

МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
имени М.В.ЛОМОНОСОВА

ФАКУЛЬТЕТ БИОИНЖЕНЕРИИ И БИОИНФОРМАТИКИ

**Отчет по качеству расшифровки структуры белка
MsKGD_{Δ360} (домена α-кетоглутаратдекарбоксилазы) из
организма *Mycobacterium smegmatis* с помощью рентгено-
структурного анализа**

Работа студентки 4-го курса

Бойко Александры

Москва

2017 г.

ОГЛАВЛЕНИЕ

Аннотация	3
Введение	3
Результаты и обсуждение	5
Общая характеристика модели	5
Характеристика качества полной модели	6
Характеристика качества отдельных фрагментов модели	13
Сравнение модели 2YIC из PDB с данными из PDB-REDO	22
Заключение	24
Список литературы	24

Аннотация

В данной работе был проведен анализ качества структуры каталитического домена α -кетоглутаратдекарбоксилазы (KGD) из прокариотического организма *Mycobacterium smegmatis* (код структуры в PDB [2YIC](#)), полученной с помощью рентгено-структурного анализа и опубликованной в 2011 году (Wagner et al., 2011). Данный белок является формой белка KGD без его первых 360 аминокислот. В этом отчете рассмотрены различные параметры оценки соответствия представленной модели экспериментальным данным для всей структуры и для отдельных фрагментов белка, и сделаны выводы о качестве полученной модели, лежащей в базе данных PDB под кодом [2YIC](#).

Введение

Белок MsKGD_{Δ360} является частью α -кетоглутаратдекарбоксилазы (KGD) организма *Mycobacterium smegmatis*, который является близким родственником возбудителя туберкулеза – *Mycobacterium tuberculosis*. Этот фермент катализирует реакцию декарбоксилирования 2-оксоглутарата (ЕС-номер 4.1.1.71) с образованием CO₂ и сукцинил-семиальдегида (Рис. 1). В качестве кофактора белку необходим тиаминдифосфат (TDP), а также ионы Mg²⁺ и Ca²⁺.

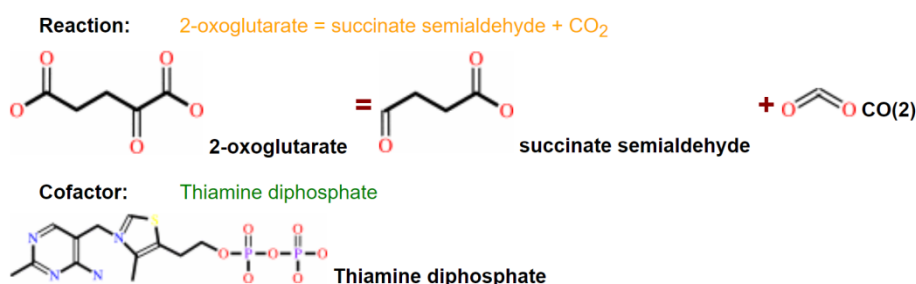


Рисунок 1. Реакция, катализируемая белком MsKGD_{Δ360}, и необходимый для реакции кофактор – ThDP).

В структуре под PDB-кодом 2YIC представлен белок MsKGD_{Δ360} длиной 868 аминокислотных остатков и приблизительной молекулярной массой 97 кДа. Полноразмерный KGD (Uniprot ID [AOR2B1](#)) состоит из 1227 аминокислот и доводит реакцию до образования сукцинил-КоА, но в рассматриваемом в данной работе варианте – MsKGD_{Δ360} – удалены первые 360 аминокислот N-конца, содержащего сукцинилтрансферазный домен. Таким образом, здесь рассматривается только каталитическая часть белка KGD, обладающая карбоксилазной и дегидрогеназной активностью, но не имеющая ацил-трансферазную активность. Белок KGD занимает важнейшее место в метаболизме *Mycobacterium smegmatis*, аналогичная роль отводится ему и в других бактериях (например, гомолог домена SucA в *Escherichia coli* (Frank et al., 2007) и в *Corinebacterium glutamicum* (Hoffelder et al., 2010). В статье, в которой впервые была

представлена структура MsKGD_{Δ360}, подробно изучены различные структурные функциональные аспекты KGD и его урезанных вариантов, исследована его активность и регуляция (Wagner et al., 2011). Предположительно, разностороннее исследование действия этого фермента и влияния на него в организме рода *Mycobacterium* поможет выявить способы борьбы с патологиями, вызываемыми этими организмами – например, туберкулезом (возбудитель *Mycobacterium tuberculosis*), лепрой (возбудитель *Mycobacterium leprae*) и другими болезнями, чьими возбудителями являются микобактерии avium-комплекса.

Результаты и обсуждение

Общая характеристика модели

Модель белка MsKGD_{Δ360} (PDB ID [2YIC](#)) была получена в 2011 году группой ученых из института Пастера под руководством Pedro M. Alzari. Дифракционные данные были получены на синхротроне ESRF (Grenoble, France) (Wagner et al., 2011). Фазовая проблема в разрешении кристаллической структуры данного белка была решена **методом молекулярного замещения** с помощью программы AMoRe (Trapani and Navaza, 2008), где в качестве гомологичной структуры была взята 2-оксоглутаратдегидрогеназа из *Escherichia coli*, EcSucA (PDB ID [2JGD](#)). Полученная модель была обработана последовательно программой REFMAC5 (Murshudov et al., 1997) в совокупности с оптимизацией вручную с использованием Coot (Emsley et al., 2010) и оптимизирована с помощью BUSTER (Bricogne et al., 2011). Также была проведена валидация модели с помощью сервера [MolProbity](#) (Davis et al., 2007). Общие характеристики эксперимента и полученных данных приведены в Таблице 1.

Таблица 1. Общая информация об экспериментальных данных для кристалла MsKGD_{Δ360}.

Параметр	Значение
Число измеренных рефлексов	263 738
Разрешение полной структуры	1.96 Å
Полнота набора структурных факторов	95.2 % ¹ 88.1 % ²
Диапазон разрешений структурных факторов	39.20-1.96 Å ¹ 49.53-1.96 Å ²
Параметры кристаллографической ячейки	a: 79.539Å b: 83.244Å c: 158.61Å α: 99.48° β: 99.06° γ: 101.25°
Кристаллографическая группа	P1
Некристаллографическая симметрия в асимметрической ячейке	Нет

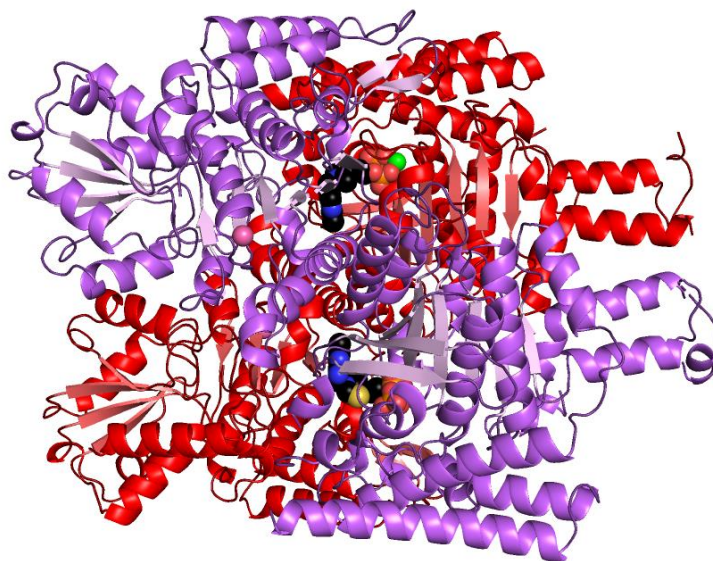
¹ – данные предоставлены авторами структуры

² – данные указаны по EDS в [полном отчете по валидации](#), приведенном на сайте [PDB](#).

В кристаллическую ячейку входит одна молекула полипептида. При этом биологически активным является гомодимер (Рис. 2А), а в файле PDB описано 2 соседних гомодимера (Рис.

2Б). На Рис. 2 представлен холофермент MsKGD_{Δ360} с необходимыми кофакторами (ThDP, а также ионы Mg²⁺ и Ca²⁺).

А



Б

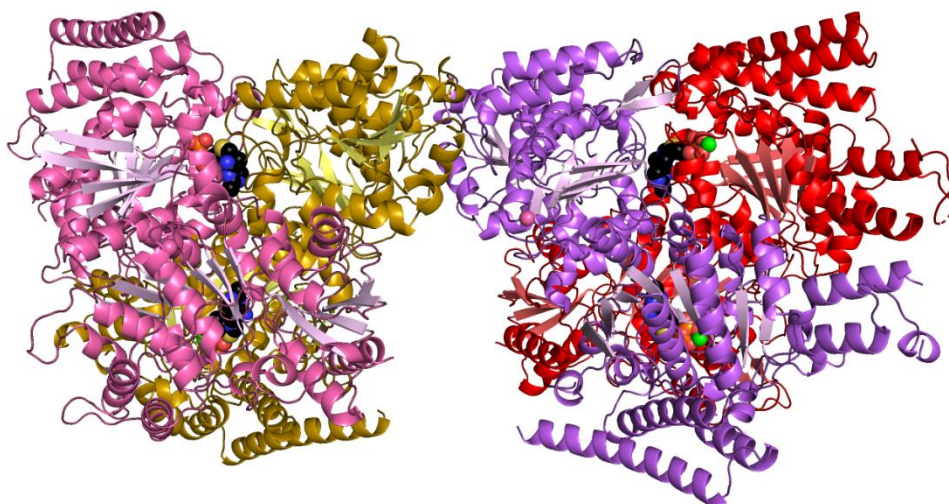


Рисунок 2. Структура белка MsKGD_{Δ360}. А – биологически активный гомодимер, Б – 2 гомодимера, содержащиеся в файле PDB 2YIC.

Характеристика качества полной модели

Рассмотрим основные индикаторы качества построенной модели в целом (Таблица 2). Одними из наиболее важных индикаторов являются R-фактор и R_{free} , отражающие качество соответствия построенной модели исходным экспериментальным данным. Значения R и R_{free} были получены авторами структуры (Wagner et al., 2011) с помощью программы [DCC](#) (Yang et al., 2016). Кроме этого, в Таблице 2 приведен общий температурный фактор B (Wilson B-factor), характеризующий степень упорядоченности кристалла и рассчитываемый по данным

дифракционного эксперимента. Для хорошей модели он не должен сильно отличаться от среднего B-фактора, вычисляемого по модели для каждого атома. Помимо этих параметров, для оценки соответствия между моделью и экспериментальными данными (наблюдаемыми структурными факторами) был также использован коэффициент корреляции между наблюдаемыми (F_o) и рассчитанными (F_c) структурными факторами.

Таблица 2. Статистика по экспериментальным данным и их оптимизации для структуры белка MsKGD_{Δ360}. Данные получены из [полного отчета по валидации](#), приведенного на сайте [PDBe](#).

Показатель	Значение
Использованная программа оптимизации модели	BUSTER (Bricogne et al., 2011)
Общее число атомов модели	26 363
Общее количество рефлексов	260 402
Количество рефлексов в тестовой выборке	13065 (5.28%)
R-factor	0.1896 (18.96%)
R_{free}	0.2108 (21.08%)
Wilson B factor	21.8 Å ²
Average B, all atoms	28.0 Å ²
Коэффициент корреляции (Fo,Fc)	0.93

Исходя из приведенных в Таблице 2 значений нескольких показателей, можно сделать вывод, что R-factor и R_{free}, во-первых, практически не различаются между собой, а во-вторых, имеют значение ниже 25%, что говорит о хорошем соответствии модели белка MsKGD_{Δ360} сырым экспериментальным данным. Об этом же свидетельствует и коэффициент корреляции между Fo и Fc (0.93 весьма близко к 1, поэтому можно говорить о хорошей линейной корреляции). При этом надо заметить, что уменьшилось количество использованных для оптимизации модели рефлексов: по экспериментальным данным их было 263 738 (Таблица 1), а в Таблице 2 приведено 260 402 (Таблица 2) – по-видимому, часть ненадежных рефлексов была отсеяна перед оптимизацией.

Была предпринята попытка получить более подробные данные с помощью сервера EDS, однако возникала проблема, показанная на Рис. 3.

Sorry...

There is no map available for this entry (2yic), because our automatic script failed to produce an electron density map with an **R-value (0.258)** within 5 percentage points of the published one (0.188). If you are the author of this entry and wish to help us remedy this situation, please [contact us](#).

Рисунок 3. Скриншот страницы EDS сервера, отражающий невозможность использования этого сервера для модели MsKGD_{Δ360}.

То есть, по Рис. 3, из данных об электронной плотности получался R-фактор = 25.8%, тогда как для данной структуры было опубликовано значение 18.96% (Таблица 2). Из-за разницы этих значений, превышающей 5%, сервер отказывается анализировать данную структуру. Тем не менее, информацию, которую предоставляет данный сервер, можно получить из других источников, которые будут использованы далее, за исключением пространственного R-фактора (RSR) для всех остатков исследуемой структуры.

Качество модели относительно других моделей банка PDB можно оценить с помощью Percentile plot, представленного на Рис. 4.

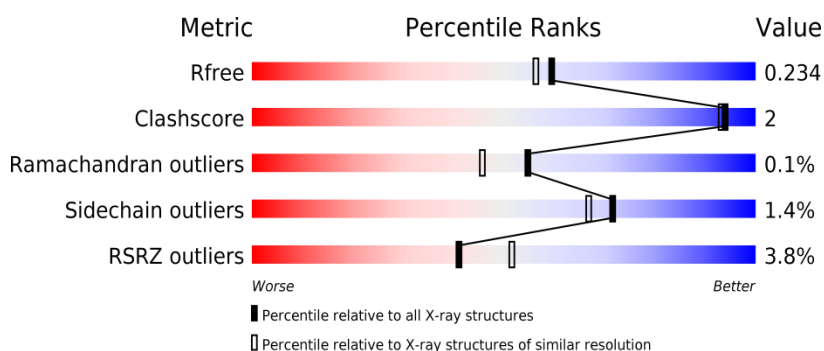


Рисунок 4. Percentile Plot для модели PDB 2YIC белка MsKGD_{Δ360} отражающий параметры качества данной модели относительно других моделей PDB.

Итак, показатели для выбранной структуры в основном среднего качества, но зато она имеет довольно высокий clashscore – настолько, что может возникнуть подозрение о подгонке модели под этот параметр. Кроме этого, из Рис. 4 можно получить данные по маргинальным остаткам по параметрам RSRZ и конформации боковой цепи – их количество 3.8% и 1.4%, соответственно. Это довольно хорошие показатели, лежащие до порога (10%).

Из [полного отчета по валидации](#) была получена информация о соответствии модели электронной плотности по геометрическим критериям (Рис. 5).

Mol	Chain	Length	Quality of chain
1	A	868	
1	B	868	
1	C	868	
1	D	868	

Рисунок 5. Таблица по качеству модели по геометрическим критериям, визуализированному на полипептидных цепях биологической единицы 2Y1C. Зеленым цветом изображены участки цепей, удовлетворяющие всем геометрическим критериям, желтым – не удовлетворяющие одному критерию (имеющие хотя бы один outlier), красным – трем и более. Серые сегменты отражают те участки цепей, которые не были смоделированы.

Для характеристики модели также необходимо отдельно проанализировать карту Рамачандрана. Ее можно получить с помощью нескольких серверов: [MolProbity](#) (Davis et al., 2007), [PROCHECK](#) (данные приведены в PDBsum) и [EDS](#) (отдельно от других параметров). Полученные карты представлены на Рис. 5.

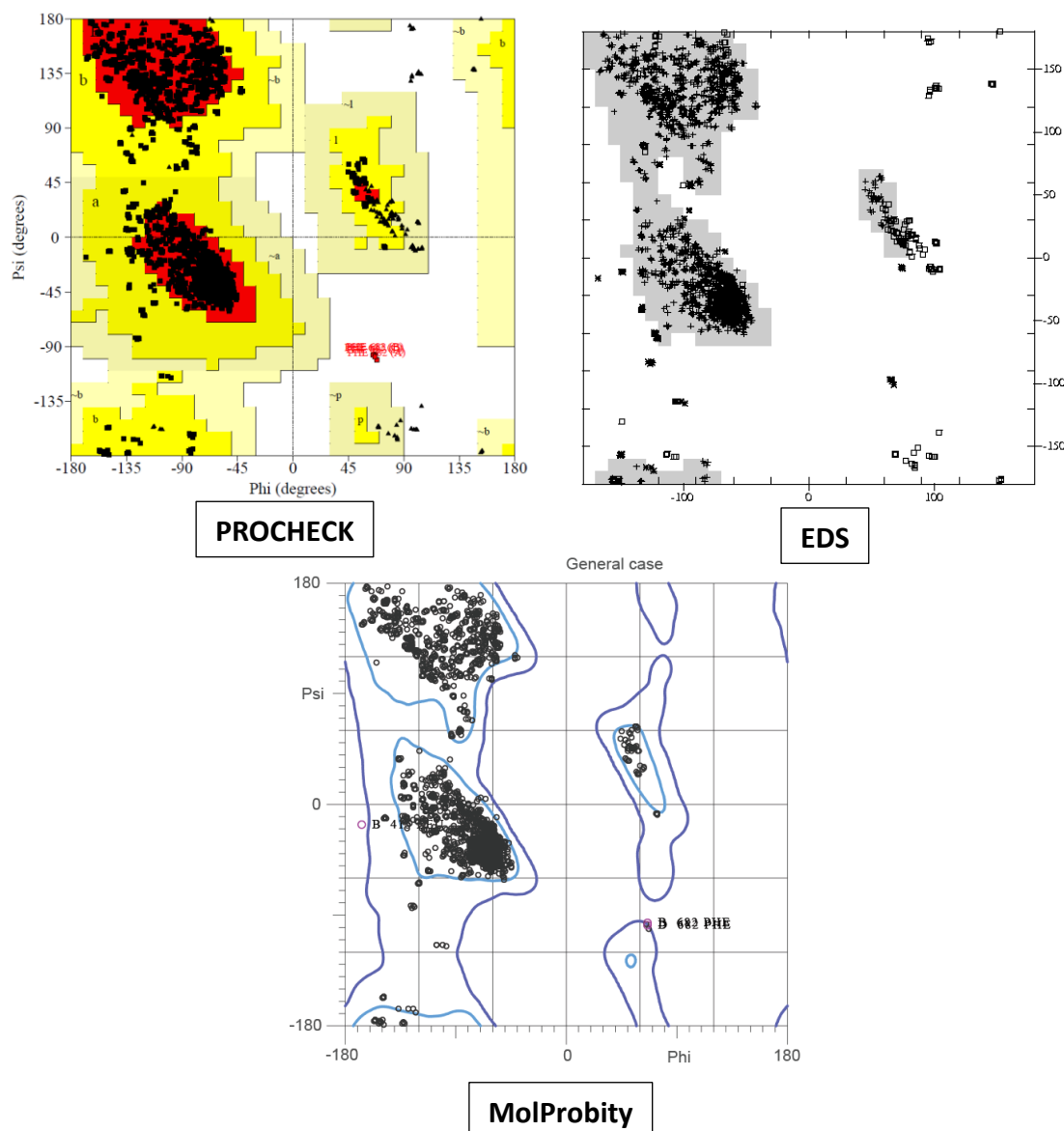


Рисунок 5. Карты Рамачандрана для всех полипептидных цепей структуры 2YIC. Рядом с каждой картой подписан сервис, с помощью которого она была получена.

Полученные карты Рамачандрана очень похожи между собой. Более подробная статистика по ним приведена в Таблице 3.

Таблица 3. Статистические параметры карт Рамачандрана, полученных с помощью сервисов MolProbity, PROCHECK и EDS.

	PROCHECK	EDS	MolProbity
Остатки в предпочитаемой области	2580/2814 (91.7%)	2911/2959 (99.5%)	3136/3200 (98%)
Остатки в разрешенной области	230/2814 (8.2%)	48/2959 (0.5%)	61/3200 (1.9%)
Остатки в запрещенной области	4/2814 (0.1%)		3/3200 (0.1%)
Общее число остатков, использованных для построения карты	2814*	2959**	3200
Общее число остатков модели	3239		

*Из общего числа исключены концевые 35 остатков, а также глицины и пролины

**Исключены глицины (240) и концевые остатки (40)

Из Таблицы 3 видно, что три сервиса используют немного разное множество остатков для построения карты Рамачандрана, однако, в целом, результаты хорошо сходятся между собой. По оценке всех трех программ более 90% остатков попадают в предпочитаемую область, что говорит о высоком качестве модели и малом количестве маргинальных по конформации остова остатков.

Отдельно в [MolProbity](#) были построены карты Рамачандрана для остатков глицина, пролина и изолейцина с валином, они приведены на Рис. 6.

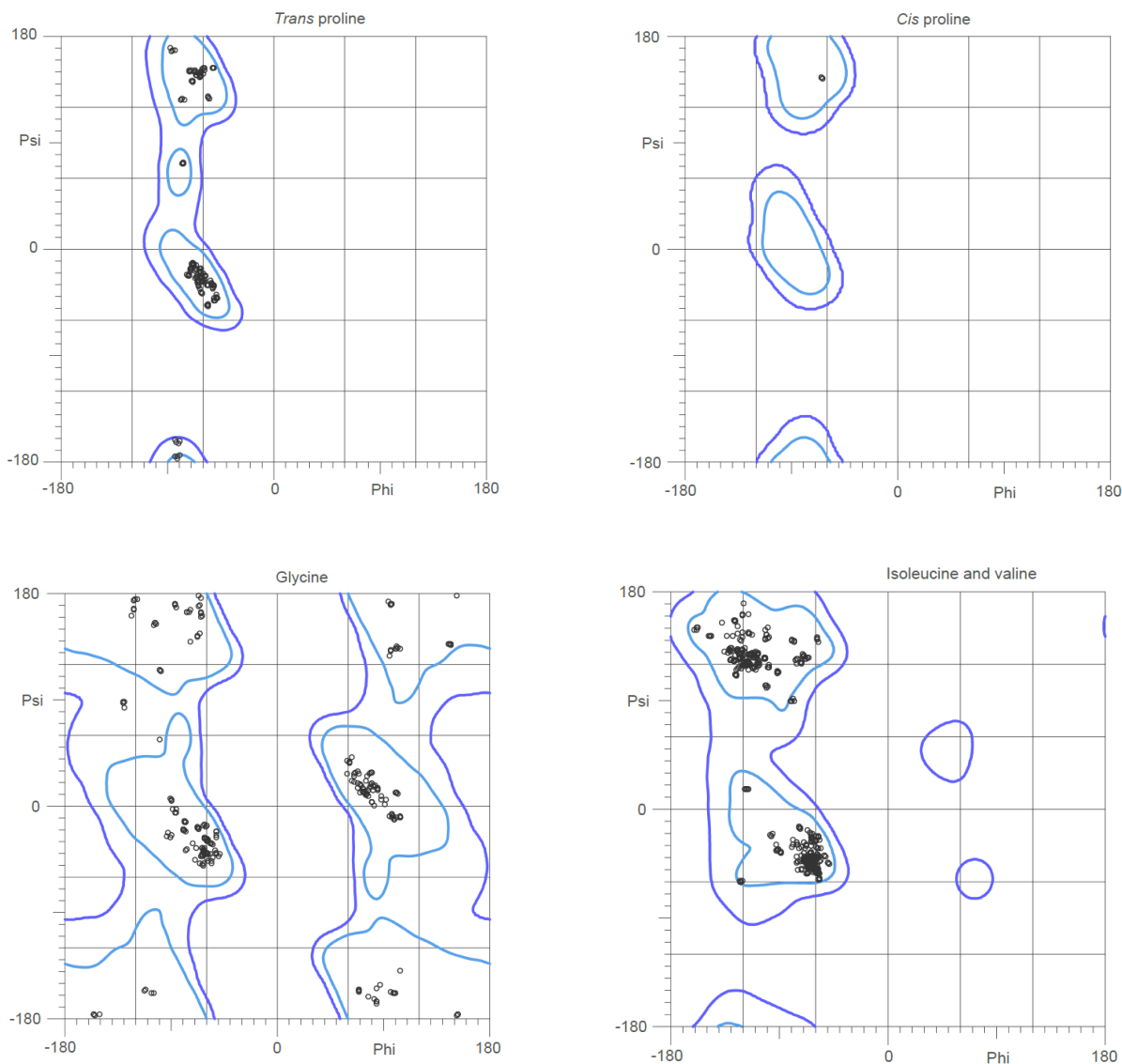


Рисунок 6. Карты Рамачандрана для транс- и цис-пролинов (верхний ряд), а также глицинов и изолейцинов с валинами (нижний ряд) структуры 2YIC. Изображения получены с помощью [MolProbity](#).

По данным картам (Рис. 6) все остатки, рассмотренные отдельно программой, попали в разрешенные области.

На Рис. 7 представлены результаты [MolProbity](#) с несколькими показателями качества модели.

All-Atom Contacts	Clashscore, all atoms:	1.6	100 th percentile* (N=721, 1.96Å ± 0.25Å)
	Clashscore is the number of serious steric overlaps (> 0.4 Å) per 1000 atoms.		
Protein Geometry	Poor rotamers	32	1.25%
	Favored rotamers	2434	94.75%
	Ramachandran outliers	3	0.09%
	Ramachandran favored	3133	98.00%
	MolProbity score [^]	0.98	100 th percentile* (N=12355, 1.96Å ± 0.25Å)
	Cβ deviations >0.25Å	1	0.03%
	Bad bonds:	25 / 25646	0.10%
Bad angles:	58 / 34808	0.17%	
Peptide Omegas	Cis Prolines:	4 / 141	2.84%
			Expected: ≤1 per chain, or ≤5%

Рисунок 7. Таблица с показателями качества модели 2YIC. Результаты получены с помощью [MolProbity](#).

По результатам [MolProbity](#) (Рис. 7), для модели было определено меньшее количество удовлетворительных ротамеров, чем должно быть у модели высокого качества (94.75% из желательных 98%). Кроме того, превышено количество допустимых плохих углов и длин связей (строки отмечены красным на Рис. 7). По-видимому, именно этим двум параметрам было уделено особое внимание в ходе оптимизации модели, поскольку в итоговом отчете о валидации на сайте [EBI](#) для структуры 2YIC указано полное отсутствие маргиналов по длинам связей и углам (Рис. 8).

Metric	Description
Bond angles in protein, DNA, RNA molecules	0 outlier(s) of 34650 (%)
Bond lengths in protein, DNA, RNA molecules	0 outlier(s) of 25544 (%)
Electron density fit in protein, DNA, RNA molecules	124 outlier(s) of 3242 (%)
Ramachandran outliers in protein molecules	3 outlier(s) of 3200 (%)
Sidechain rotamer outliers in protein molecules	37 outlier(s) of 2569 (%)

Рисунок 8. Таблица с показателями валидации, приведенными на сайте [EBI](#) для структуры 2YIC.

Помимо этого, на Рис. 8 приведено количество маргинальных остатков по конформации боковой цепи – 37 outliers of 2569.

Характеристика качества отдельных фрагментов модели

В данном разделе будут рассмотрены маргинальные остатки модели 2YIC белка MsKGD_{Δ360}, которые приведены в Таблице 4.

Три маргинальных остатка были в явном виде получены из карты Рамачандрана, построенной сервисом MolProbity (Рис.5), тогда как другие маргиналы по этому параметру были показаны на картах Рамачандрана для каждого типа аминокислотного остатка сервисом PROCHECK (Рис.9).

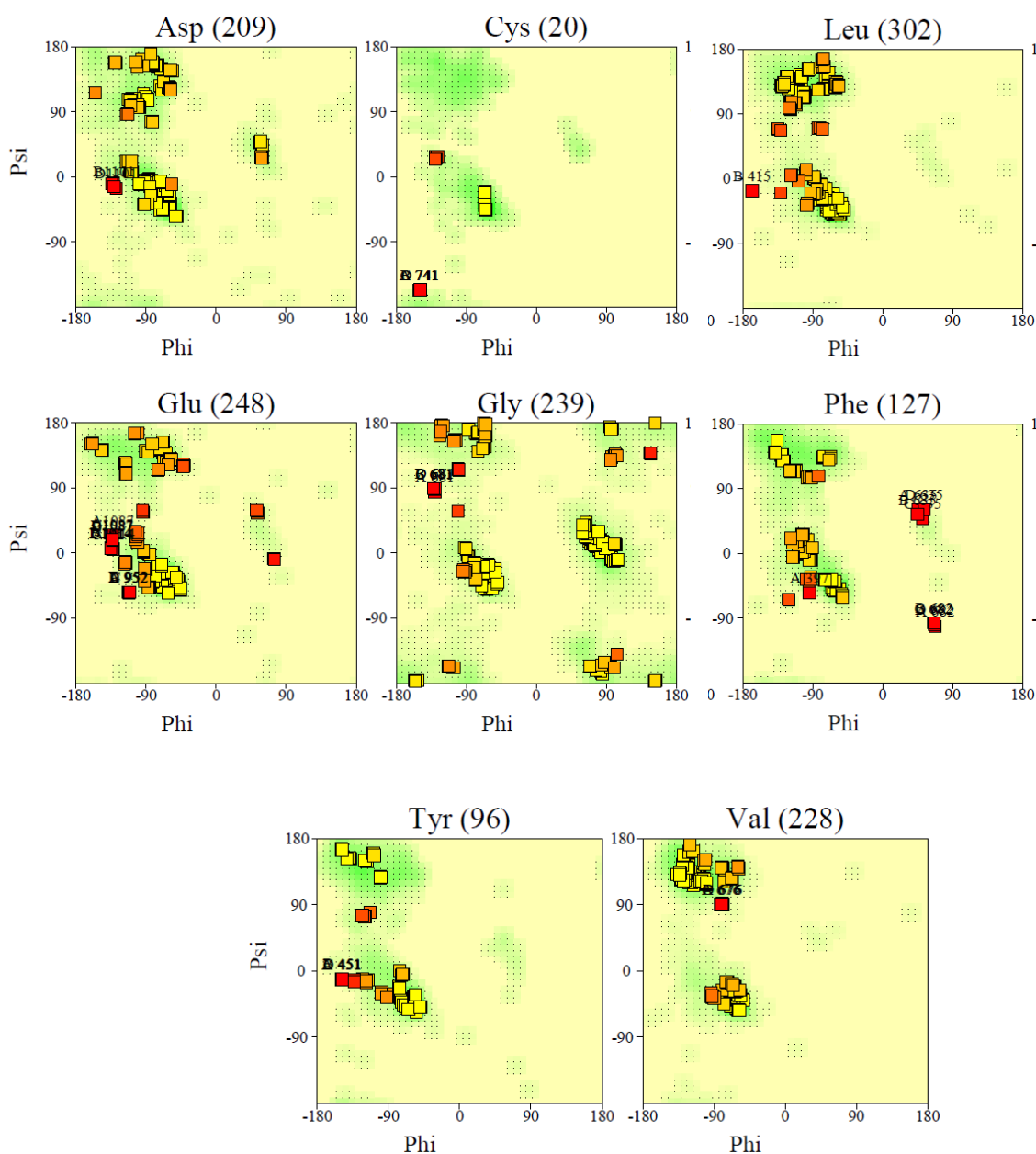


Рисунок 9. Карты Рамачандрана для каждого типа аминокислотных остатков из программы PROCHECK. Остатки структуры 2YIC, попавшие в запрещенные области, подписаны.

Кроме карт Рамачандрана для отдельных остатков, из программы PROCHECK были добыты маргинальные остатки по торсионным углам главной и боковых цепей (Рис. 10).

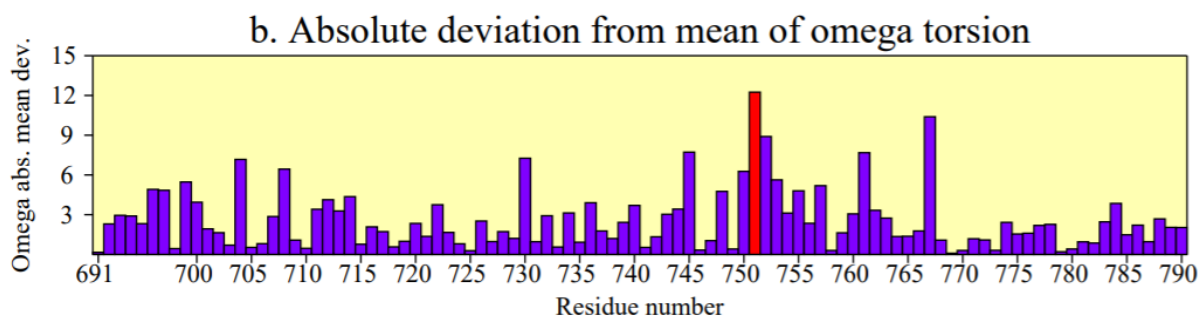
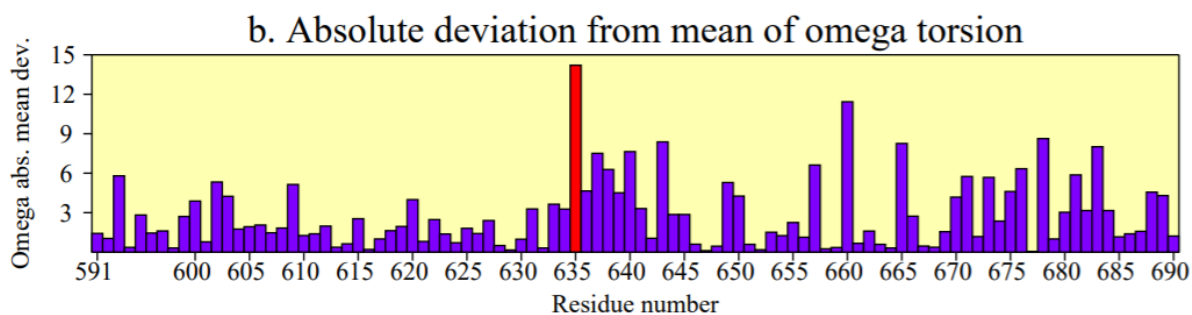
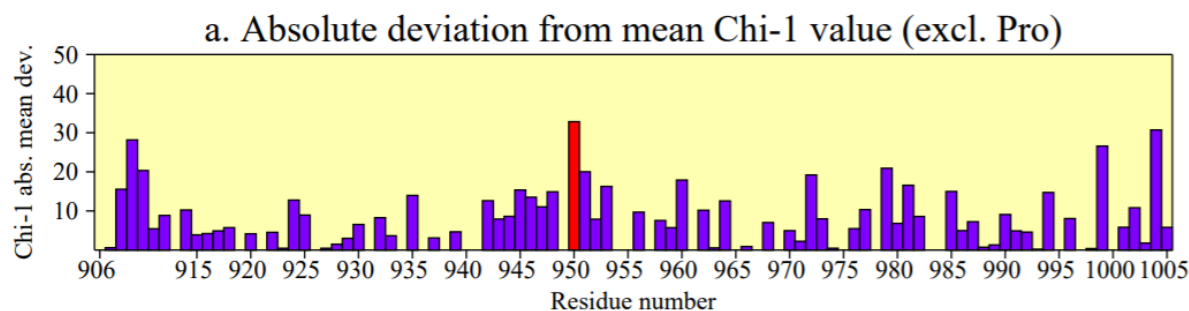
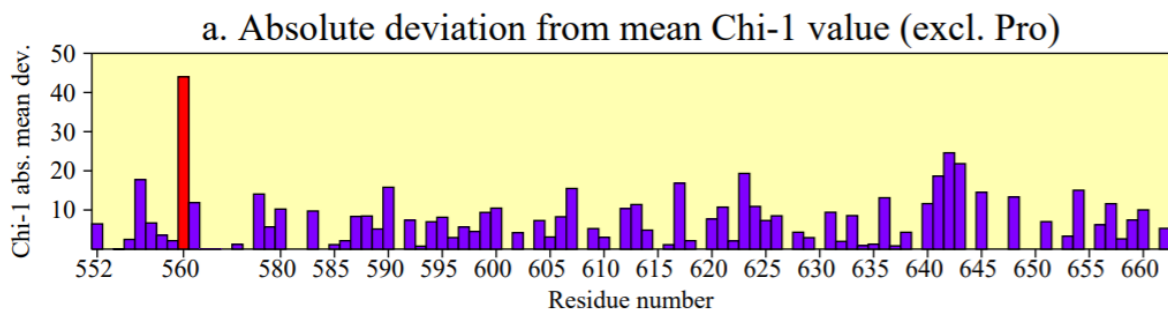
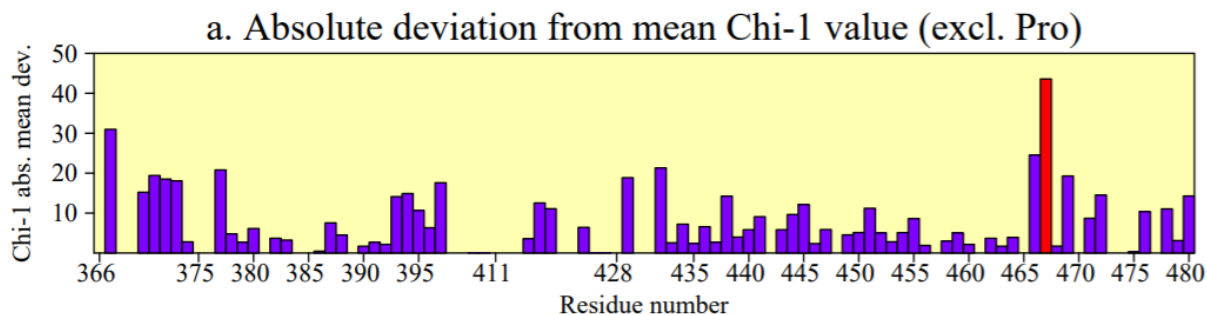


Рисунок 10. Графики с отклонениями значений углов χ_1 (верхние три) и ω (нижние три) для каждого остатка структуры MsKGD_{Δ360}. Красным отмечены остатки, отличающиеся от идеальных значений больше, чем на 2 Z. (Изображены не все графики.)

Наконец, из той же программы PROCHECK были получены данные об остатках с нестандартными длинами связей в остовах (Рис. 11)

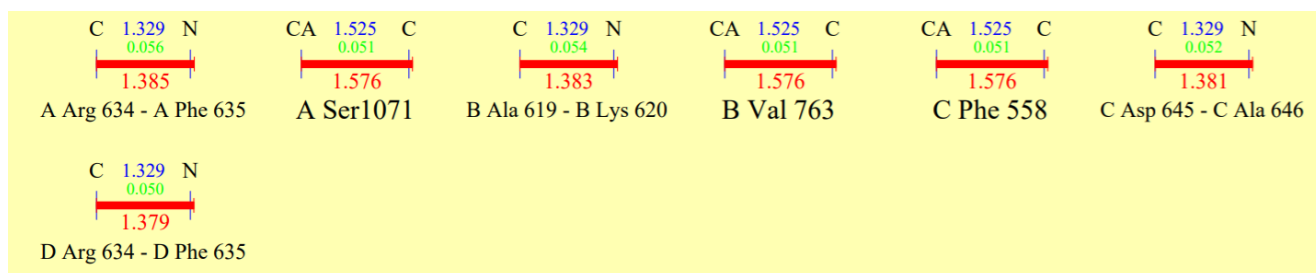


Рисунок 11. Остатки с отличающимися от идеальных значениями длин связей остова: синее число – «идеальное» значение, зеленое – разница, красное – реальное значение по структуре MskGD_{Δ360}.

Таким образом, исходя из представленной выше информации, был составлен список некоторых маргинальных остатков структуры MskGD_{Δ360}. Он приведен в Таблице 4.

Таблица 4. Список нескольких маргинальных остатков структуры белка MskGD_{Δ360}. Цветом выделены остатки, которые будут подробнее рассмотрены далее.

№	Цепь	Название остатка	Критерий
1	B	415 LEU	Маргинал по карте Рамачандрана (MolProbity)
2	B, D	682 PHE	Маргинал по карте Рамачандрана (MolProbity)
3	...	451 TYR	Маргинал по карте Рамачандрана (PROCHECK)
4	...	741 CYS	Маргинал по карте Рамачандрана (PROCHECK)
5	...	952 GLU	Маргинал по карте Рамачандрана (PROCHECK)
6	...	676 VAL	Маргинал по карте Рамачандрана (PROCHECK)
7	...	467 VAL	Торсионный угол χ_1 (ротамеры)
8	B	560 GLU	Торсионный угол χ_1 (ротамеры)
9	A	635 PHE	Торсионный угол ω
10	A	634 ARG – 635 PHE	Длина связи C – N
11	A	1071 SER	Длина связи C α - C
12	B	619 ALA – 620 LYS	Длина связи C – N

Рассмотрим подробнее несколько маргинальных остатков из Таблицы 4.

1. LEU 415

Остаток лейцина Leu415, являющийся маргиналом по карте Рамачандрана, построенной программой MolProbity (Рис. 5), показан на Рис. 12 с электронной плотностью по уровню подрезки 2σ .

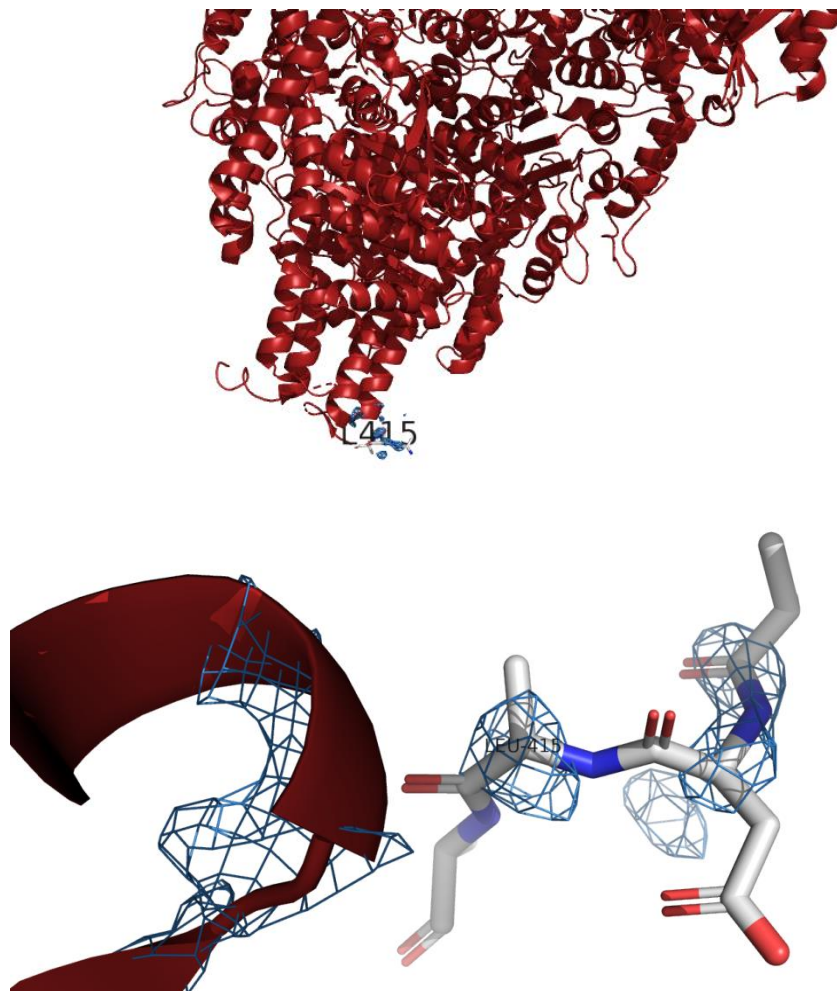


Рисунок 12. Остаток LEU415, цепь В с электронной плотностью на уровне подрезки 2σ . Вся структура $MskGD_{\Delta 360}$ изображена *cartoons* темно-красного цвета. Сверху: общее положение остатка в структуре. Снизу: отдельно остаток LEU415 и его ближайшие соседи (изображены палочками и покрашены по элементам).

Во-первых, из верхнего изображения на Рис. 12 видно, что остаток LEU415 находится на конце цепи В, выходящем на периферию белка. Во-вторых, в электронную плотность на уровне подрезки 2σ данный остаток вписан с плохим качеством. Таким образом, можно сказать, что в этом случае остаток действительно является маргиналом из-за плохого качества расшифровки.

2. GLU 952

Остаток глутамата GLU952 всех цепей также был определен сервисами PROCHECK и MolProbity как попадающие на границу запрещенной и разрешенной области карты Рамачандрана (Рис. 9). Он с электронной плотностью изображен на Рис. 13.

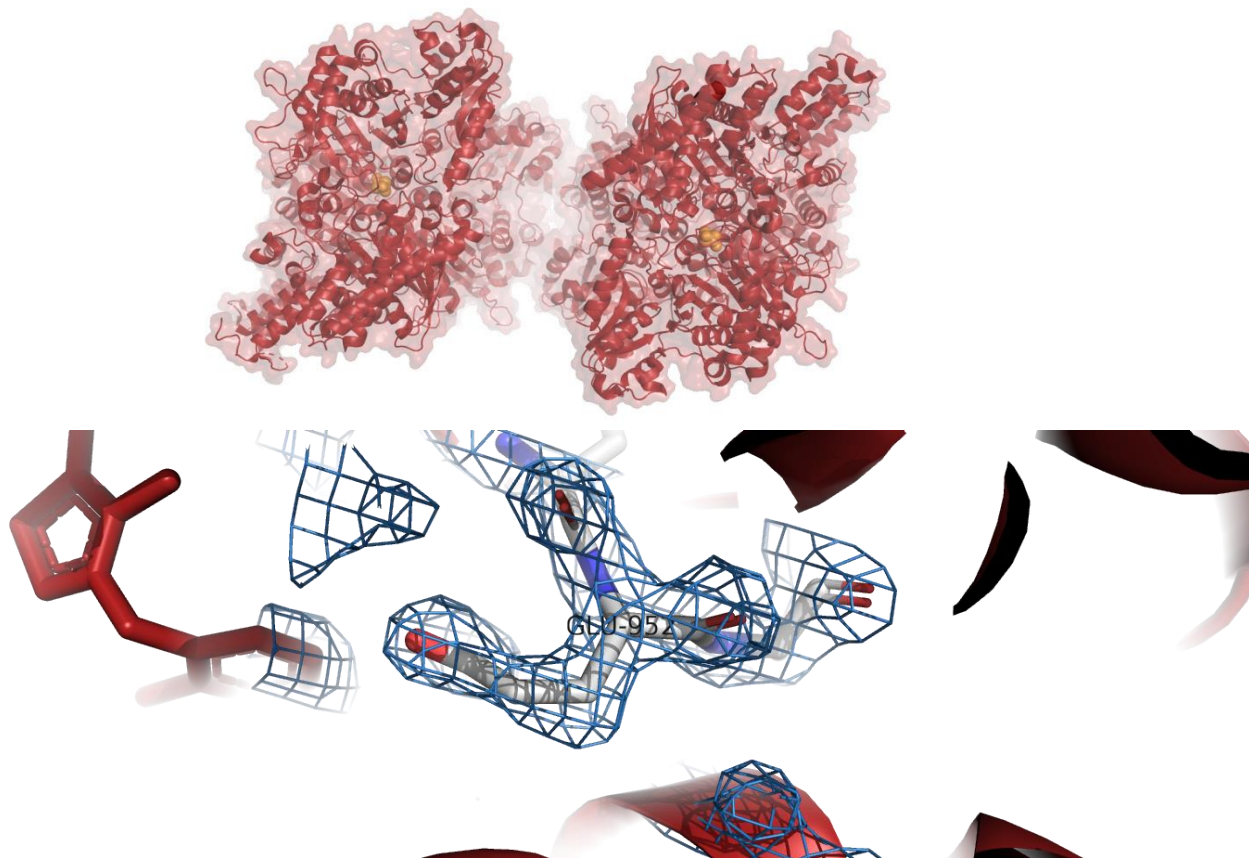


Рисунок 13. Остаток GLU952. Вся структура $MskGD_{\Delta 360}$ изображена *cartoons* темно-красного цвета. Сверху: общее положение остатка в структуре. Остаток изображен в цепях B и D оранжевыми шарами. Снизу: отдельно остаток GLU952 цепи B и его ближайшие соседи (изображены палочками и покрашены по элементам), также отражена электронная плотность на уровне подрезки 2σ .

Исходя из Рис. 13, смоделированный остаток GLU952 хорошо вписан в электронную плотность на выбранном уровне подрезки. В связи с этим сложно сказать, если ли проблемы с качеством определения положений атомов данного остатка. Вероятно, такое значение торсионных углов остова для GLU952 является особенностью белка в данном месте.

3. VAL 467

Следующий остаток – VAL467 – является маргинальным по значению угла χ_1 или другого(их) угла(ов) боковой цепи, которое, вероятно, не соответствует типичному значению для ротамеров данного остатка (Таблица 4). Согласно Supplementary Materials из (Kleywegt and

Jones, 1998) с данными по углам ротамеров, наиболее предпочтительным углом χ_1 для валинов является 180° , менее комфортными (но возможными) являются углы 60° и 300° .

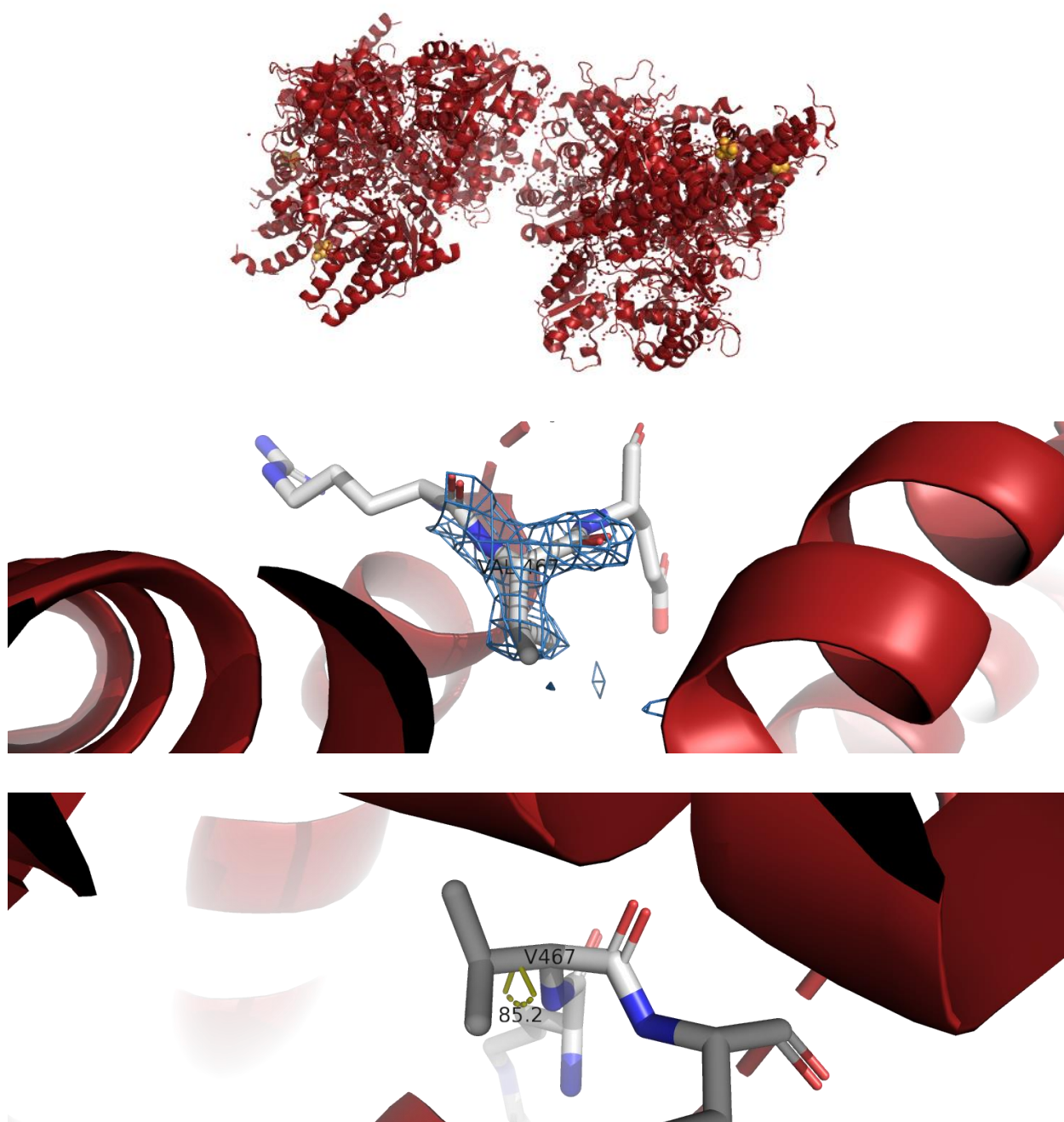


Рисунок 14. Остаток VAL467. Вся структура MsKGD $_{\Delta 360}$ изображена cartoons темно-красного цвета. Сверху: общее положение остатка в структуре. Остаток изображен во всех цепях оранжевыми шарами. В середине: отдельно остаток VAL467 цепи В и его ближайшие соседи (изображены палочками и покрашены по элементам), также отражена электронная плотность на уровне подрезки 2σ . Снизу: остаток VAL467 цепи В и его ближайшие соседи (изображены палочками и покрашены по элементам), обозначен торсионный угол χ_1 исследуемого остатка.

Как следует из Рис. 14, остаток VAL467 хорошо вписывается в электронную плотность на уровне подрезки 2σ . При этом, так как остаток подозревается на маргинальность по торсионному углу

χ_1 , он был измерен в PyMol (как угол между атомами N-Ca-Cb-Cg1), полученное значение – 85.2°, что отличается от предпочитаемого значения 0°(180°). Предположить происхождение такого отклонения трудно, возможно это зависит от окружения (Рис. 15)

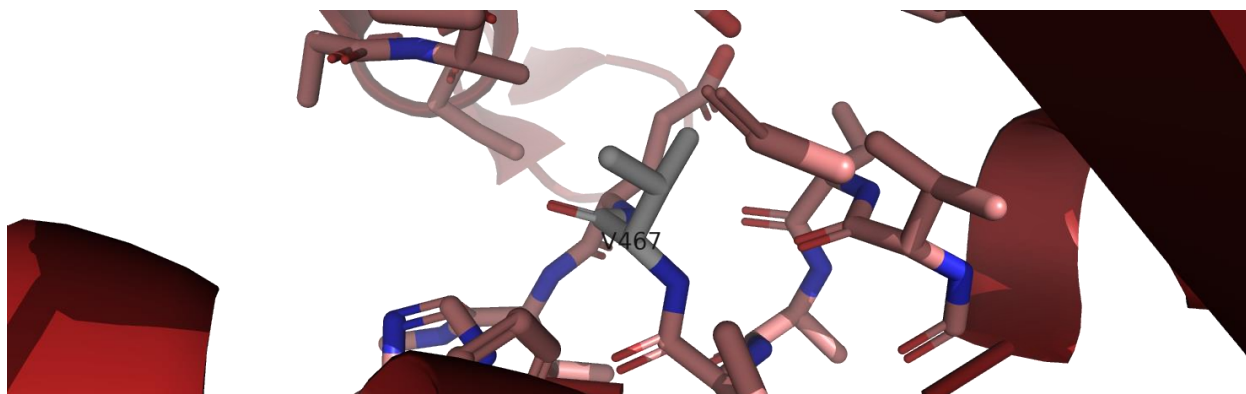
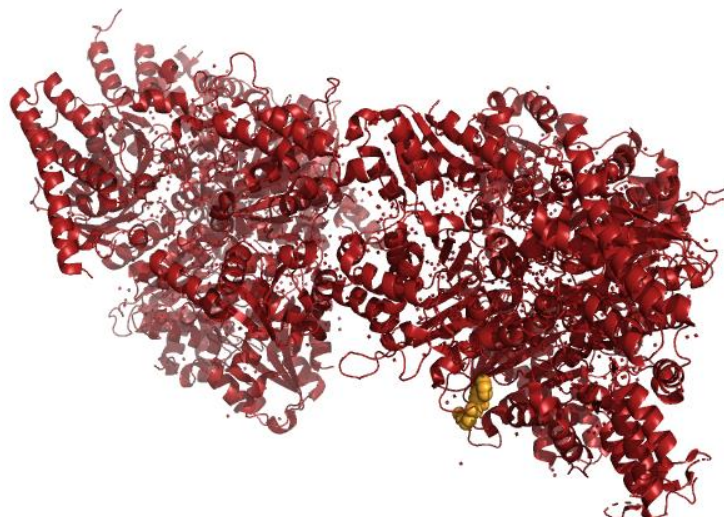


Рисунок 15. Остаток VAL467 цепи В (изображен палочками и покрашен по элементам) и его окружение в пределах 7 Å (изображено палочками и покрашено розовым). Вся структура MsKGD_{Δ360} изображена cartoons темно-красного цвета.

Окружение остатка VAL467 достаточно гидрофобно – теоретически, можно предположить, что ради создания гидрофобного кармана он и принял такую конформацию. Однако также можно предположить погрешность эксперимента.

4. ARG634 – PHE635

Остаток фенилаланина PHE635 маргинален по величине угла ω , связь C(O)-N(H) (Таблица 4). Этот угол может принимать значения 0° (цис-форма) и 180° (транс-форма). Кроме этого, эта же связь маргинальна по длине. Ситуация рассмотрена подробнее на Рис. 16.



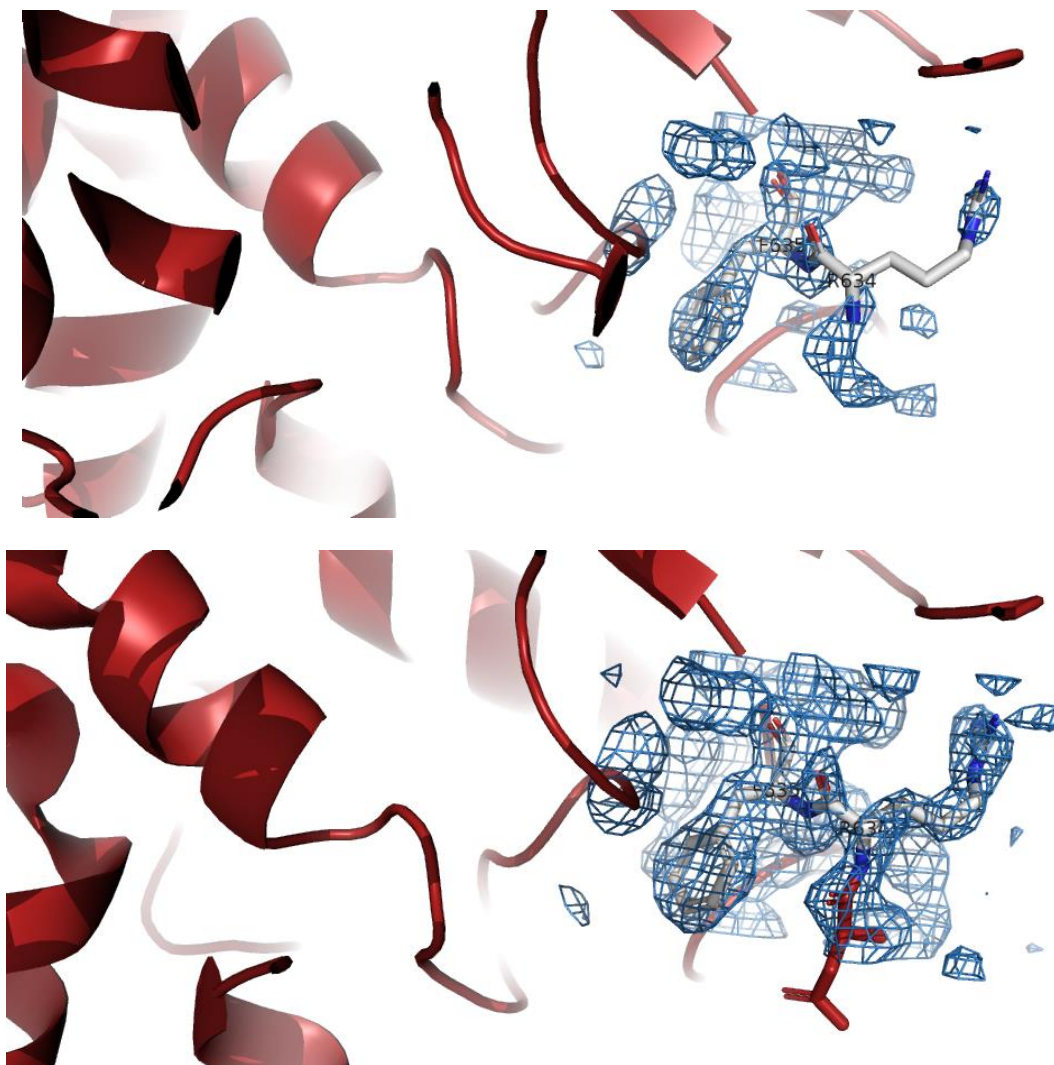


Рисунок 16. Остатки ARG634 и PHE635 цепи A. Вся структура MsKGD_{Δ360} изображена cartoons темно-красного цвета. Сверху: общее положение остатка в структуре. Остаток изображен оранжевыми шарами. В середине: остатки PHE635 и ARG634 цепи A (изображены палочками и покрашены по элементам), также отражена электронная плотность на уровне подрезки 2σ. Снизу: Аналогично, добавлен остаток ASN633 цепи A и электронная плотность на уровне подрезки 1σ.

Как и в большинстве предыдущих случаев, остаток располагается на поверхности белка, что, вероятно, может приводить к каким-то аномалиям в конформациях. Из анализа электронной плотности (уровень подрезки 2σ), соответствующей этим двум остаткам, можно заметить, что PHE635 хорошо вписан в плотность, тогда как соседний аргинин на уровне подрезки 2σ совсем не попадает в электронную плотность (за исключением гуанидиновой группы), и только на уровне 1σ вписывается в нее (Рис. 16). При этом по связи C(O)-N(H) получается действительно цис-форма. Можно сделать вывод о неоднозначном расположении остатка ARG634 в синтезированной электронной плотности. Обращая внимание на среднее изображение Рис. 16, можно заметить другой возможный вариант размещения данного аргинина. Эти остатки также отмечены в протоколе [WhatCheck](#) – однако Z-скор для конформации остова признан нормальным.

5. SER1071

Данный остаток является маргиналом по длине связи $C\alpha - C$ (Рис.11, Таблица 4). Он изображен на Рис. 17.

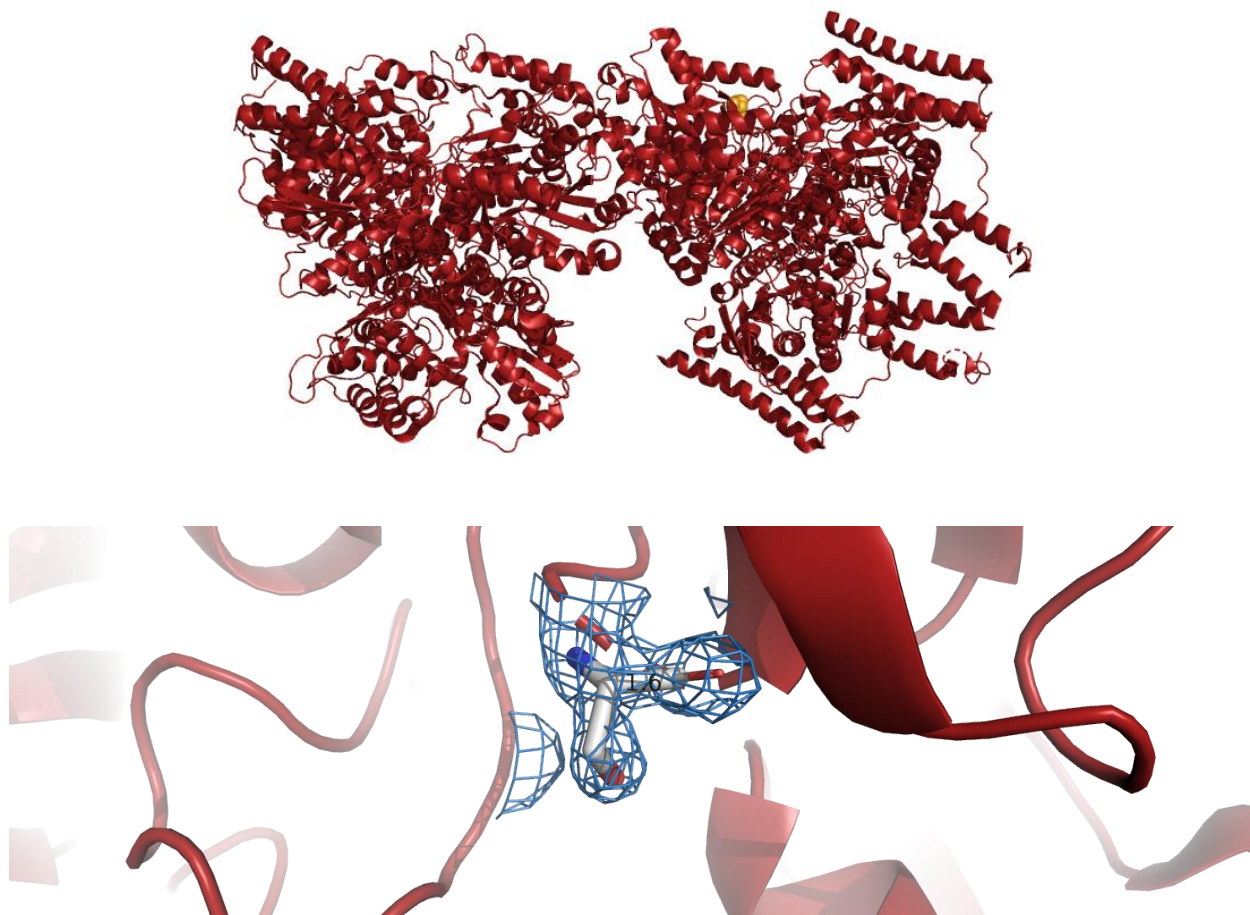


Рисунок 17. Остаток SER1071 цепи A. Вся структура $MsKGD_{\Delta 360}$ изображена *cartoons* темно-красного цвета. Сверху: общее положение остатка в структуре. Остаток изображен оранжевыми шарами. Внизу: остаток SER1071 цепи A (изображен палочками и покрашен по элементам), также отражена электронная плотность на уровне подрезки 1.5σ .

По Рис. 17 SER1071 неплохо вписан в полученную электронную плотность (уровень подрезки 1.5σ). Длина данной связи в среднем должна быть около 1.51 \AA , но в рассматриваемой случае ее длина ближе к 1.6 \AA . В целом, не очень понятно, откуда мог взяться такой маргинальный показатель, и, по-видимому, это просто особенность структуры. Этот остаток упоминается в отчете программы [WhatCheck](#) как возможно входящий в состав странной петли, а возможно, являющийся ошибкой идентификации.

Сравнение модели 2YIC из PDB с данными из PDB-REDO

На сайте [PDB-REDO](#) хранятся оптимизированные модели структур из PDB. В данной работе было проведено сравнение параметров модели из PDB с параметрами из PDB-REDO. Они выделены цветом в Таблице 5.

Таблица 5. Параметры качества неоптимальной модели MsKGD_{Δ360} из PDB и модели из PDB-REDO. Цветом обозначены значимые несоответствия: зеленым – в положительную сторону (то есть модель PDB-REDO по этому параметру лучше модели PDB), красные – в отрицательную.

Validation metrics from PDB-REDO		
	PDB	PDB-REDO
Crystallographic refinement		
<i>R</i>	0.2065	0.2039
<i>R-free</i>	0.2514	0.2297
<i>Bond length RMS Z-score</i>	0.628	0.676
<i>Bond angle RMS Z-score</i>	0.850	0.824
Model quality (raw scores percentiles)		
<i>Ramachandran plot appearance</i>	47	55
<i>Rotamer normality</i>	32	47
<i>Coarse packing</i>	N/A	N/A
<i>Fine packing</i>	15	30
<i>Bump severity</i>	93	81
<i>Hydrogen bond satisfaction</i>	76	72

Как можно заметить из Таблицы 5, PDB-REDO понизил *R-free*, увеличил количество немаргинальных остатков на карте Рамачандрана (с 47% до 55%) и значительно оптимизировал положения ротамеров. Тем не менее, автоматически программа снизила количество оптимизированных водородных связей с 76 для модели, лежащей в PDB, до 72 (Таблица 5).

Далее было проделано наложение структур исходной модели и оптимизированной с помощью REDO (Рис. 18). В целом две модели хорошо соответствуют друг другу (Рис. 18, сверху). Тем не менее, можно обнаружить фрагменты, в которых модели не так хорошо друг другу сопоставляются.

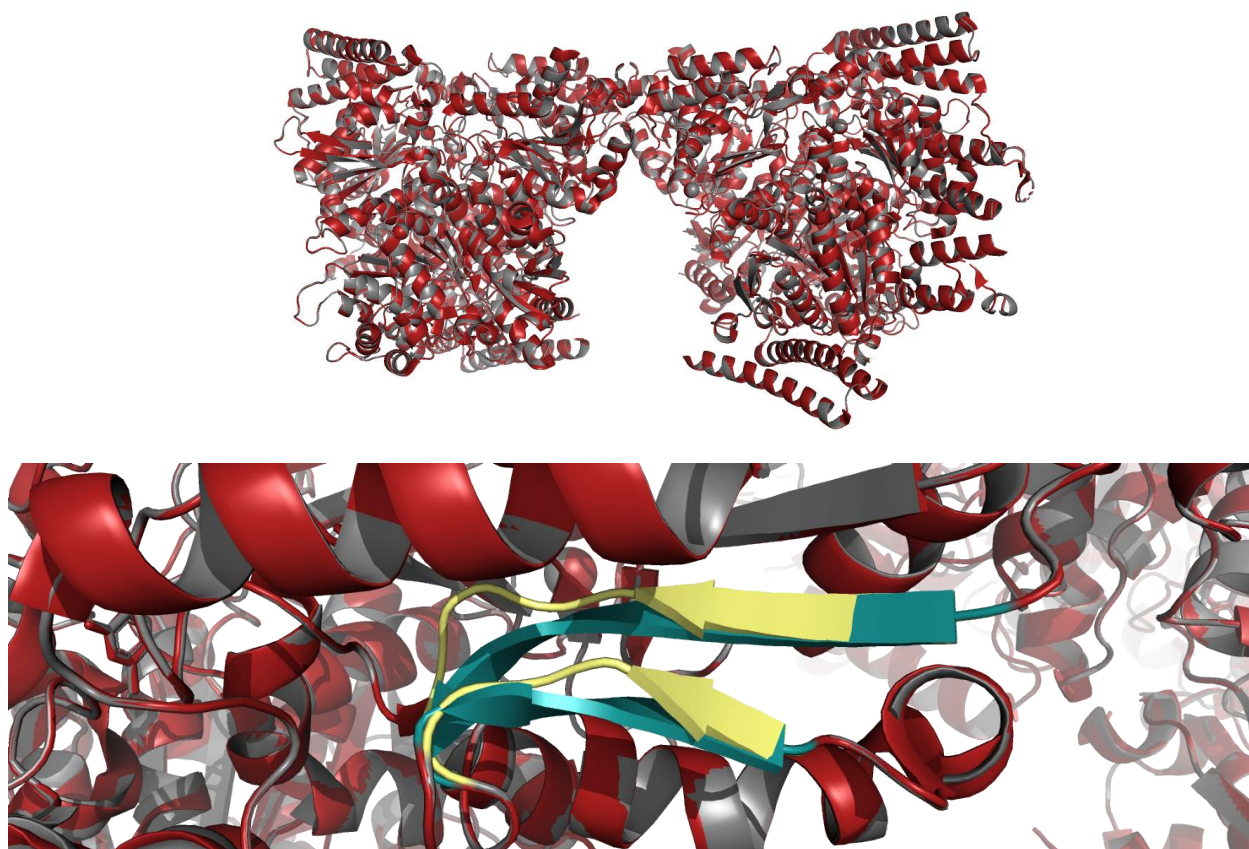


Рисунок 18. Сопоставленные модели PDB 2YIC и PDB-REDO 2YIC. Розовым обозначена исходная (неоптимизированная) модель 2YIC, серым – модель после обработки REDO. Сверху: общий вид структур, Снизу: отдельный участок цепи А, на котором произошло расхождение структур: бирюзовым обозначен фрагмент структуры до обработки REDO, желтым – после REDO.

На Рис. 18 показаны остатки цепи А 1163 – 1171 и 1196 – 1200, в которых структуры не были наложены друг на друга. Интересно, что на обоих участках для исходной структуры был смоделирован более длинный β -тяж, чем аналогичный участок в модели REDO, то есть там, где в неоптимальной структуре продолжается β -тяж, в оптимизированной модели показан неупорядоченный участок полипептидной цепи.

Заключение

Исходя из приведенных в этом отчете данных, можно сделать вывод о том, что построенная модель в целом хорошо соответствует результатам PCA-анализа. Основные параметры качества отражают высокую степень качества расшифровки структуры белка MsKGD_{Δ360}. Тем не менее, удалось обнаружить некоторое количество маргинальных остатков и неточностей. В ходе рассмотрения нескольких маргинальных остатков были обнаружены как особенности структуры, так и неоднозначности расшифровки структуры белка MsKGD_{Δ360}. В частности, рассчитываемый сервером EDS R-factor существенно отличается от указанного для структуры 2YIC в PDB, причем, вообще говоря, лежит на пороге, по которому оценивается качество модели (~25%). Кроме того, можно предположить «подгонку» по некоторым параметрам с удивительно хорошими показателями – например, “clashscore” и карта Рамачандрана. Также хотелось бы отметить не очень понятное несоответствие данных, представленных на странице структуры [2YIC](#), с результатами использованных для оптимизации и проверки сервисов – например, отсутствие маргиналов по длинам связей и углам, хотя таковые были обнаружены и даже рассмотрены подробнее в ходе работы.

Тем не менее, указанные недостатки выглядят незначительными, поскольку в целом структура белка расшифрована с высокой точностью, соответствующей указанному разрешению – 1.96 Å.

Список литературы

- Bricogne, G., Blanc, E., Brandl, M., Flensburg, C., Keller, P., Paciorek, W., et al. (2011). *BUSTER v.2.11.2*. [Online]. BUSTER v.2.11.2. Available: <http://www.globalphasing.com> [Accessed].
- Davis, I.W., Leaver-Fay, A., Chen, V.B., Block, J.N., Kapral, G.J., Wang, X., et al. (2007). MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res* 35(Web Server issue), W375-383. doi: 10.1093/nar/gkm216.
- Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010). Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* 66(Pt 4), 486-501. doi: 10.1107/S0907444910007493.
- Frank, R.A., Price, A.J., Northrop, F.D., Perham, R.N., and Luisi, B.F. (2007). Crystal structure of the E1 component of the Escherichia coli 2-oxoglutarate dehydrogenase multienzyme complex. *J Mol Biol* 368(3), 639-651. doi: 10.1016/j.jmb.2007.01.080.

- Hoffelder, M., Raasch, K., van Ooyen, J., and Eggeling, L. (2010). The E2 domain of OdhA of *Corynebacterium glutamicum* has succinyltransferase activity dependent on lipoyl residues of the acetyltransferase AceF. *J Bacteriol* 192(19), 5203-5211. doi: 10.1128/JB.00597-10.
- Kleywegt, G.J., and Jones, T.A. (1998). Databases in protein crystallography. *Acta Crystallogr D Biol Crystallogr* 54(Pt 6 Pt 1), 1119-1131.
- Murshudov, G.N., Vagin, A.A., and Dodson, E.J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* 53(Pt 3), 240-255. doi: 10.1107/S0907444996012255.
- Trapani, S., and Navaza, J. (2008). AMoRe: classical and modern. *Acta Crystallogr D Biol Crystallogr* 64(Pt 1), 11-16. doi: 10.1107/S0907444907044460.
- Wagner, T., Bellinzoni, M., Wehenkel, A., O'Hare, H.M., and Alzari, P.M. (2011). Functional plasticity and allosteric regulation of alpha-ketoglutarate decarboxylase in central mycobacterial metabolism. *Chem Biol* 18(8), 1011-1020. doi: 10.1016/j.chembiol.2011.06.004.
- Yang, H., Peisach, E., Westbrook, J.D., Young, J., Berman, H.M., and Burley, S.K. (2016). DCC: a Swiss army knife for structure factor analysis and validation. *J Appl Crystallogr* 49(Pt 3), 1081-1084. doi: 10.1107/S1600576716004428.