

# Оценка качества расшифровки белка COX-2 из Homo Sapiens (PDB ID:5IKQ)

Азбукина Надежда

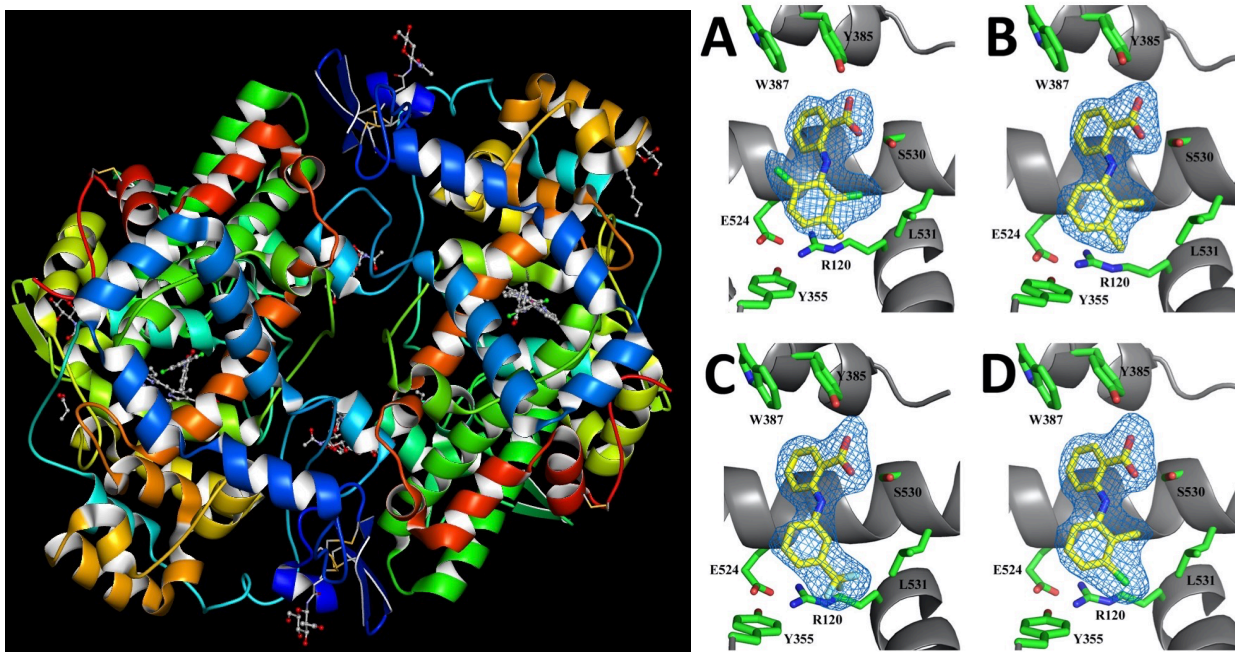
## Аннотация

В работе был проведен анализ качества расшифровки структуры белка COX-2 (5IKQ). Данные о структуре белка были получены методом рентгеноструктурного анализа. Были оценены основные параметры оценки качества полученной модели и подробно рассмотрены некоторые маргинальные остатки.

## Введение

Циклооксигеназа-2 (COX-2) катализирует реакцию окисления арахидоновой кислоты и эндоканнабиноидных производных, продукты этих реакции являются важными сигнальными медиаторами. COX-2 в клетке существует в виде гомодимера, однако измеренная ферментативная активность показывает, что в один момент времени активен только один мономер. Показано, что некоторые нестероидные противовоспалительные препараты ингибируют COX-2 субстрат-специфичным способом, при этом связывание ингибитора с одним из мономеров достаточно, чтобы ингибировать окисление эндоканнабиноидов, но не арахидоновую кислоту. При этом механизм, обеспечивающий субстратную специфичность ингибирования остается неизвестным. В данной работе авторы получили структуры COX-2 с четырьмя различными ингибиторами. Получено, что эти ингибиторы связываются с циклооксигеназным каналом в инвертированной ориентации и взаимодействуют с Tyr-385 и Ser-530, расположенными в начале канала. Кроме того в статье было показано, что субстратная специфичность ингибирования зависит от «пероксидной нагрузки» окружения, что позволяет подбирать лекарства в зависимости от ткани – мишени. Например, известно, что в тканях мозга низкая «пероксидная нагрузка», поэтому для ингибирования образования окисленных производных арахидоновой кислоты в тканях мозга, не затрагивая метаболизм эндоканнабиноидов, необходимо использовать определенный спектр ингибиторов, например, ацетаминофен [1].

Циклооксигеназа содержит два типа доменов: EGF – подобные домены (PF00008) и домен из семейства гем-зависимых пероксидаз животных (PF03098). Циклооксигеназный активный сайт, который катализирует реакцию циклизации, расположен на конце длинного гидрофобного канала, который продолжается из мембран-связывающего домена в центре молекулы. Пероксидазный активный сайт, катализирующий окислительную реакцию, расположен с другой стороны молекулы, в зоне связывания с гемом. Активные центры белка преимущественно образованы различными альфа-спиралями (всего белок содержит 19 альфа-спиралей) [2].



**Рисунок 1. (Слева) 5ikq. (Справа) Связь производных фенаминовой кислоты с каналом COX-2 [1].**

## Результаты

### *Общая информация о модели*

Структура находится в банке PDB с 3-его марта 2016 года. Для наработки белка для получения кристалла использовалась экспрессионная система *SPODOPTERA FRUGIPERDA* [3]. Всего в работе получено 4 структуры циклооксигеназы с разными ингибиторами: 5IKV, 5IKQ, 5IKR и 5IKT.

Разрешение структуры - 2.41 Å, диапазон разрешений составляет 2.41-39.20 Å, полнота набора гармоник – 94.8%. Число измеренных рефлексов – 53402. Для решения фазовой проблемы исследователи использовали метод молекулярного замещения с использованием программы PHASER, в качестве референса была взята молекула 3HS5 [4].

Файл содержит 18634 атома, содержит два полимера, то есть две белковые цепи: А и В, являющиеся идентичными молекулами длиной 551 а.о. Помимо основных белковых цепей в файле содержится несколько простетических групп: бета-октилглюкозид (1 шт), 2-[(2,6-дихлоро-3-метил-фенил)амино]бензойная кислота, являющаяся ингибитором (2 шт), N-ацетил-D-глюкозамин (8 шт), альфа-D-манноза (4 шт), протопорфириновое кольцо (2 шт), акриловая кислота (3 шт), вода (416 шт). В статье указывается, что протопорфириновое кольцо содержало ионы кобальта, а не железа, в данной экспрессионной системе для повышения устойчивости. Протопорфириновое

кольцо входит в состав каталитического центра, а сахара попали в кристалл из-за технических особенностей выделения белка.

Биологическая единица представляет собой гомодимер. Асимметрическая единица (белковой цепи) не совпадает с биологической. Имеется некристаллографическая симметрия внутри ячейки (цепи А и В). Тип кристаллической решетки  $I_{222}$ . Молекул в кристалле – 16 штук.

#### *Качество модели*

Общее качество модели классически оценивается по R-фактору, R-free, RMS Z-score для длин и углов. В таблице 1 приведены значения этих параметров. R-фактор отражает соответствие рассчитанных и экспериментальных структурных факторов. Хорошим можно считать показатель меньше 0,25. У рассматриваемой модели он 0,173, что говорит о хорошем качестве. R-free является контролем переоптимизации. Считается, что он не должен быть больше R-фактора на 10%. У данной модели они различаются на 8,15%, что является вполне допустимым значением, но все же близким к критической границе. Среднеквадратичные отклонения длин и углов в связях являются довольно маленькими, что свидетельствует о качестве модели.

**Таблица 1.** Параметры качества модели

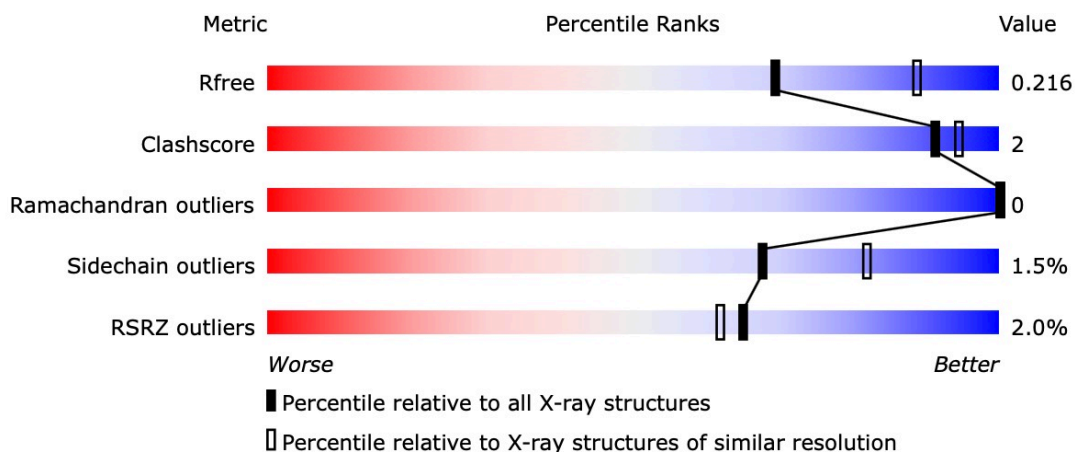
<b>Параметр</b>	<b>Значение</b>
R	0.1728
R-free	0.21170
Bond length RMSD (в ангстремах)	0.010
Bond angle RMSD (в градусах)	0.854

Помимо оценки общего качества модели при валидации структуры обычно оцениваются маргинальные остатки.

**Таблица 2.** Оценка маргинальных остатков (согласно MolProb)

<b>Параметр</b>	<b>Число маргиналов</b>	<b>% маргиналов</b>
Длины связей	8	0,7
Углы связей	8	0,7
Цис-пептиды	2	0,2
Карты Рамачандрана	0	0
Ротамеры	17	1,7
Перекрытия (>0.4Å)	48	4,5

На рисунке 2 представлена оценка качества исследуемой структуры относительно всех моделей банка PDB. Видно, что по всем параметрам качество представленной модели выше среднего по PDB (смещено в синюю область).



**Рисунок 2.** Оценка качества модели 5ikq относительно банка PDB. Красный цвет – смещение в худшую сторону, синий – в лучшую, белый – на уровне среднего. Черная метка показывает расположение относительно всех PCA структур, белая – относительно структур, имеющих такое же разрешение. Источник: <https://pdbj.org/mine/summary/5ikq>

Для более детального анализа использовался сервис MolProbity [5]. С его помощью в структуру были добавлены атомы водорода, оптимизированы водородные связи, произвелись инверсии боковых цепей аспарагина, глутамина и гистидина. Изначально в модели не было водородов, программа добавила 8995 атомов, позиции 255-ти из были изменены для оптимизирования водородных связей. Результат работы программы представлен на рисунке 3.

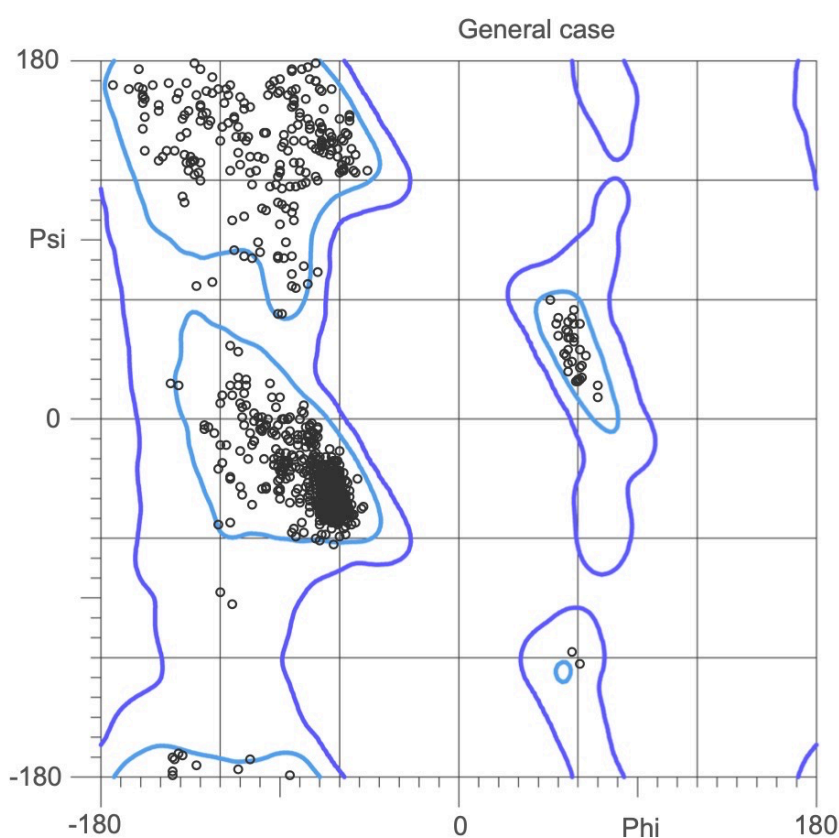
All-Atom Contacts	Clashscore, all atoms:	1.88	100 <sup>th</sup> percentile* (N=328, 2.41Å ± 0.25Å)
	Clashscore is the number of serious steric overlaps (> 0.4 Å) per 1000 atoms.		
Protein Geometry	Poor rotamers	17	1.73% Goal: <0.3%
	Favored rotamers	939	95.82% Goal: >98%
	Ramachandran outliers	0	0.00% Goal: <0.05%
	Ramachandran favored	1075	97.91% Goal: >98%
	MolProbity score <sup>^</sup>	1.15	100 <sup>th</sup> percentile* (N=7825, 2.41Å ± 0.25Å)
	Cβ deviations >0.25Å	0	0.00% Goal: 0
	Bad bonds:	28 / 9458	0.30% Goal: 0%
	Bad angles:	10 / 12828	0.08% Goal: <0.1%
Peptide Omegas	Cis Prolines:	2 / 76	2.63% Expected: ≤1 per chain, or ≤5%

**Рисунок 3.** Индикаторы качества структуры 3o48, определенные сервисом MolProbity

Clashscore показывает число перекрываний на 1000 атомов. В качестве перекрываний рассматривается пересечение Ван-дер-Ваальсовых радиусов более, чем 0,4 ангстрема. Для данной модели этот параметр равен 1.88, входит в 100-ый перцентиль, что соответствует наилучшему качеству для моделей данного разрешения.

Для оценки конформации боковых цепей используют ротамеры, которые представляют собой наборы разрешенных положений боковых цепей для данного остатка, которые заданы допустимыми значениями торсионных углов. Согласно MolProbity «хорошей» можно назвать модель, содержащую менее 0,3% недопустимых остатков. В исследуемой структуре только 95,82% остатков являются «хорошими» и 1,73% - «плохими» (их 17 штук), то есть завышен в 6 раз, что свидетельствует о том, что модель требует доработки «вручную».

Для оценки качества восстановления остовов используются карты Рамачандрана, отражающие допустимые углы вращения между атомами остова. На рисунке 4 показана карта Рамачандрана для всех аминокислот.



**Рисунок 4.** Карта Рамачандрана для структуры 5ikq. Голубые линии ограничивают участки предпочтительных областей, синие – допустимых областей. Точками представлены аминокислотные остатки.

97,9% остатков модели лежат в предпочтительной области, 2,1% - в разрешенной, 0%- аутлайнеры. Таким образом можно сделать вывод, что по аутлайнеров по этому параметру в модели нет. Кроме того сервис отдельно показывает карты для изолейцина и валина, глицина, цис- и транс- пролинов ввиду особенностей возможностей их торсионных углов. Было обнаружено 2

цис-пролина, однако значения их торсионных углов лежат в предпочтительных областях.

Параметр RSR характеризует соответствие а.о. экспериментальным данным. Традиционно оценка производится в виде Z-score, показывающую, насколько RSR для остатка данной модели отличается от среднего значения RSR по структурам PDB такого же разрешения. Маргинальным считается остаток, обладающий Z-score >2.

Доля маргинальных остатков по RSR для данной модели составляет 1% (22/1102), что лучше, чем средний показатель по PDB.

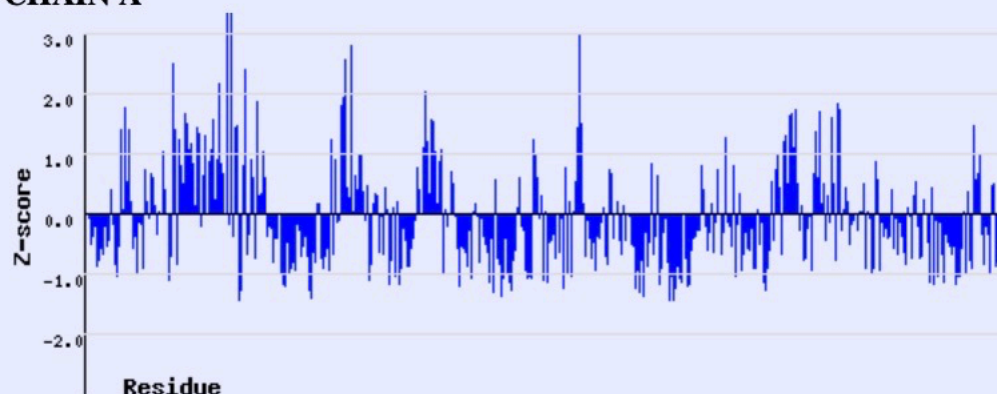
Значения Z-score для отдельных остатков показаны на рисунке 5. Список аутлайнеров приведен ниже.

**Таблица 3.** Маргинальные остатки согласно критерию RSRZ. Источник: <https://pdj.org/rest/displayPDBfile?format=val-pdf&id=5ikq>

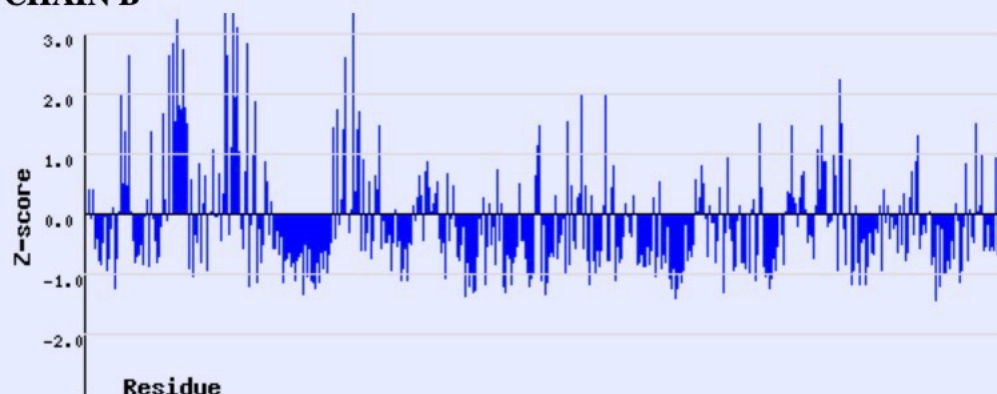
Mol	Chain	Res	Type	RSRZ
1	B	108	PHE	3.2
1	A	206	THR	3.1
1	B	106	ILE	3.1
1	B	81	PHE	2.8
1	B	74	PHE	2.7
1	A	96	PHE	2.7
1	A	202	ALA	2.6
1	A	530	SER	2.6
1	A	99	PHE	2.5
1	A	387	TRP	2.4
1	A	98	GLY	2.4
1	A	205	PHE	2.3
1	A	160	PRO	2.3
1	A	103	VAL	2.3
1	A	381	PHE	2.2
1	A	526	GLY	2.2
1	A	385	TYR	2.1
1	B	78	ILE	2.1
1	A	108	LEU	2.1
1	A	515	ASP	2.1
1	B	490	ALA	2.1
1	B	82	LEU	2.0

Chain	Nres	<Z>	Sigma	%<-1	%<0	%>1	%>2
A	551	-0.08	0.87	10.71	62.07	12.52	2.18
B	551	-0.19	0.90	12.34	67.70	8.71	2.54
All	1102	-0.13	0.89	11.52	64.88	10.62	2.36

#### CHAIN A



#### CHAIN B



**Рисунок 5.** Значения Z-score для а.о. 5ikq. Источник: <http://eds.bmc.uu.se/cgi-bin/eds/zscore?5ikq>

С помощью сервиса WHAT IF [6] был проведен анализ исследуемой модели. Согласно проведенному анализу не было задетектировано ошибочный ковалентных связей (все в пределах одной полипептидной цепи).

Для некоторых атомов боковых цепей аминокислот 74 Phe A, 75 Leu A, 78 Ile A, 81 Phe A, 82 Leu A, 102 Leu A, 121 Ser A, 158 Asp A, 215 Lys A, 53 Asp B, 75 Leu B, 109 Leu B, 159 Asp B, 170 Lys B, 400 Asp B, 486 Lys B была обнаружен низкий уровень заполненности близкий к нулю. Это связано с техническими сложностями кристаллографического эксперимента. Кроме того можно заметить, что проблемы связаны преимущественно с лейцином, аспарагином и лизином – остатками, имеющие длинные боковые цепи.

Атомы остова, напротив, все имеют высокий уровень заполненности, свидетельствующий о хорошей расшифровке модели. Кроме того все В-факторы лежат в диапазоне [0-100], что также свидетельствует о высоком уровне расшифровки модели.

Далее был оценен уровень «упаковки» аминокислот, показывающий оптимальность их окружения. Для каждой аминокислоты указана мера качества окружения остов-остов, боковая цепь-остов, боковая цепь – боковая цепь. Если это число ниже -2,5, то остаток считается «подозрительным» и требующим дополнительного внимания. Было найдено всего 4 «проблемных» остатка: LEU ( 124 )В, ILE ( 378 )В, LYS ( 244 )В, ASN ( 376 )В. Ни один из этих остатков не связан с лигандами и не является каталитически активным остатком, что позволяет отнести их к маргинальным. Z-score для всех взаимодействий оказался -1,68, что является показателем хорошей структуры.

На последнем этапе была посчитана доступная молекулярная поверхность для каждого остатка. Было найдено несколько гидрофобных остатков с аномально большой доступной площадью:

PHE ( 52 )А  
PHE ( 74 )А  
LEU ( 75 )А  
PHE ( 81 )А  
PHE ( 107 )А  
LEU ( 224 )А  
ILE ( 256 )А  
PHE ( 52 )В  
PHE ( 74 )В  
PHE ( 81 )В  
PHE ( 108 )В  
ILE ( 275 )В  
PRO ( 278 )В  
PHE ( 578 )В

В дальнейшем они будут рассмотрены подробнее.

Примечательно также, что модель хорошо построена не только для пептидной цепи, но и для определенных лигандов. Все лиганды имеют заполненность > 0.9, что является индикатором хорошего восстановления структуры.

**Таблица 4.** Соответствие экспериментальных и модельных данных для лигандов структуры. Источник: <https://pdj.org/rest/displayPDBfile?format=val-pdf&id=5ikq>

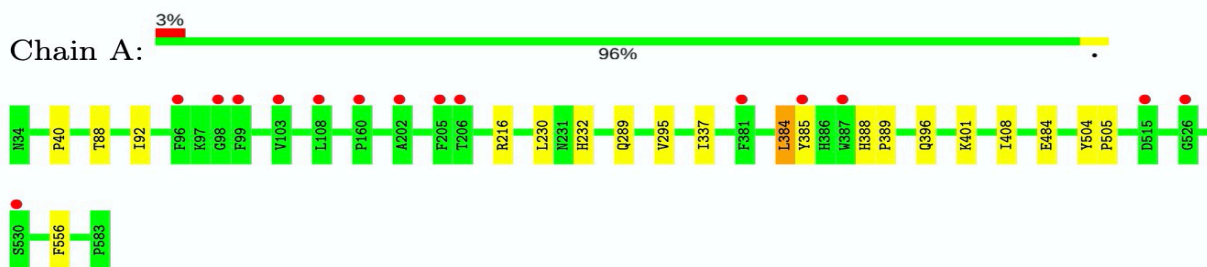


Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å <sup>2</sup> )	Q<0.9
5	MAN	B	607	11/12	0.53	0.33	89,94,112,116	0
5	MAN	A	607	11/12	0.64	0.32	89,109,129,132	0
5	MAN	B	606	11/12	0.69	0.25	70,84,101,101	0
5	MAN	A	606	11/12	0.71	0.33	72,91,109,113	0
6	COH	A	609	43/43	0.77	0.28	57,79,93,98	0
4	NAG	A	603	14/15	0.79	0.38	51,70,84,86	0
4	NAG	B	603	14/15	0.79	0.35	61,77,93,93	0
4	NAG	A	608	14/15	0.82	0.29	56,69,81,81	0
6	COH	B	601	43/43	0.83	0.23	53,82,99,101	0
7	AKR	B	610	5/5	0.83	0.21	60,61,72,74	0
4	NAG	B	608	14/15	0.89	0.26	45,57,68,69	0
2	BOG	A	601	20/20	0.92	0.19	41,64,79,81	0
7	AKR	B	609	5/5	0.93	0.23	40,46,60,60	0
4	NAG	A	605	14/15	0.93	0.23	43,57,70,73	0
7	AKR	A	610	5/5	0.93	0.23	41,48,61,61	0
3	JMS	A	602	19/19	0.95	0.21	32,44,53,56	0
4	NAG	B	605	14/15	0.96	0.12	30,44,57,64	0
4	NAG	B	604	14/15	0.96	0.13	19,35,42,43	0
4	NAG	A	604	14/15	0.96	0.13	27,41,46,49	0
3	JMS	B	602	19/19	0.97	0.15	28,35,43,70	0

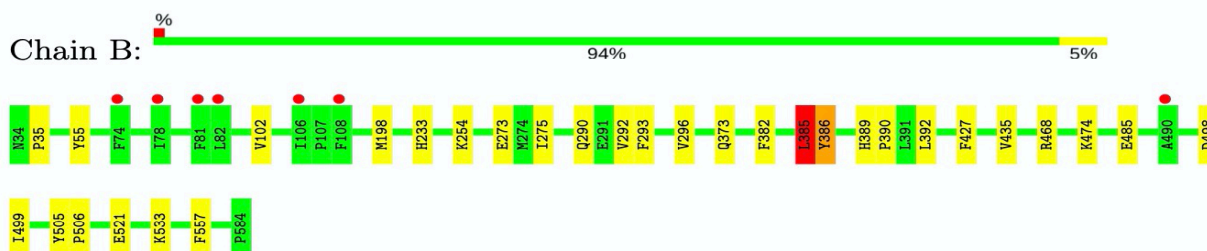
### Анализ маргинальных остатков

На рисунке 6 представлена визуализация качества восстановления а.о. остатков в цепях белка. Остатки раскрашены относительно маргинальности по геометрическим параметрам. Зеленым – 0 параметров, желтым – 1, оранжевым – 2, красным – 3 и более. Красная точка сверху остатка свидетельствует о маргинальности по RSR. В целом картина довольно хорошая, всего около 4% маргинальных остатков.

- Molecule 1: Prostaglandin G/H synthase 2



- Molecule 1: Prostaglandin G/H synthase 2



**Рисунок 6.** Маргинальные остатки. Источник:  
<https://pdj.org/rest/displayPDBfile?format=val-pdf&id=5ikq>

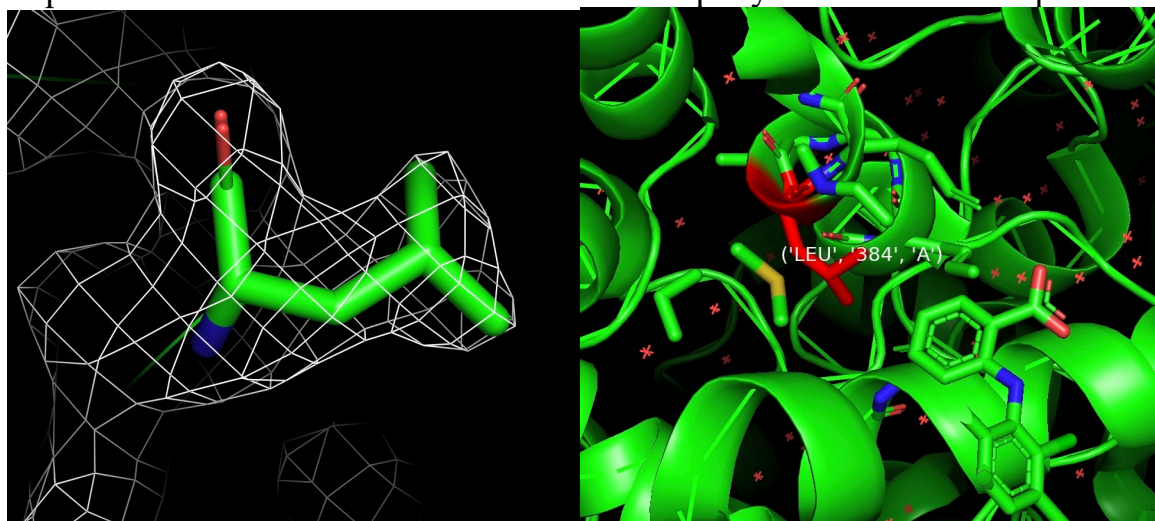
В таблице 5 приведены выборочные маргинальные остатки с указанием параметров маргинальности.

**Таблица 5.** Некоторые маргинальные остатки

Остаток	Причина маргинальности
A384 Leu	«плохой» ротамер; недопустимый угол связи CA-CB-CG
B385 Leu	«плохой» ротамер; недопустимый угол связи CA-CB-CG, недопустимая площадь перекрывания с B 382 PHE HA
B386 Tyr	«плохой» ротамер; недопустимая площадь перекрывания с B 602 JMS OXT
B198 Met	недопустимая площадь перекрывания с B 424 VAL HA
A295 Val	«плохой» ротамер

1) A384 Leu

На рисунке 7 показан 384 лейцин и его электронная плотность (сигма = 0.5). Видно, что атомы остатка вписаны в нее довольно-таки хорошо. На рисунке справа показано окружение этого остатка. Видно, что он входит в состав альфа-спирали. Кроме того относительно недалеко от этого остатка (менее 5 ангстрем) расположена молекула ингибитора JMS. Я предполагаю, что именно и привело к изменению угла в боковой цепи лейцина и является нормальным положением этого остатка в присутствии ингибитора.

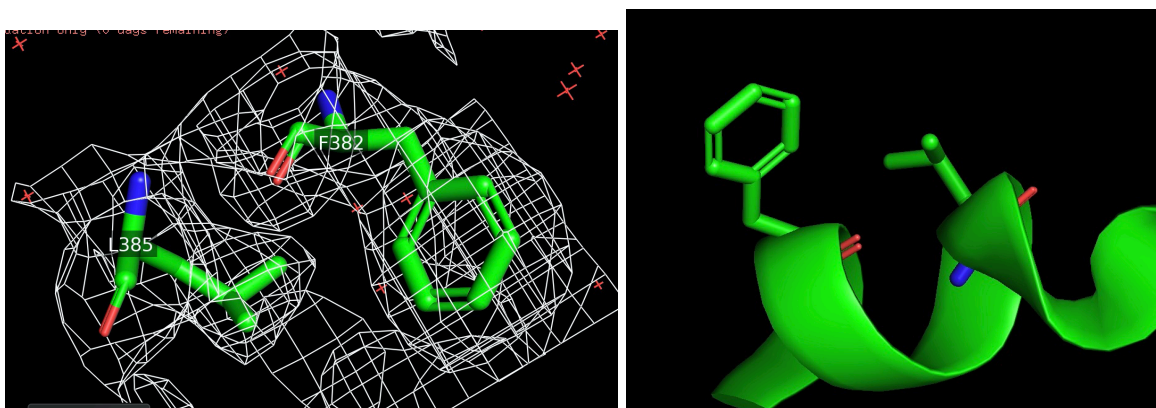


**Рисунок 7.** A384 Leu. На рисунке справа обсуждаемый остаток обозначен красным цветом. Показан в контексте альфа-спиралей.

2) B385 Leu

На рисунке 8 показаны B385 Leu и B382 PHE. Видно, что электронные плотности указанных молекул очень сближены и перекрываются, что является недопустимым. Допустимое значение составляет 0,40 ангстрема, а данные остатки пересекаются 0,44 ангстрема. Учитывая, что они находятся в одной альфа-спирали (рисунок 8 справа) друг над другом, я считаю, что их

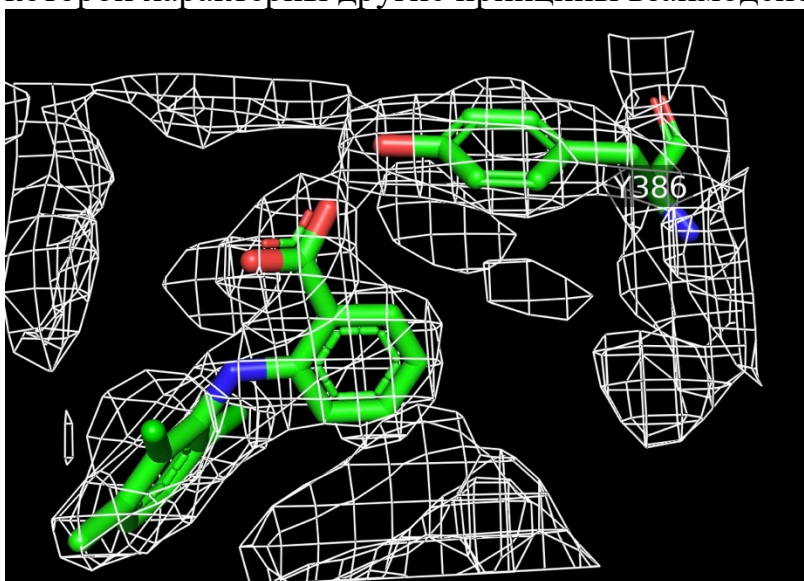
тоже можно исключить из числа маргинальных. Кроме того выглядит это немного похожим на T-стекинг, так как боковая цепь лейцина направлена прямо в ароматическое кольцо фенилаланина.



**Рисунок 8.** Слева показаны электронные плотности B385 Leu и B382 PHE (сигма = 0,5). На рисунке справа обсуждаемые остатки показаны в контексте альфа-спирали.

### 3) B386 Tyr

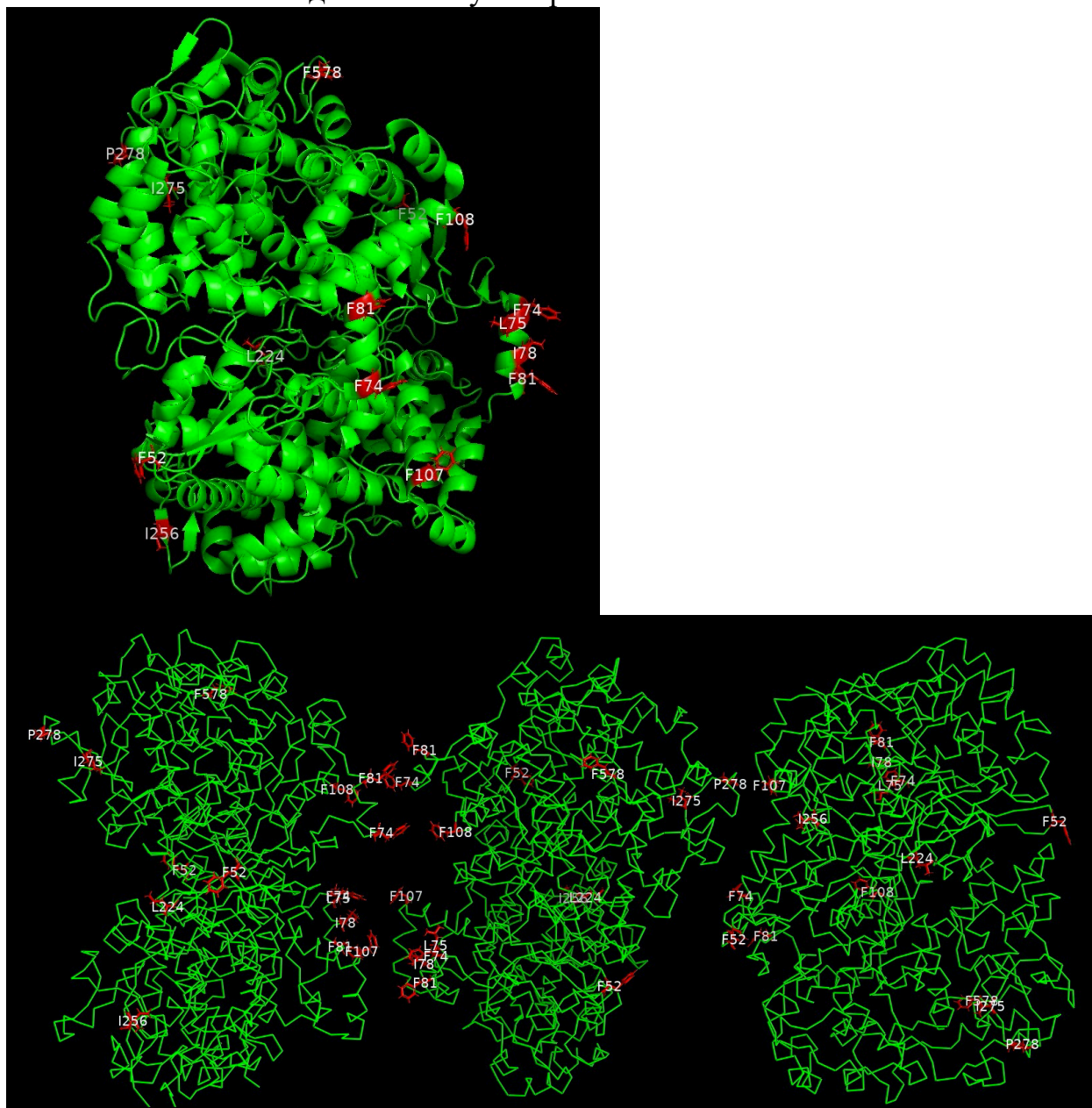
На рисунке 9 показан тирозин и ингибитор меклофенамовая кислота (JMS), окруженные экспериментальными электронными плотностями, сигма = 0.5. Данный остаток был отнесен к маргинальным, так как имеет слишком большую площадь перекрывания с молекулой ингибитора, однако, я подозреваю, что такое перекрывание является нормальным и обеспечивает взаимодействие ингибитора с каналом белка. Кроме того порог для перекрывания в 0,4 ангстрема был выбран исходя из положений аминокислот в белке, а здесь мы имеем дело взаимодействия белка с малой молекулой, для которой характерны другие принципы взаимодействия.



**Рисунок 9.** Показаны электронные плотности B386 Tyr и JMS (сигма = 0,5).

#### 4) Сильно экспонированные гидрофобные остатки

Следующим шагом я провела анализ anomalно экспонированных гидрофобных остатков. На рисунке 10 показана визуализация молекулы белка, где красным отмечены гидрофобные остатки, упомянутые выше при обсуждении программы What if. Чтобы попытаться объяснить причину такой anomalности я восстановила несколько соседних молекул в кристалле. На нижнем рисунке видно, что на самом деле большинство гидрофобных остатков не являются экспонированными в раствор, а взаимодействуют с аминокислотами соседних молекул в кристалле.

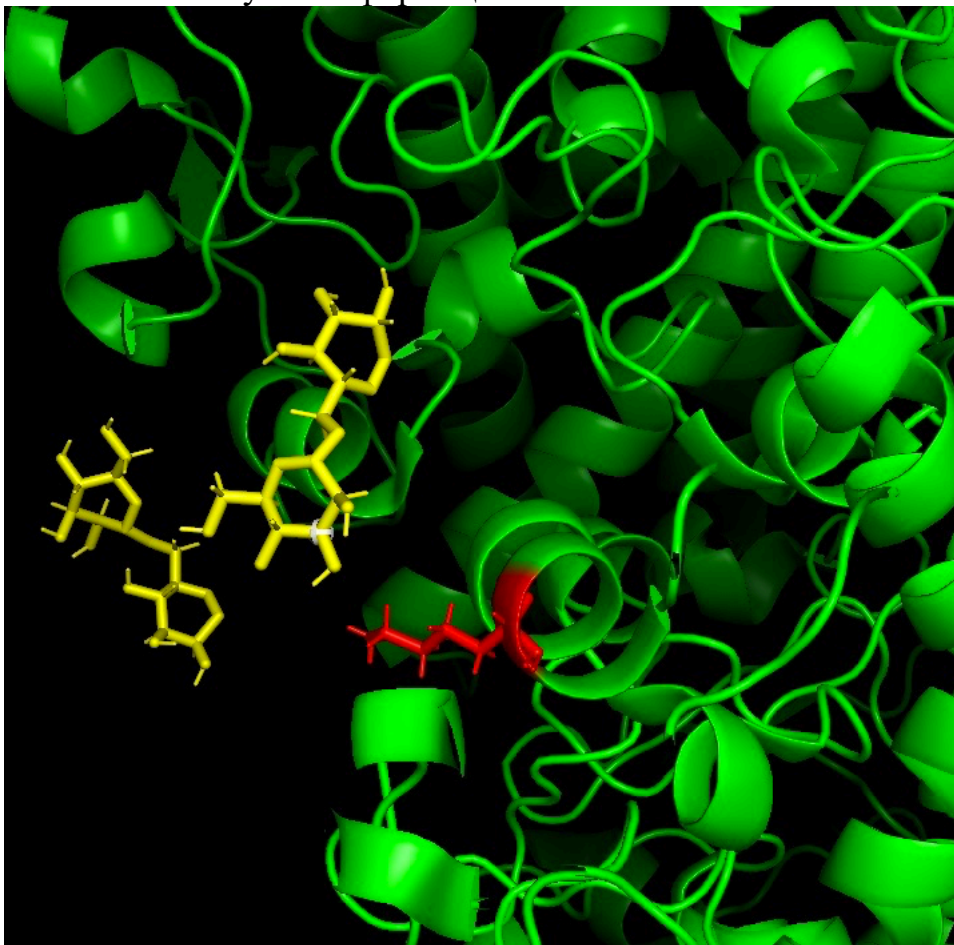


**Рисунок 10.** Anomalно экспонированные аминокислоты показаны красным. На нижнем рисунке показано восстановление нескольких молекул кристалла.

#### 5) B124 LEU

Данный атом был аннотирован как «подозрительный» программой What If в связи с окружением. При более внимательном рассмотрении было

обнаружено, что данная аминокислота связана с молекулой маннозы, из-за этого имеет необычную конформацию.

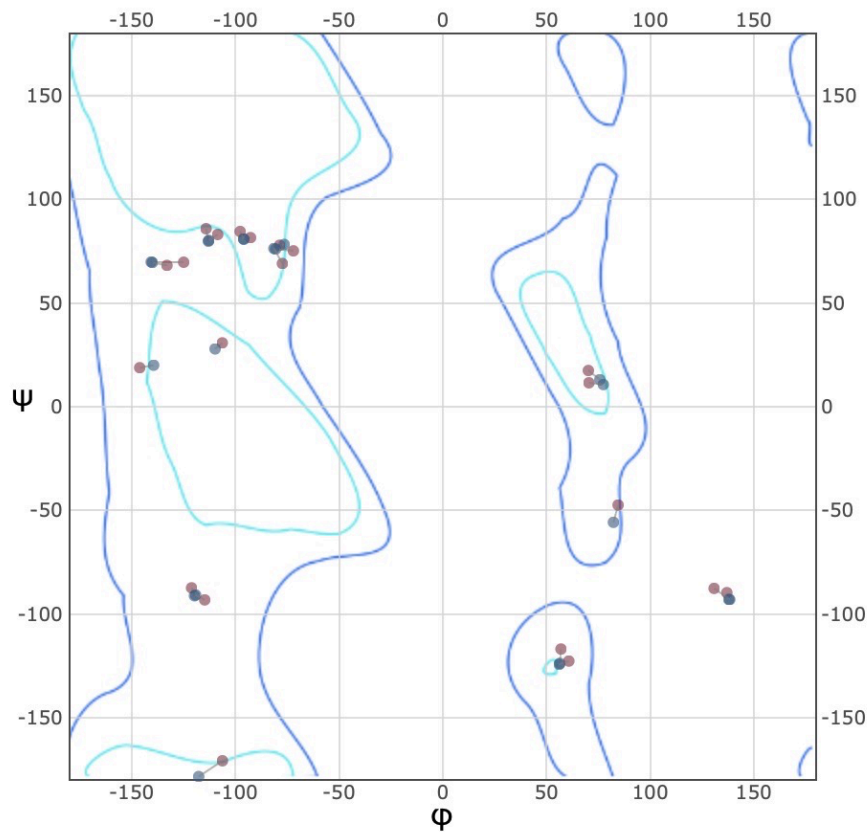


**Рисунок 11.** Лизин 124 В показан красным, манноза – желтым, остальная структура белка – зеленым.

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

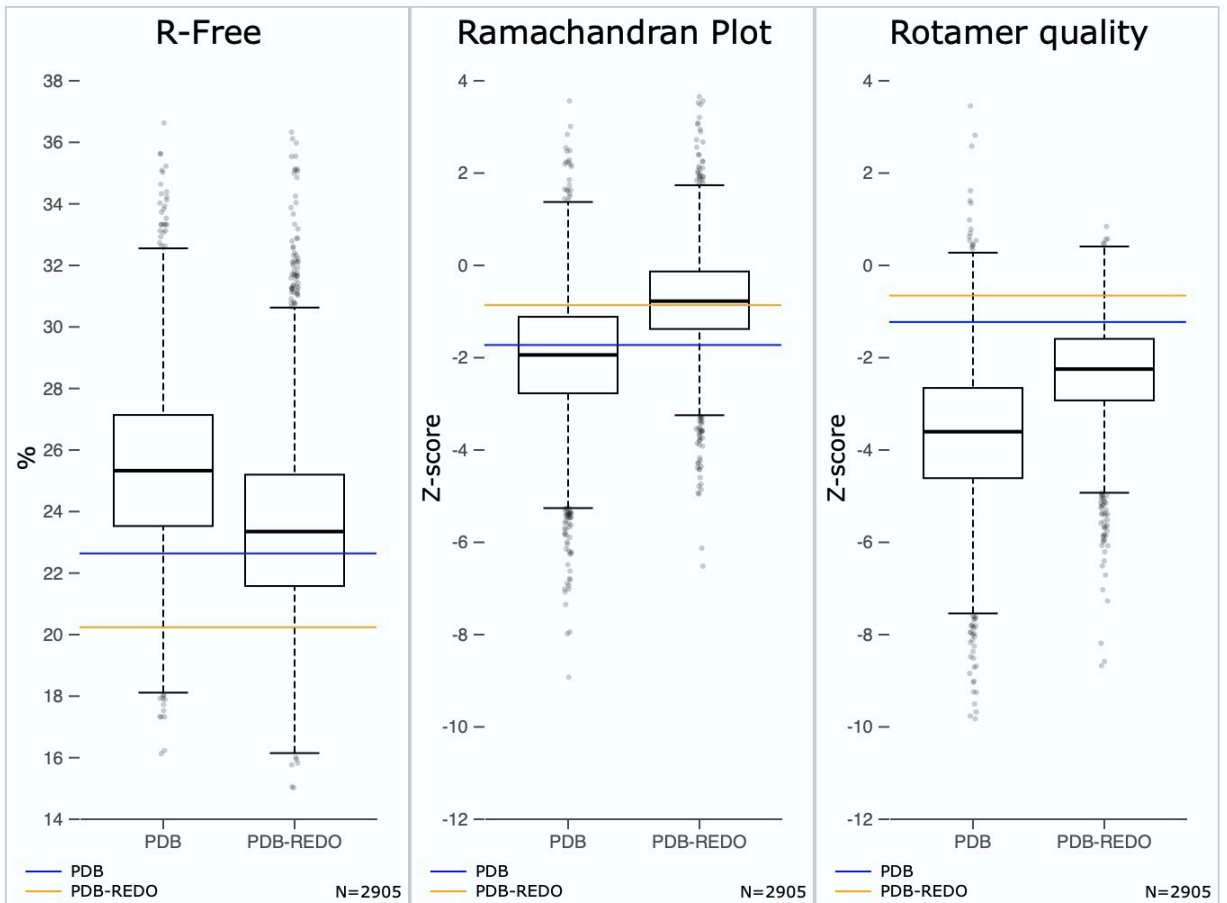
База данных PDB\_redo [7] содержит оптимизированные версии моделей, существующих в PDB. Для структуры 5ikq было проведено сравнение ее исходной и «улучшенной» версий.

Изменений по «улучшению» было проведено не так уж и много: улучшился R-factor, R-free, нормальность ротамеров, оптимизированы некоторые водородные связи. Кроме того было оптимизировано положение некоторых аминокислот относительно допустимых торсионных углов. На рисунке 12 приведены основные параметры, которые улучшились у оптимизированной модели.



**Description**                      **PDB** ●                      **PDB-REDO** ●

**Model quality compared to resolution neighbours**



<b>Significant model changes</b>	
<b>Description</b>	<b>Count</b>
<i>Rotamers changed</i>	21
<i>Side chains flipped</i>	16
<i>Waters removed</i>	32
<i>Peptides flipped</i>	0
<i>Chiralities fixed</i>	0
<i>Residues fitting density better</i>	93
<i>Residues fitting density worse</i>	2

**Рисунок 11.** Сравнение основных параметров качества в моделях PDB и PDB\_redo

## **Выводы**

Структура человеческой циклооксигеназы, связанной с селективным ингибитором, 5ikq имеет достаточно высокое разрешение и хорошо построенную модель. Все индикаторы качества имеют значения равные или выше среднего для структур с аналогичным разрешением из PDB. Большинство маргинальных остатков, определенных по формальным параметрам, таковыми не являются и могут быть объяснены «природными» факторами или техническими тонкостями кристаллографического эксперимента. Дополнительная оптимизация структуры с помощью PDB\_redo не требуется.

## **Список источников**

- [1] Orlando B.J., Malkowski M.G. Substrate-selective Inhibition of Cyclooxygenase-2 by Fenamic Acid Derivatives Is Dependent on Peroxide Tone \*. 2016.
- [2] <https://pfam.xfam.org/family/PF03098>
- [3] <https://wistar.org/baculovirus-expression-systems>
- [4] <https://www.rcsb.org/structure/3HS5>
- [5] <http://molprobit.biochem.duke.edu>
- [6] <https://swift.cmbi.umcn.nl/servers/html/>
- [7] <https://pdb-redo.eu/db/5ikq>