

# Практикум 6. Валидация

## Задание 1

### Информация о белке

Мне была выдана структура с PDB ID **1DMN** (<https://www.rcsb.org/structure/1DMN>). Один из наиболее эффективных ферментов,  $\Delta 5$ -3-кетостероидизомераза (KSI), обеспечивает уникальную и высоко оптимизированную среду и геометрию активного центра. KSI от двух разных видов бактерий, *Pseudomonas putida* и *Comamonas testosteroni*, наиболее интенсивно исследовались с точки зрения взаимосвязи структура – функция. Хотя два KSI имеют только 34% идентичности последовательностей, трехмерные структуры двух KSI, определенные с высоким разрешением с помощью рентгеновской кристаллографии и KSI *C. testosteroni* с помощью ЯМР, показали, что их общие структуры удивительно похожи друг на друга.

Было обнаружено, что сетка водородных связей имеет решающее значение для KSI для поддержания правильной геометрии активного центра как для функции, так и для стабильности.



Рис 0. Структура 1DMN.

Трехмерные структуры обоих ферментов KSI, недавно определенные с помощью рентгеновской кристаллографии и ЯМР-спектроскопии, показывают, что этот небольшой белок свернут в шестицепочечный бета-лист и три альфа-спирали в каждом мономере. Большая часть активного сайта KSI является гидрофобной по природе, чтобы быть эффективной для связывания субстрата. Одной из наиболее заметных особенностей геометрии активного центра KSI является то, что полярные функциональные группы поддерживаются в высоко аполярном активном центре за счет образования сети водородных связей.

Оригинальная публикация PubMed: [10769113](https://pubmed.ncbi.nlm.nih.gov/10769113/)

### Оценка качества структуры

Качество структуры будем оценивать глядя на отчет о валидации.  
Разрешение нашей структуры 2.05Å.

**R-фактор** характеризует соответствие модели ее экспериментальным значениям (модулям структурных факторов)

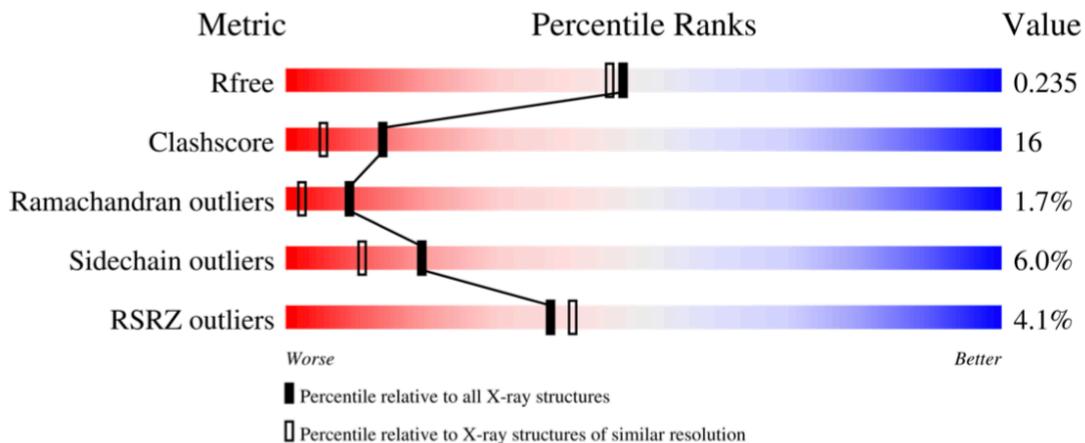


Рис 1. Оценка (в процентилях 0-100) глобальных метрик валидации структуры.

**R-factor** для нашей структуры равен 0.195, а **R-free** 0.244(из картинки 0.235). Значения факторов считаются хорошими, так как  $R < 0.25$  значит хорошее соответствие. Таким образом результаты достаточно приближены к экспериментальным данным. Разница между значениями маленькая (<10%), что говорит об отсутствии переобучения.

**Clashscore**- значение для количества атомов, для которых перекрываются Ван-дер-Вальсовы радиусы. Для нашей структуры значение равно 16, что мне кажется в целом не очень большим.

**Ramachandran outliers**- число маргиналов по картам Рамачандрана. Значение этого параметра 1.7%, что означает количество остатков, находящихся в не предпочитаемой зоне конформации.

**Sidechain outliers**- доля маргинальных остатков в боковых цепях. У нас 6%, что я считаю небольшим процентом.

При загрузке структуры в *rufo* видно, что многие остатки не покрыты электронной плотностью. Ниже пример таких остатков.

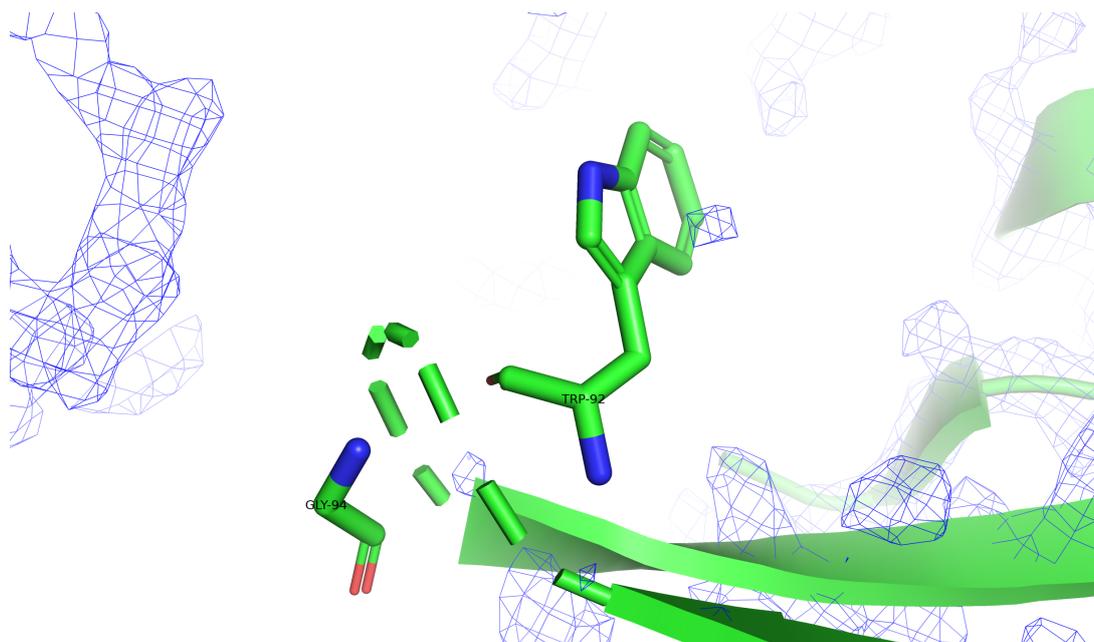


Рис 2. Примеры остатков, которые не покрывает ЭП(mesh 2, carve 1.5).

Также с помощью базы данных Uniprot я попыталась понять, значимы ли остатки, не покрытые ЭП. Из раздела Mutagenesis я посмотрела для всех остатков, покрыты ли они ЭП. Для остатков Val127 и Trp92 видим, что нет. Таким образом, эти остатки без ЭП важные для структуры белка при мутагенезе.

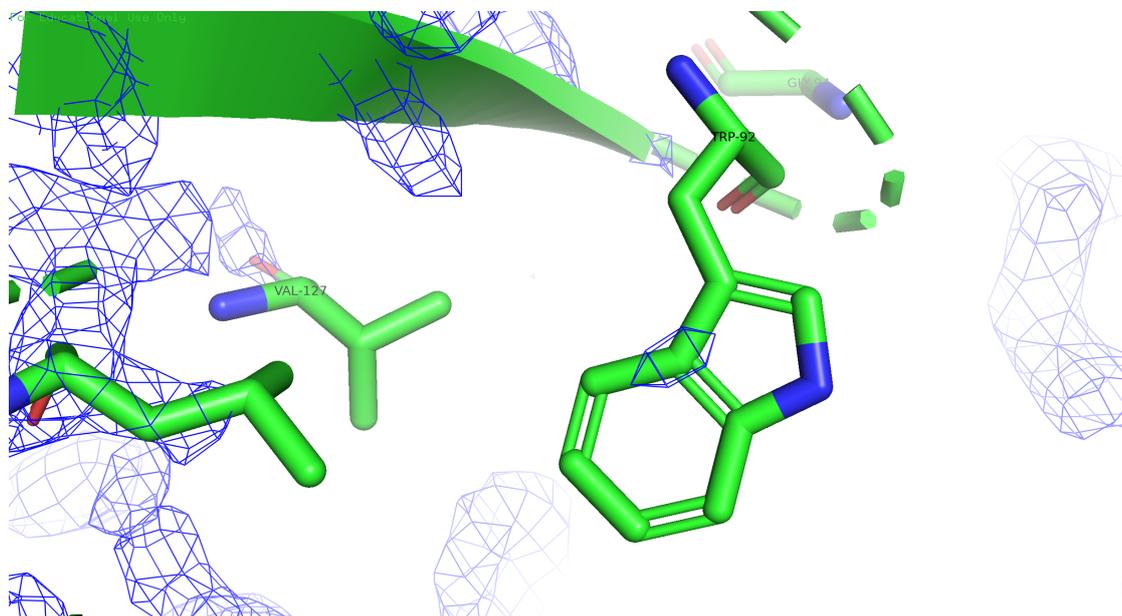


Рис 3. VAL127 и TRP92, не подходящие под электронную плотность(mesh 2, carve 1.5)

## Задание 2

Рассматриваем маргинальные остатки в MolProbity.

### 1. Clashes

Я добавила в модель водороды и посчитала число пар атомов, для которых перекрываются Ван-дер-Вальсовы радиусы. Таких получилось 30 штук. В качестве примера приведу пару 97-CYS атом SG и 127-VAL атом CG2. Как показано на картинке, расстояние между атомами 2.8 (mesh 3), а перекрытие 0.6. Значение меньше, чем сумма Ван-дер-Ваальсовых радиусов атомов.

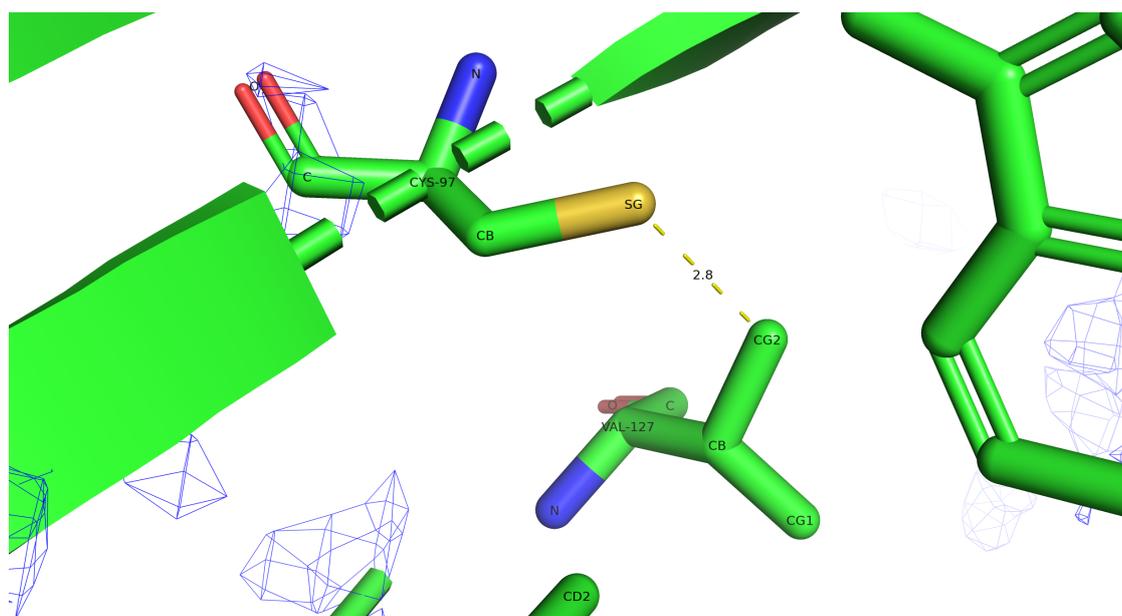


Рис 4. Расстояние(желтый пунктир) между атомами (97-CYS) SG и (127-VAL) CG2 (mesh 3, carve 1.5)

## 2. Bond angles

Всего обнаружено 4 маргинальных остатка по углам ковалентных связей.  
Z-оценка длины связи (или угла) - это количество стандартных отклонений, на которое наблюдаемое значение отклоняется от ожидаемого значения.  
В качестве примера возьму 61-LEU, атомы угла N-CA-C. Z-score=6.41, observed=128, ideal=111.

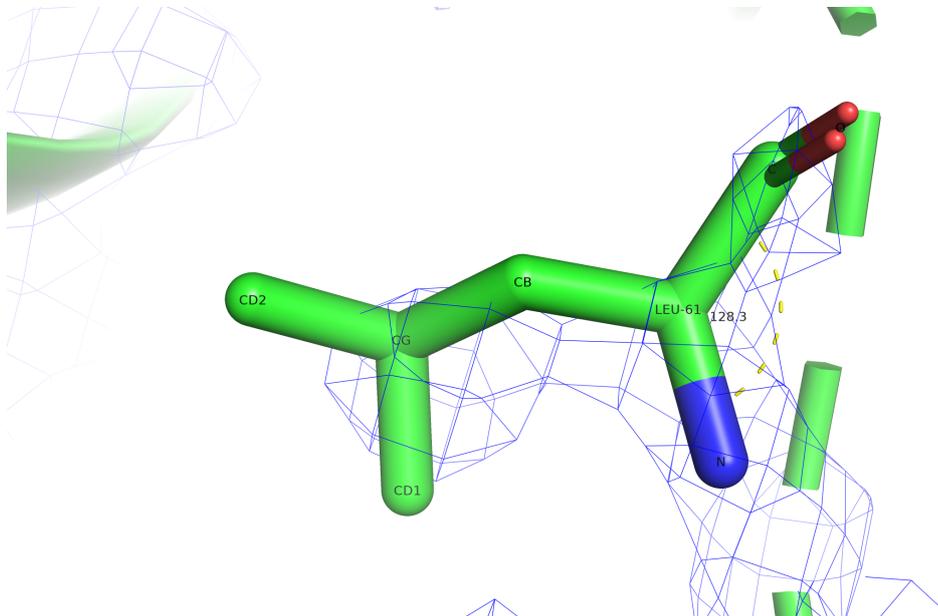


Рис 5. Угол LEU-61 между атомами N-CA-C, равный 128.3, показан желтым пунктиром. Синим- не полностью покрывающая остаток ЭП (mesh 2).

## 3. Bond length

Обнаружена 1 пара атомов CB-SG в 97-CYS с отличной длиной связи. Z-score=-5.34, observed=1.73Å, ideal=1.81Å.

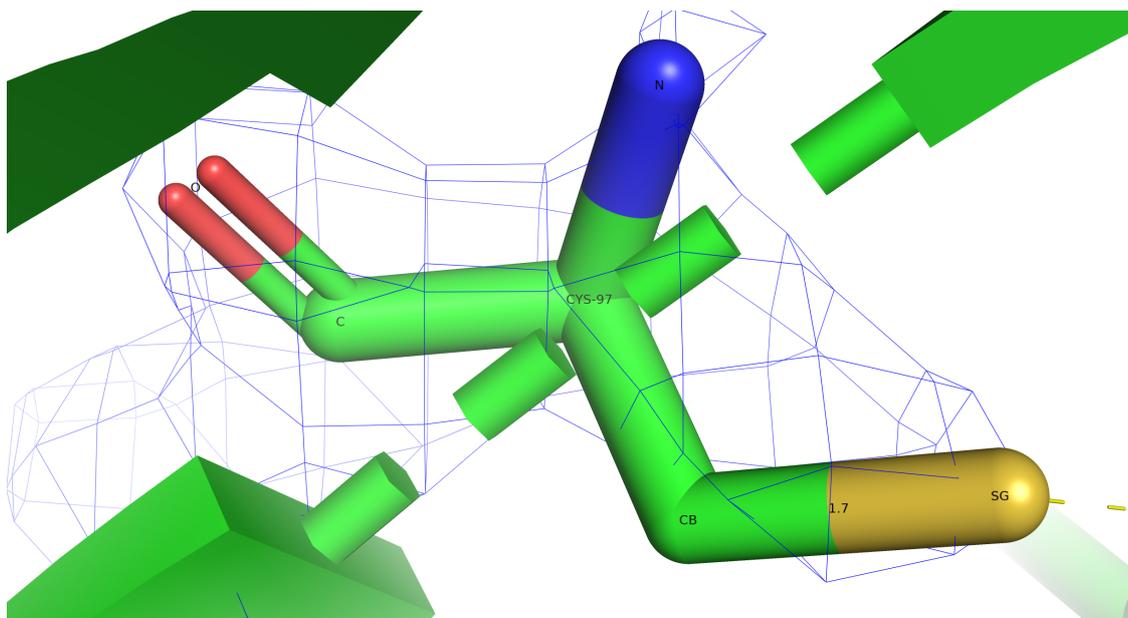


Рис 6. Длина связи 1.7 между атомами CB и SG. Синим- не полностью покрывающая остаток ЭП (mesh 2).

#### 4. Torsion angles (Protein backbone)

Были обнаружены маргиналы по торсионным углам остова.  
В благоприятную область попали 115 (95%) остатков.  
В разрешенную 4 (3%), маргиналами оказались 2 (2%).  
Пример одного из маргиналов на картинке. Это 109-GLU.

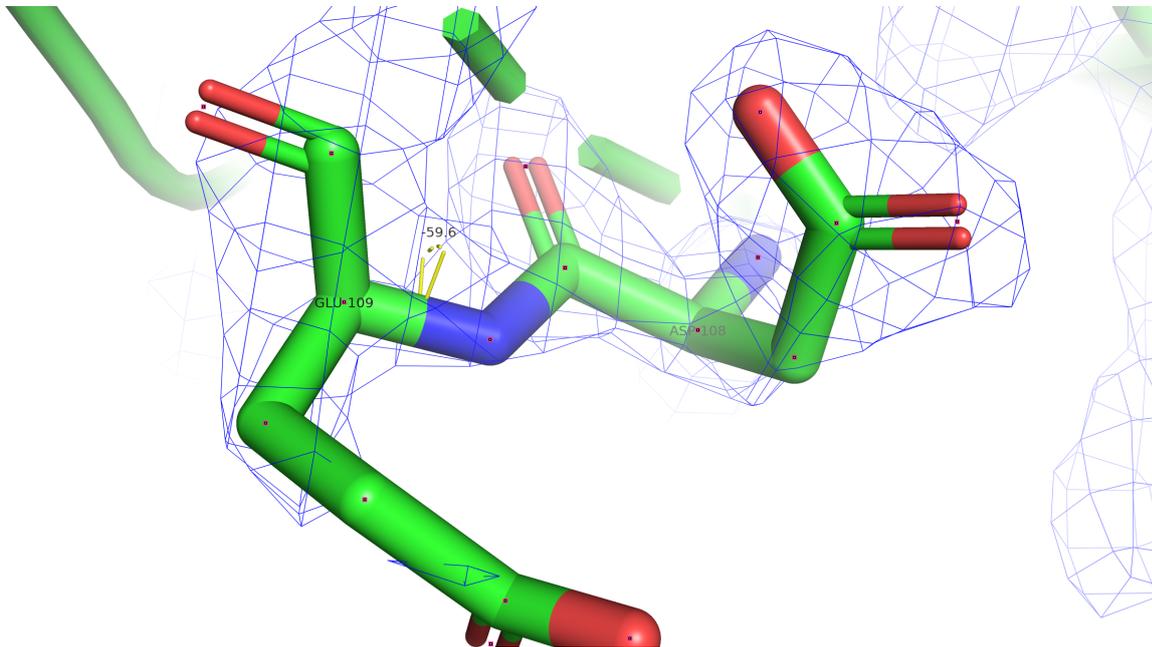


Рис 7. Маргинальный 109-GLU по торсионному углу со значением -59.6. Синим- не полностью покрывающая остаток ЭП (mesh 2).

Также этот остаток мы видим на карте Рамачандрана.

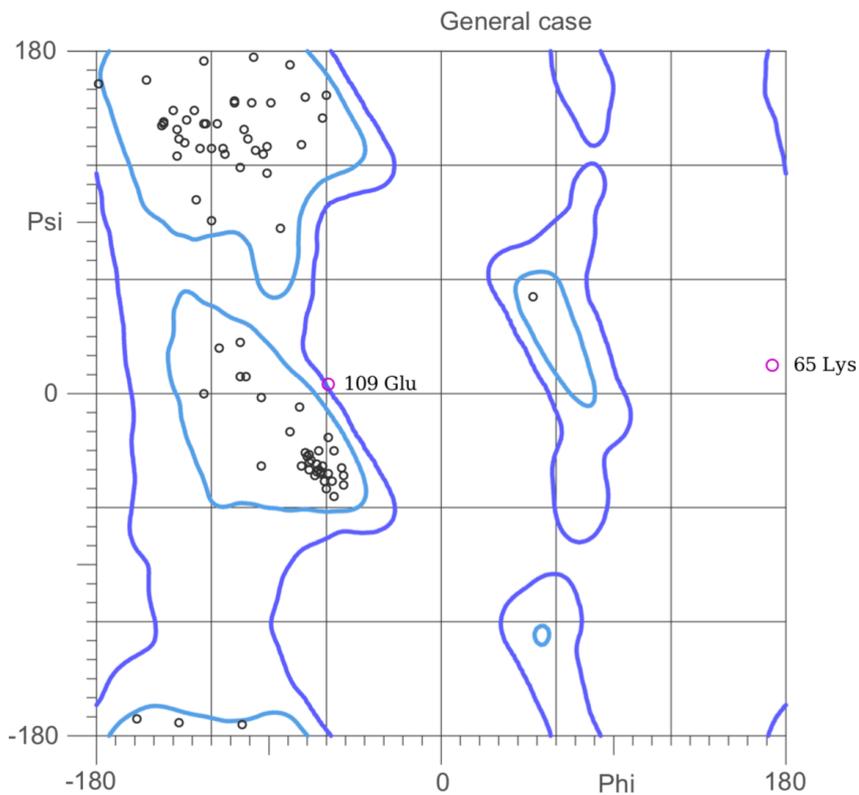


Рис 8. Карта Рамачандрана. Остатки 109 Glu и 65 Lys маргинальные. Черным отмечены остатки, которые попали в разрешенные области.

## 5. Torsion angles (Protein sidechains )

Также рассмотрены маргиналы по торсионным углам остатков.

Ротамеры- 94 (94%)

Маргиналы- 6 (6%)

Например, 37-THR.

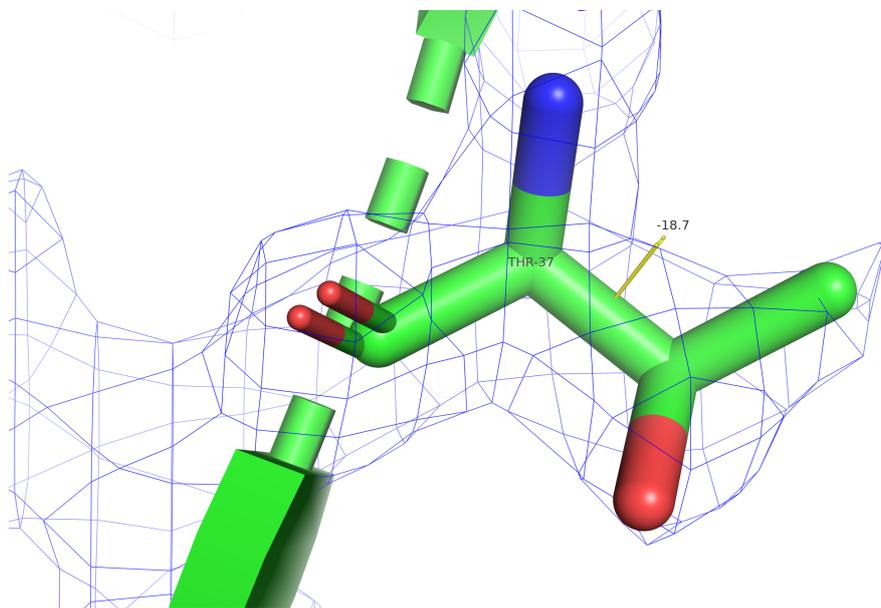


Рис 9. Остаток THR-37. Торсионный угол показан желтым пунктиром. Полное соответствие ЭП(синим, mesh 2, carve 1.5)

## Задание 3.

Я думаю, что данная модель неидеальна для изучения особенностей структуры, так как далеко не все остатки покрыты электронной плотностью, включая значимые остатки, а также присутствует достаточное количество маргиналов по перекрыванию Ван-дер-Ваальсовых радиусов.

С другой стороны, если говорить о маргиналах по углам и длинам связей, их не так много. Поэтому в целом можно использовать структуру с такой геометрией для определения стабильности и функций белка, для чего и нужна была данная структура (исходя из оригинальной статьи).

## Задание 4.

Теперь проведем анализ структуры с помощью PDB redo( <https://pdb-redo.eu/db/1dmn> ). Банк данных PDB-REDO содержит оптимизированные версии существующих записей PDB с картами электронной плотности, описанием изменений модели и большим количеством данных проверки модели.

Значение для R и R-free теперь равны 0.157(было 0.1879) и 0.2054 (было 0.224) соответственно. Однако PDB-REDO выдает, что значения на PDB отличаются от значений, которые действительно были на PDB(0.195 и 0.244 соответственно ). Но в любом случае значения улучшились.

Для многих остатков качество улучшилось, и нет таких, для которых качество PDB-REDO было бы хуже. Ниже пример наиболее улучшенных остатков: LYS65, LEU61, VAL127.

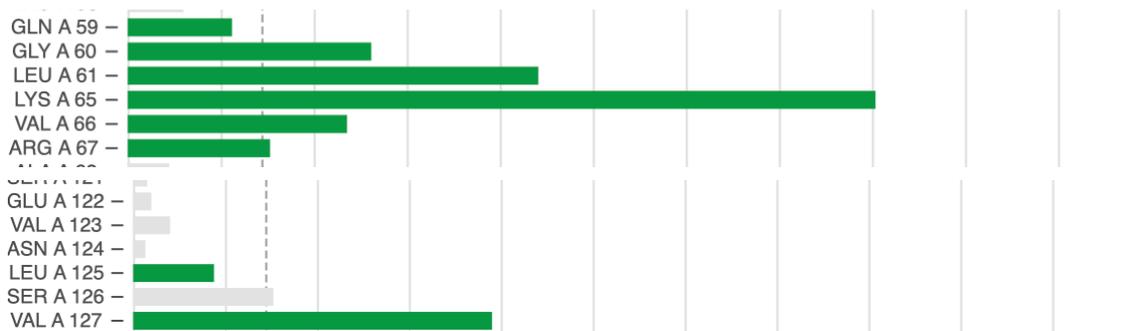


Рис 10. Остатки из выдачи PDB-REDO. Зеленым-улучшенные, серым-не изменившиеся.

Посмотрим на подробнее на структуру этих остатков.

Видим, что LEU61 стал лучше вписываться в электронную плотность (сравним с рисунком 5).

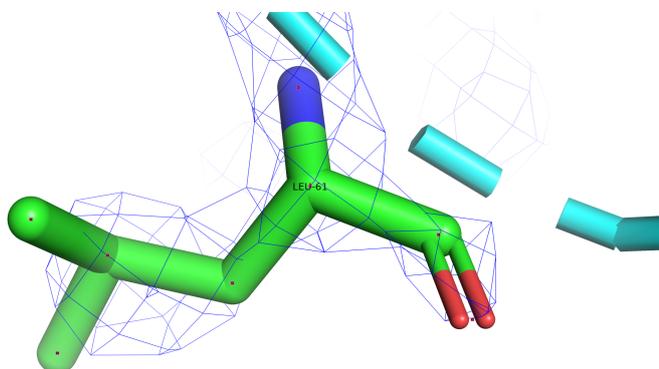


Рис 11. Остаток LEU-61 в PDB-REDO.

LYS65 как не обладал электронной плотностью, так и не обладает. Однако он, судя по новой карте Рамачандрана, больше не является маргиналом и попадает в разрешенную область, как и 109-GLU (рис 13).

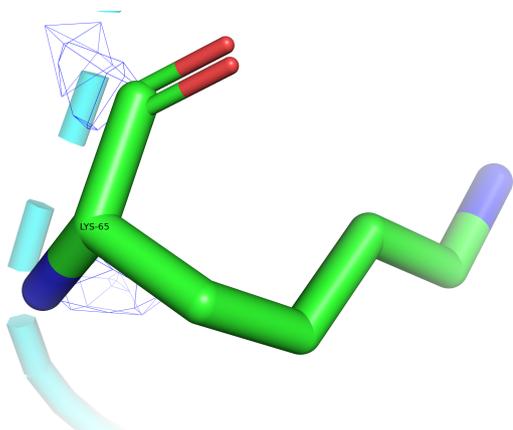
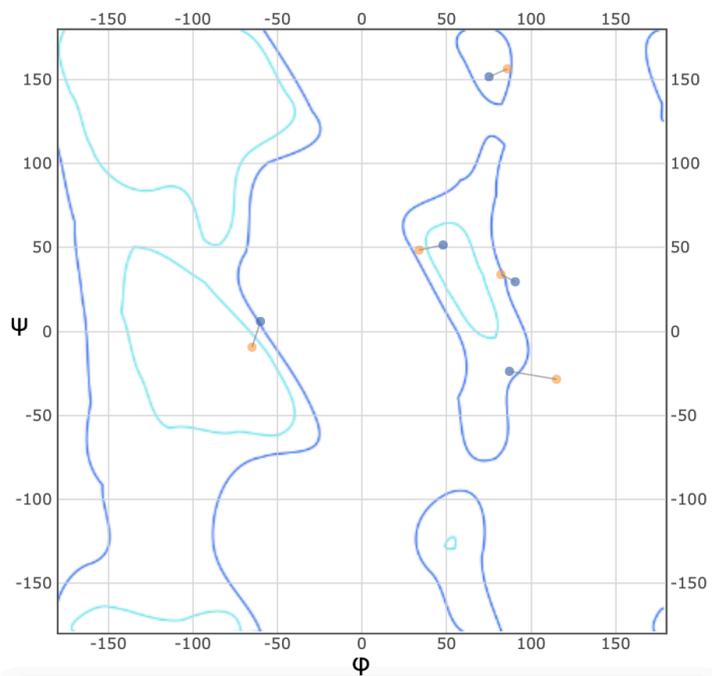


Рис 12. Остаток LYS-65 в PDB-REDO.



Description	PDB ●	PDB-REDO ●
Ramachandran Z-score	-1.432	0.136
<input type="checkbox"/> Preferred regions	116	119
<input checked="" type="checkbox"/> Allowed regions	3	5
<input checked="" type="checkbox"/> Outliers	0	0

Рис 13. Карта Рамачандрана.

VAL127 стал чуть лучше вписываться в электронную плотность.

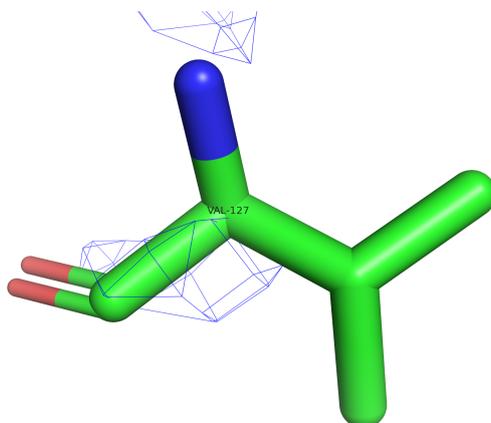


Рис 14. Остаток VAL-127. Синим-ЭП (mesh 2, carve 1.5)

Так как для многих остатков качество улучшилось и нет ни одного, для которого оно бы стало хуже, я считаю, что можно использовать модель PDB-REDO для изучения.

## Литература

- 1) Structural double-mutant cycle analysis of a hydrogen bond network in ketosteroid isomerase from *Pseudomonas putida* biotype B, Do Soo Jang
- 2) Pseudoreversion of the catalytic activity of Y14F by the additional substitution(s) of tyrosine with phenylalanine in the hydrogen bond network of delta 5-3-ketosteroid isomerase from *Pseudomonas putida* biotype B
- 3) Contribution of the Hydrogen-Bond Network Involving a Tyrosine Triad in the Active Site to the Structure and Function of a Highly Proficient Ketosteroid Isomerase from *Pseudomonas putida* Biotype B†,‡. *Biochemistry*