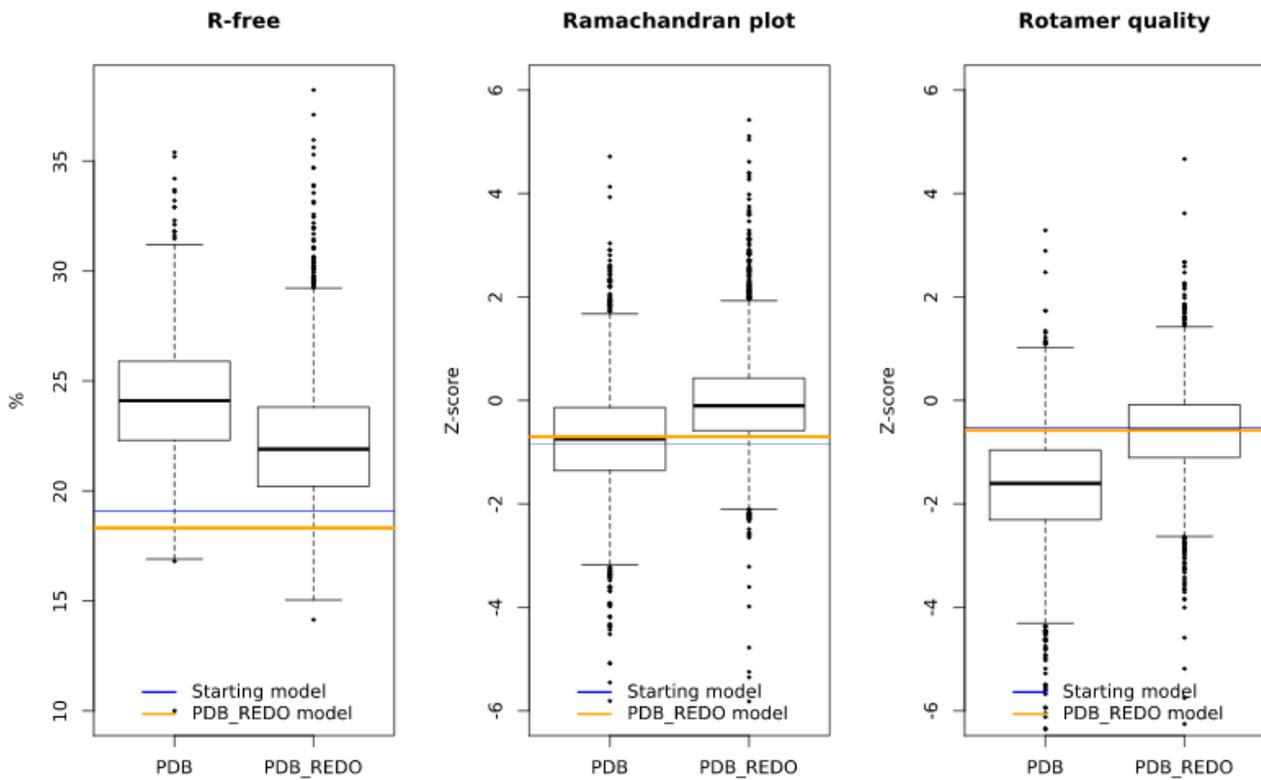


Рис.12 Качество модели по сравнению с моделями, сходными по разрешению

## *Заключение*

Рассмотренная модель обладает приемлимым качеством разрешения, наиболее оптимальным способом для нахождения маргинальных остатков скорее всего является комбинация представленных выше методов. Сервис PDB\_redo позволил улучшить показатели модели, что является несомненным плюсом.

Список литературы:

- 1) <http://eds.bmc.uu.se/cgi-bin/eds/uusfs?pdbCode=3und>
- 2) <http://eds.bmc.uu.se/cgi-bin/eds/zscore?3und>
- 3) <http://eds.bmc.uu.se/cgi-bin/eds/sign?3und&sco=2.0>
- 4) <http://eds.bmc.uu.se/cgi-bin/eds/rama?pdbCode=3und>
- 5) [https://cdn.rcsb.org/etl/ramachandran/img/un/3und/3und\\_rama.png](https://cdn.rcsb.org/etl/ramachandran/img/un/3und/3und_rama.png)
- 6) [https://www.ebi.ac.uk/thornton-srv/databases/cgi-bin/pdbsum/GetPage.pl?pdbcode=1und&template=procheck\\_summary.html](https://www.ebi.ac.uk/thornton-srv/databases/cgi-bin/pdbsum/GetPage.pl?pdbcode=1und&template=procheck_summary.html)
- 7) <https://pdb-redo.eu/db/3und>