

# Оценка качества структуры белка пептидогликановой гликозилтрансферазы из протеобактерии *Burkholderia ambifaria* (PDB ID: 5uy7)

## Аннотация

В отчете была проведена оценка качества расшифровки структуры белка с PDB ID: 5uy7, полученной методом рентгеноструктурного анализа. Были изучены значения идентификаторов качества модели и несколько маргинальных остатков. А также произведено сравнение модели из PDB с моделью из PDB-redo.

## Введение

Пептидогликановая гликозилтрансфераза (PGT) – фермент, который необходим для синтеза клеточной стенки бактерий. Он катализирует гликозилтрансфер, образование пептидогликана из субъединиц дисахарида в межклеточном пространстве. Субъединицы дисахарида (NAG-NAM) транслоцированного липида II (дифосфолипид дисахарид-пентапептид) соединяются с помощью PGT, образуя линейные углеводородные цепи, которые соединены через пептидные фрагменты. NAG – N-ацетилглюкозамин, NAM – N-ацетилмураминовая кислота. Механизм гликозилтрансфера еще не до конца изучен, лучше понять его поможет расшифровка структуры PGT белка [3].

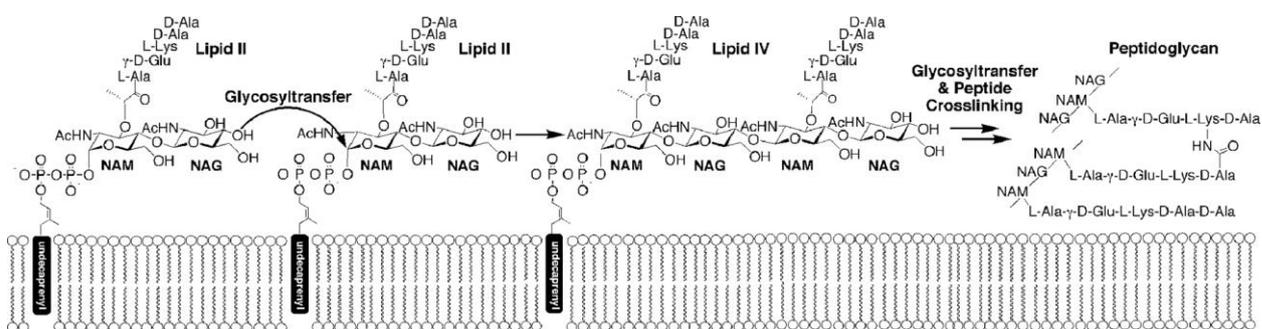


Рис. 1 Синтез пептидогликана.

PGTs определяются по 5ти консервативным мотивам в последовательности. Также белок может находиться в двух состояниях:

1. Когда в бифункциональных белках N концевой гликозилтрансферазный домен, также содержит C концевой транспептидазный домен, такие белки названы РВРа (пенициллин связывающие белки класса А)
2. В виде монофункционального белка (MGT), когда не содержит транспептидазного домена.

В разных бактериях разное количество и типы PGT. У некоторых PGT играет основную роль в процессе элонгации в клеточном цикле, у других в делении при построении перегородки.

Изучаемая, структура 5uy7 представляет собой кристаллическую структуру PGT белка, выделенного из протеобактерии *Burkholderia ambifaria*. Это грамм отрицательная, аэробная, палочковидная, подвижная бактерия. Среди бактерий этого

есть опасные патогены, кроме того, они имеют хорошую устойчивость к антибиотикам, что затрудняет борьбу с ними. Это увеличивает необходимость в их изучении.

## Результаты

### *Общая информация о модели*

Структура 5cu7 состоит из одной цепи белка PGT. 03.08.2017 эта структура была размещена в PDB (Protein Data Bank) группами исследователей из центра инфекционных заболеваний Сиэттла (Edwards, T.E., Abendroth, J., Lorimer, D., Seattle Structural Genomics Center for Infectious Disease) [1]. Статья авторов еще не вышла в публикацию.

Структура была получена методом PCA. Она имеет разрешение 1.65 Å, диапазон разрешений 1.65 – 47.38 Å. Число измеренных рефлексов модели равно 42 618. Полнота данных - 99.9%. Для решения фазовой проблемы был использован метод аномального рассеивания сигналов SAD.

Параметры кристаллографической ячейки: тип кристаллографической симметрии -  $C22_1$ ; длины направляющих векторов кристалла - **a**: 68.8 Å **b**: 74.31 Å **c**: 137.17 Å; углы между векторами - **α**: 90° **β**: 90° **γ**: 90°. Число молекул в ячейке равно 8. Ассиметрическая единица совпадает с биологической. Некристаллографических симметрий в ассиметрической ячейке нет.

Полученная структура состоит из одной цепи А, содержащей 579 аминокислотных остатков в ассиметрической ячейке, и является мономером. Также в структуре присутствуют лиганды: 5 сульфат ионов. Одна молекула состоит из 316 аминокислотных остатка, из них 4 остатка имеют альтернативные конформации. Молекулы воды присутствуют в виде 299 атомов О, с 4мя альтернативными конформациями. Однако авторы предполагают, что белок был протеолизирован в течении кристаллизации, и поэтому остался только пенициллин-связывающий домен. [6]

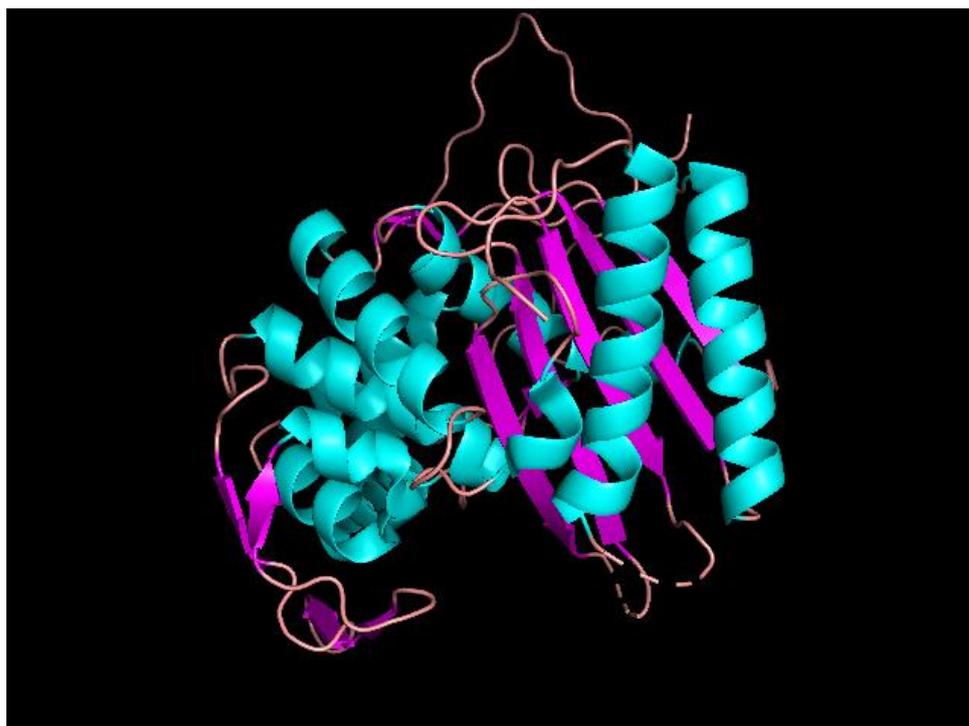


Рис. 2 Структура 5uу7.

### **Индикаторы качества модели**

Ниже приведена таблица индикаторов R и R-free. R-фактор характеризует соответствие модели экспериментальным данным. По мере оптимизации модели он минимизируется. Показатели меньше 0.25 считаются хорошими. R-free фактор защищает от переоптимизации. Он вычисляется как R-фактор, но по небольшой выборке случайно выбранных структурных факторов. Он должен быть примерно равен R-фактору, отличаться не более чем на 10%. Иначе велика вероятность переоптимизации. [2]

Таблица 1. Значения R-фактора и R-free для 5uу7.

Параметр	Значение
R-фактор	0.159
R-free	0.187
Выборка структурных факторов для вычисления R-free	2049 (4.81%)

Для этой модели значение R-фактора равно 0.159. Это меньше, чем 0.25, поэтому он считается хорошим, что говорит о приличном соответствии модели эксперименту.

Количество рефлексов, взятых для расчета R-free = 2049, что составляет 4.81% от общего числа. R-free равен 0.187, что немного больше чем значение R-фактора. Разница R-free – R-фактор = 0.028 небольшая (2.8%), поэтому вероятность переоптимизации мала.

Ниже представлена таблица с количеством маргинальных остатков в структуре по разным критериям.

Таблица 2. Маргинальные остатки в структуре 5uу7.

Критерий	Количество маргинальных остатков/процент
Длины связей	0/0
Углы связей	0/0
RSR	14/4.4% (14 из 316)
Карта Рамачандрана	0/0
Ротамеры	0/0
Перекрытия (clashes)	6/1% (6 из 579)

Рисунок ниже показывает качество модели 5uу7 по сравнению с другими моделями из PDB. Градиент цвета показывает, насколько лучше или хуже значения критериев для данной модели средних значений. Синий показывает лучшие значения, а красный худшие.

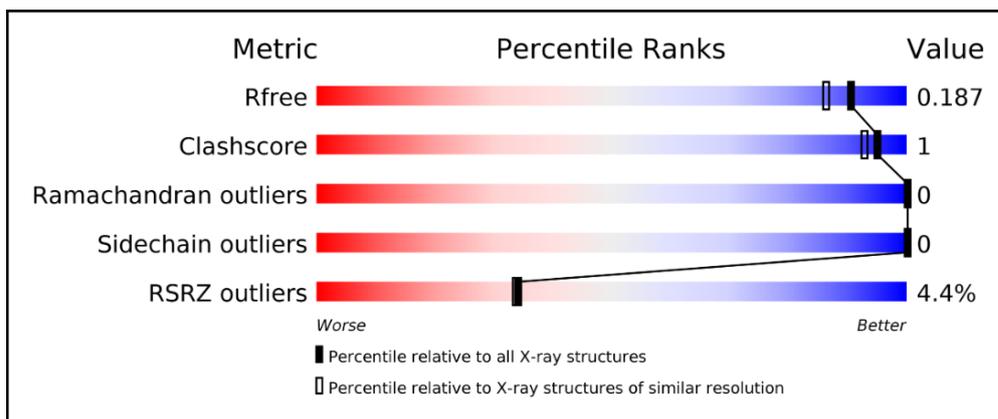


Рис. 3 Качество модели 5uy7 в сравнении с другими моделями из PDB.

Из рисунка 3 видно, что модель 5uy7 имеет качество выше среднего по всем параметрам, кроме RSRZ.

Параметр RSR используется для оценки соответствия каждого аминокислотного остатка экспериментальной электронной плотности. Принято производить оценку через Z-score (RSRZ). RSRZ показывает насколько значение RSR для остатка этой модели отличается от среднего RSR по моделям из PDB со схожим разрешением. Для модели 5uy7 RSRZ равен 4.4%, это хуже, чем в среднем по моделям из PDB. Далее маргинальные остатки будут рассмотрены подробнее.

Для более подробного анализа индикаторов качества модели был использован сервис MolProbity [4], который позволяет строить карты Рамачандрана, выявлять маргинальные остатки, инверсии боковых цепей Asn, Gln, His и недопустимых наложений атомов. Используя этот сервер было добавлено 2298 атомов водорода, из которых положение 67 были скорректированы для оптимизации водородных связей. Ниже представлена таблица суммарных характеристик, выданная данным сервисом.

All-Atom Contacts	Clashscore, all atoms:	1.51		99 <sup>th</sup> percentile* (N=829, 1.65Å ± 0.25Å)
	Clashscore is the number of serious steric overlaps (> 0.4 Å) per 1000 atoms.			
Protein Geometry	Poor rotamers	0	0.00%	Goal: <0.3%
	Favored rotamers	236	98.33%	Goal: >98%
	Ramachandran outliers	0	0.00%	Goal: <0.05%
	Ramachandran favored	303	98.38%	Goal: >98%
	MolProbity score <sup>^</sup>	0.89		100 <sup>th</sup> percentile* (N=9033, 1.65Å ± 0.25Å)
	Cβ deviations >0.25Å	0	0.00%	Goal: 0
	Bad bonds:	0 / 2399	0.00%	Goal: 0%
Bad angles:	0 / 3283	0.00%	Goal: <0.1%	
Peptide Omegas	Cis Prolines:	3 / 23	13.04%	Expected: ≤1 per chain, or ≤5%

Рис. 4 Индикаторы качества модели 5uy7.

Рассмотрим полученные параметры. Как видно из рисунка 4 все параметры входят в необходимый предел, что говорит о высоком качестве модели. Первый параметр ClashScore - это число недопустимых наложений атомов на 1000. Наложением является пересечение Ван-дер-Ваальсовых радиусов более, чем на 0.4 Å. Для 5uy7 ClashScore равно 1.51, что является 99-тым перцентилем. Самый лучший перцентиль 100, значит по этому параметру модель 5uy7 имеет почти наилучшее качество среди моделей с таким же разрешением (в диапазоне 1.65Å ± 0.25Å).

Далее следуют несколько нулевых параметров. Poor rotamers - число остатков с маргинальными по отклонению от ротамеров боковыми цепями. Ramachandran outliers - полные маргиналы по карте Рамачандрана те, которые лежат вне допустимой области. C $\beta$  deviations  $>0.25\text{\AA}$  - число C $\beta$ -атомов с неприемлемым отклонением от ожидаемого положения. Bad backbone bonds - число ковалентных связей, существенно отклоняющихся от теории. Bad backbone angles - число валентных углов, существенно отклоняющихся от теории. То, что эти показатели равны 0, говорит о том, что нарушений по длине химических связей и величинам валентных углов нет.

Ramachandran favored - число и процент остатков в предпочитаемой области. В модели 5uy7 таких остатков 303, что составляет больше 98% от общего числа и считается хорошим. MolProbity score - интегральная оценка структуры по данным этого сервиса. Она составляет 0.89 и является идеальным перцентилем.

Favored rotamers – число остатков с боковыми цепями не маргинальным по отклонению от ротамеров. Ротамеры - это наборы предпочитаемых положений боковых цепей для остатка, которые задаются допустимыми значениями торсионных углов. В структуре 5uy7 все 236 остатков имеют “хорошие” ротамеры.

Карта Рамачандрана позволяет оценить, насколько хорошо конформация остова модели соответствует предпочитаемому положению торсионных углов  $\phi$  и  $\psi$ . С помощью сервиса MolProbity были построены карты Рамачандрана для структуры 5uy7, они представлены на рисунках 5-7 ниже. Голубыми линиями ограничены участки предпочитаемых областей, синими допустимых областей. Точками являются отдельные аминокислотные остатки.

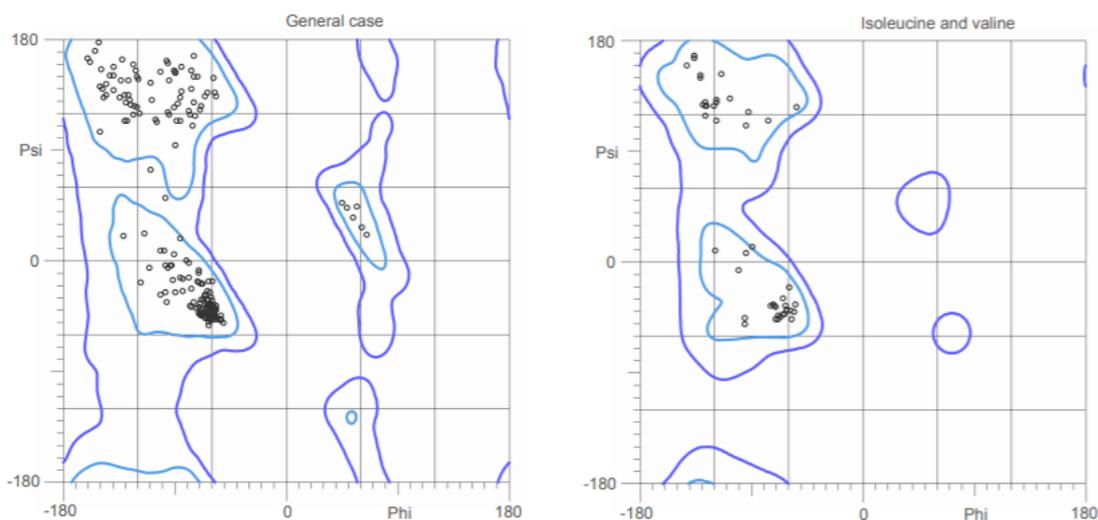


Рис. 5 Карта Рамачандрана для структуры 5uy7 (слева). Карта Рамачандрана для остатков изолейцина и валина для структуры 5uy7 (справа).

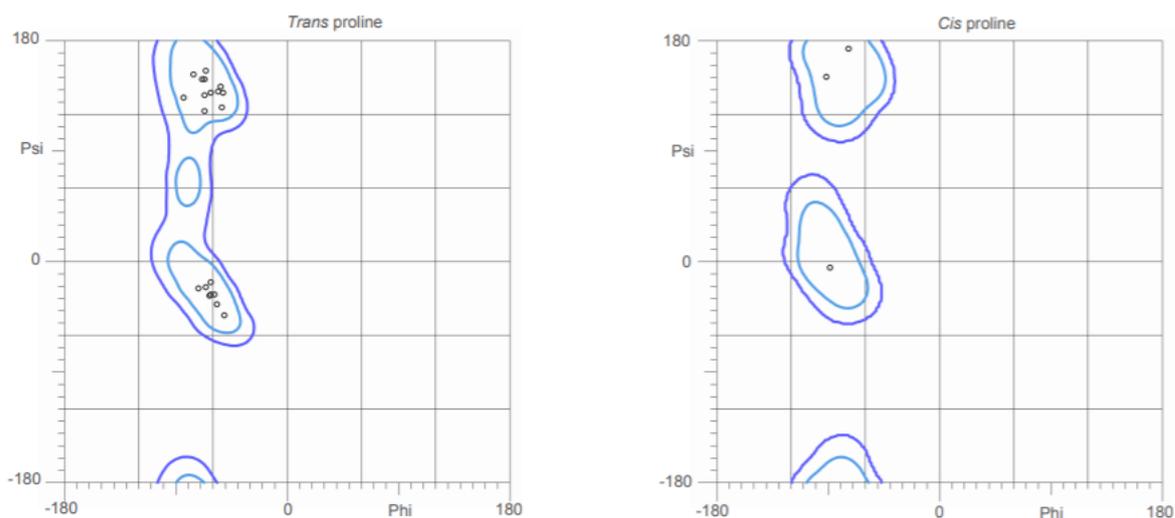


Рис. 6 Карта Рамачандрана для остатков тран и цис пролина для структуры 5uу7 (справа и слева соответственно).

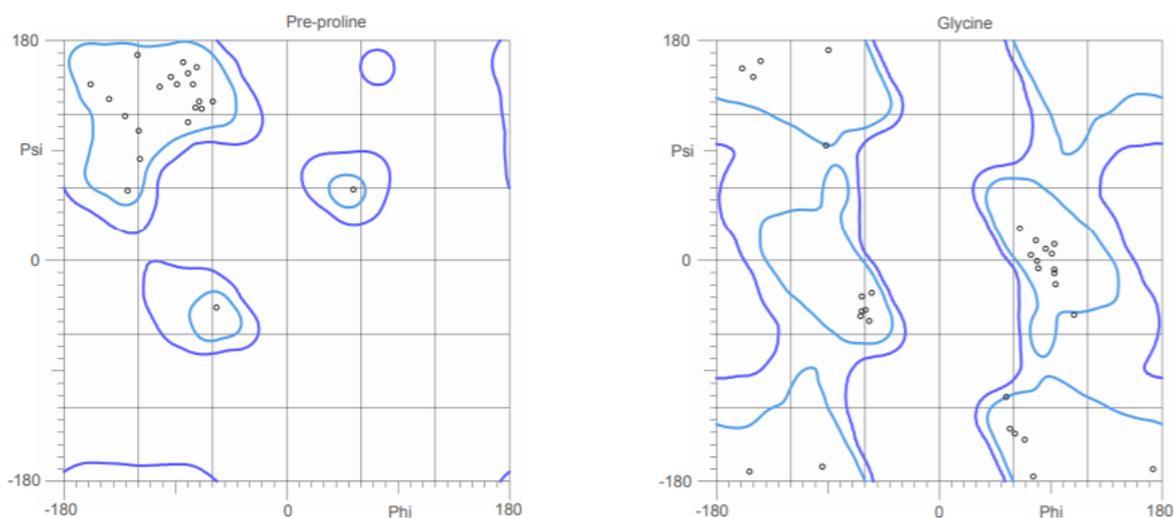


Рис. 7 Карта Рамачандрана для остатков пре-пролина и глицина для структуры 5uу7 (справа и слева соответственно).

Из представленных рисунков 5-7 видно, что практически все аминокислотные остатки лежат в предпочитаемой области, кроме некоторых лежащих в допустимой области. Из выдачи MolProbity 98.4%, то есть 307 из 312, остатков лежат в предпочитаемой области. И 100% (312 из 312) – в допустимой. Это говорит о том, что маргиналов по значениям торсионных углов нет.

Остатки изолейцина, валина и глицина известны тенденцией принимать необычные значения торсионных углов, поэтому из карты были построены отдельно.

Были обнаружены остатки пролина в cis конформации 3 из 23, что составляет 13.04%. На цепь ожидается  $\leq 1$  или  $\leq 5\%$ . В пептидных связях чаще участвует тран-пролин, потому что цис конформация имеет стерическое отталкивание альфа атомов.

## Анализ маргинальных остатков

На рисунке 8 ниже показаны графики качества остатков в полипептидной цепи молекулы белка PGT. Красная линия и процент над ней (2%) отображают процент остатков, плохо вписывающихся в электронную плотность. Линия ниже показывает соотношение в процентах остатков разного качества в цепи. Ниже отображена последовательность аминокислотных остатков цепи, окрашенных по геометрическим критериям: зеленый – не маргинальный остаток, желтый – маргинальный по одному критерию, оранжевый – по 2м, красный – по 3м и более. Красная точка сверху над остатком указывает на  $RSRZ > 2$ , то есть плохое соответствие остатка электронной плотности. Серым покрашены остатки, которые были в кристалле, но не вошли в модель. Зеленые стретчи изображают 2 и более не маргинальных остатка.

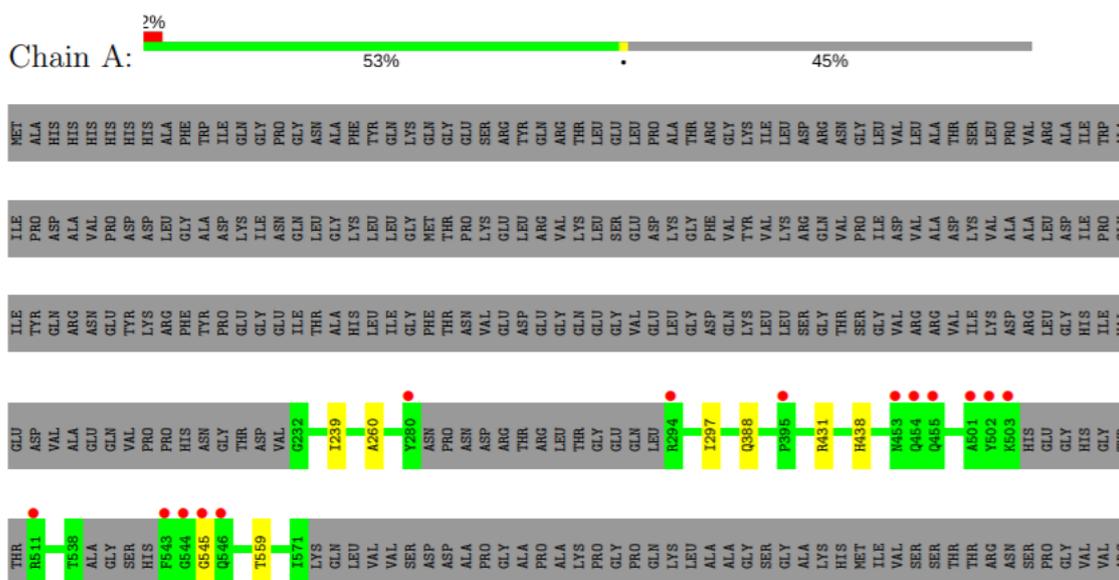


Рис. 8 График качества остатков в структуре 5cy7.

Из рисунка 8 видно, что довольно многие остатки были в кристалле, но не вошли в модель (серые) – 45%. Из вошедших в модель остатков, многие не являются маргинальными ни по одному критерию. 8 остатков являются маргинальными по одному критерию. 14 остатков имеют плохое соответствие электронной плотности, один из них еще и маргинален по одному критерию. Однако процент плохо вписывающихся остатков небольшой – 2%.

В таблице ниже представлены маргинальные остатки структуры с указанием критерия. Таблица содержит 15 остатков, в нее не вошли 6 остатков с  $RSRZ > 2$ , приведены с наибольшими  $RSRZ$ .

Таблица 3. Список маргинальных остатков в структуре 5cy7.

Остаток	Объяснение маргинальности	Значение
Ala 260 с Ile 297	Перекрытие Ван-дер-Ваальсовых радиусов между атомами HB1 и HG12	0.62 Å (размер перекрытия)
Arg 431[A] с HOH 801	Перекрытие Ван-дер-Ваальсовых радиусов между атомами NH2 и O	0.55 Å

Gly 545 с HOH 815	Перекрытие Ван-дер-Ваальсовых радиусов между атомами HA3 и O	0.49 Å
Gln 388 с HOH 801	Перекрытие Ван-дер-Ваальсовых радиусов между атомами O и O	0.47 Å
Ile 239[B] с Thr 559	Перекрытие Ван-дер-Ваальсовых радиусов между атомами HD12 и HG23	0.46 Å
His 438	Перекрытие Ван-дер-Ваальсовых радиусов между атомами этого остатка CD2 и O	0.41 Å
Gln 546	Подозрение на необходимость в инверсии боковой цепи Плохое соответствие экспериментальной электронной плотности	RSRZ = 2.6
Phe 543	Плохое соответствие экспериментальной электронной плотности	RSRZ = 5.8
Tyr 502	Плохое соответствие экспериментальной электронной плотности	RSRZ = 5.5
Gln 454	Плохое соответствие экспериментальной электронной плотности	RSRZ = 3.2
Gly 544	Плохое соответствие экспериментальной электронной плотности	RSRZ = 3.1
Asn 453	Плохое соответствие экспериментальной электронной плотности	RSRZ = 3.0
Gln 455	Плохое соответствие экспериментальной электронной плотности	RSRZ = 3.0
Arg 294	Плохое соответствие экспериментальной электронной плотности	RSRZ = 2.5

Рассмотрим некоторые маргинальные остатки подробно. Для этого был взят PDB файл с добавленными атомами водорода.

Для начала рассмотрим перекрытие аминокислотных остатков Ala 260 и Ile 297. Изображение этих остатков на рисунке 9 ниже.

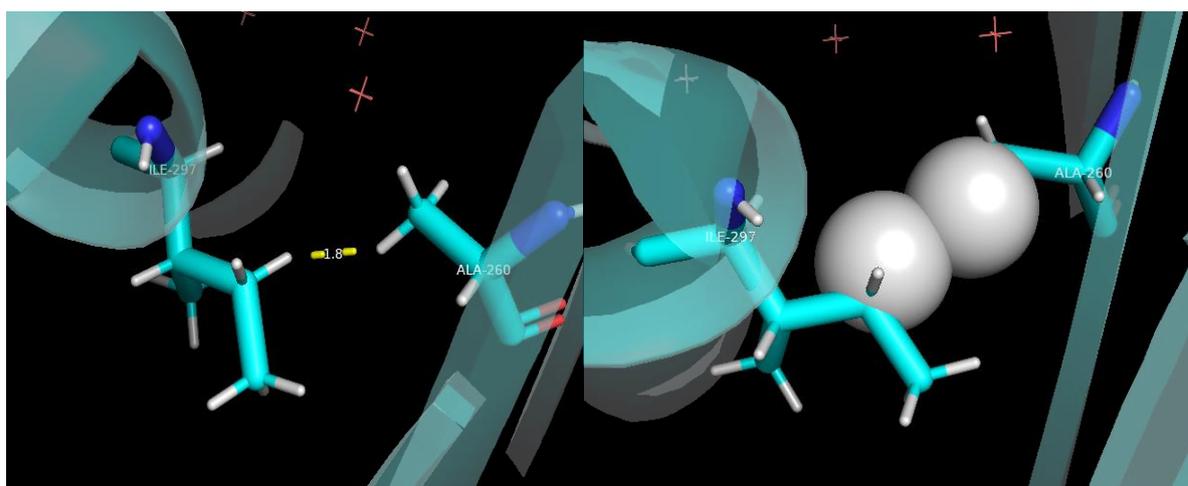


Рис. 9 Остатки Ala 260 (справа) и Ile 297 (слева). Вверху изображена плохая водородная связь, внизу перекрытие радиусов атомов.

Между остатками Ala 260 с Ile 297 должна была образовываться водородная связь (3.5 Å), однако вместо этого связь составляет лишь 1.8 Å. То есть произошло перекрытие Ван-

дер-Ваальсовых радиусов атомов HB1 и HG12 (составляют  $0.5 \text{ \AA}$ ) на  $0.62 \text{ \AA}$ . Ala 260 расположен в бета-листе, а Пе 297 – в бета-тяже. Вероятно, существует подвижность между альфа спиралью и бета-тяжом, поэтому расположение атомов можно считать приемлемым.

Далее рассмотрим перекрывание Ван-дер-Ваальсовых радиусов атомов между остатками Arg 431[A] и HON 801. Их изображение на рисунке 10 ниже.

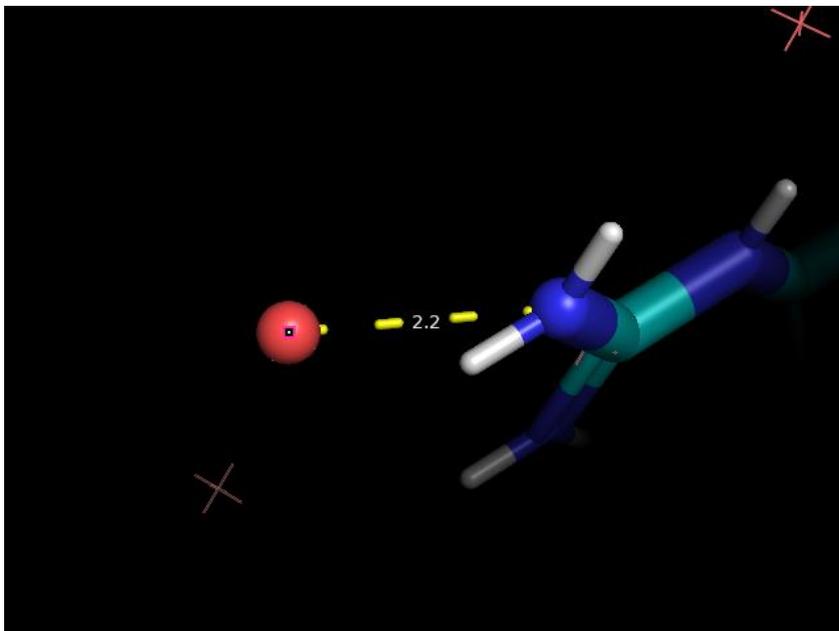


Рис. 10 Остатки Arg 431[A] и HON 801 (красная сфера слева). Желтым пунктиром изображена водородная связь.

Перекрывание между атомами NH<sub>2</sub> и O остатков Arg 431[A] и HON 801 составляет  $0.55 \text{ \AA}$ . В этом случае мы наблюдаем ту же картину, что и в предыдущем, между этими атомами должна была образоваться водородная связь обычно с расстоянием  $3.5 \text{ \AA}$ , но из-за сближения атомов образуется связь только  $2.2 \text{ \AA}$ . Однако остаток Arg 431 имеет 2 изоформы: А и В. Вероятно более правильной является изоформа В, при которой не образуется перекрытия.

Затем рассмотрим аминокислотный остаток Gln 546, который изображен на рисунке 11. Для него есть подозрение о необходимости инверсии боковой цепи, и он плохо вписан в экспериментальную электронную плотность,  $RSRZ = 2.6$  (хорошим считается  $RSRZ = 2.0$ ). На рисунке 11 видно, что несколько атомов боковой цепи совсем не вписаны в электронную плотность. С помощью сервиса MolProbity был получен файл с инверсией этого остатка (слева) (атомы N и O боковой цепи поменяны местами). Видно, что после инверсии остаток стал немного лучше вписан в электронную плотность. Однако электронное облако есть только для одного атома. В левой части рисунка 11, для модели без инверсии, видно, что облако немного не соответствует атому O, и намного лучше для атома N. Вариант расположения с инверсией выглядит более правдоподобным. Однако то, что остаток плохо вписан в электронную плотность говорит о том, что положения атомов практически не определены, поэтому делать какие-то выводы на этой основе сложно.

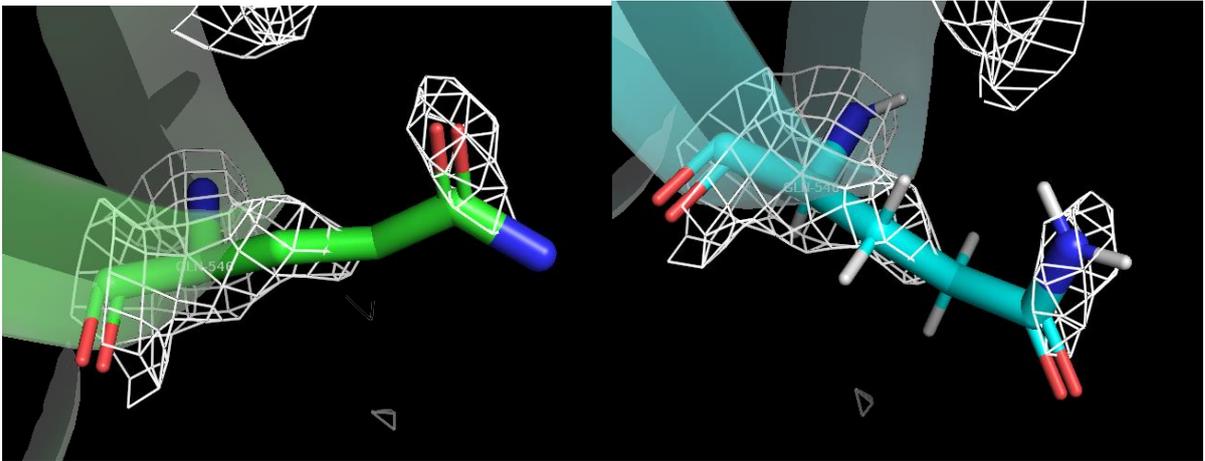


Рис. 11 Остаток Gln 546. Сетка изображена на уровне подрезки 1.0. Справа без инверсии, слева с инверсией.

Рассмотрим остаток Phe 543 с самым большим RSRZ = 5.8 или, другими словами, хуже всего вписанный в экспериментальную электронную плотность.

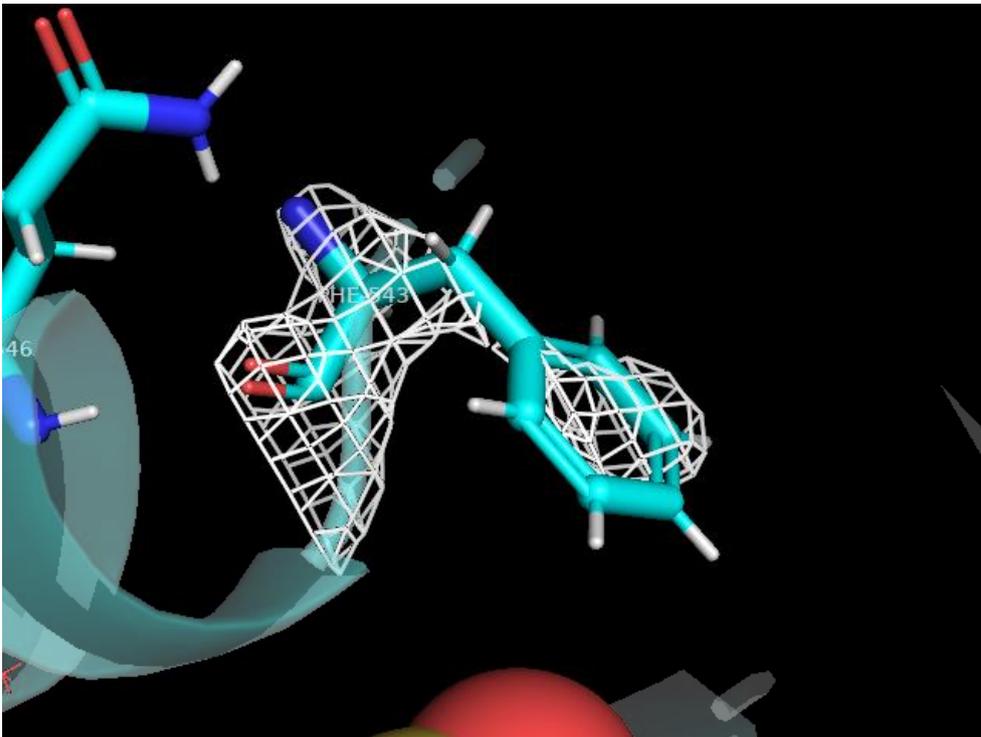


Рис. 12 Остаток Phe 543. Сетка изображена на уровне подрезки 1.0.

Видно, что даже на низком уровне подрезки (1.0) электронная плотность хорошо вписана в остовные атомы и атомы кислорода и азота. Ароматическое кольцо практически не покрыто электронной плотностью. Это видимо связано с низким качеством карты электронной плотности на этом участке цепи. Кроме того, остаток расположен в петле, и возможно обладает некоторой подвижностью.

Рассмотрим еще один остаток, плохо вписанный в экспериментальную электронную плотность – Туг 502 с RSRZ = 5.5. Он изображен на рисунке 13.

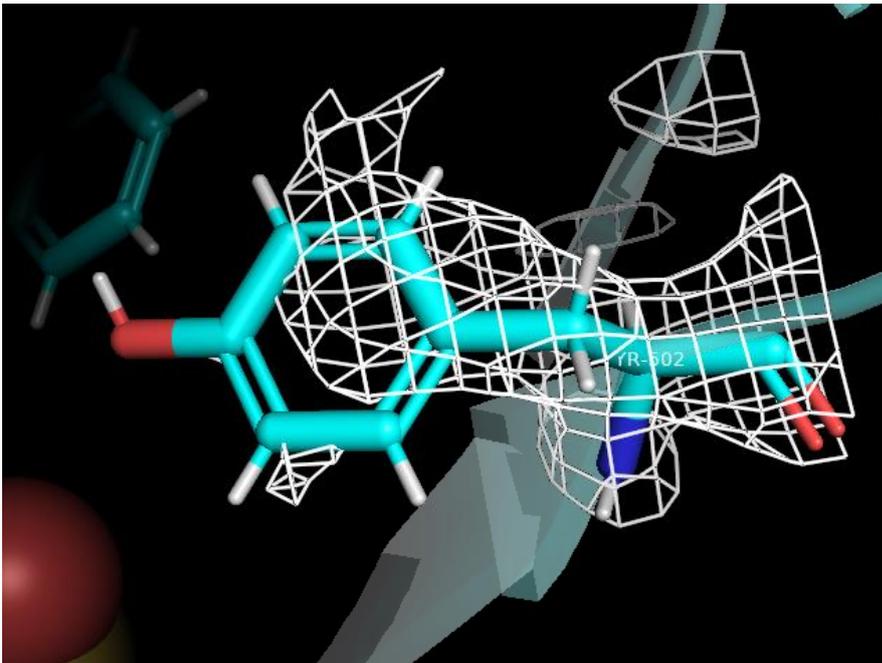


Рис. 13 Остаток Туг 502. Сетка изображена на уровне подрезки 1.0.

Ситуация аналогична, основные атомы и чуть большая часть атомов боковой цепи хорошо вписаны в электронную плотность, в отличие от ароматического кольца. Также мы видим, что атом кислорода на конце ароматического кольца совершенно не вписан, хотя и электронноплотный. Остаток Туг 502 находится на краю бета-тяжа и петли, возможно у него есть некоторая подвижность, осложняющая точное определение его атомов.

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

База данных PDB\_redo содержит оптимизированные версии моделей, существующих в PDB. Проведем сравнение для структуры 5цy7 модели PDB\_redo с PDB [5].

В PDB\_redo было изменено положение 4х ротамеров, произведена инверсия 14 боковых цепей, удалены 34 остатка воды. В результате таких изменений упали значения индикаторов качества. R-фактор снизился с 0,1552 до 0,1513. R-free – с 0,1824 до 0,1746.

Z-score по картам Рамачандрана немного ухудшился, а качество ротамеров немного улучшилось.

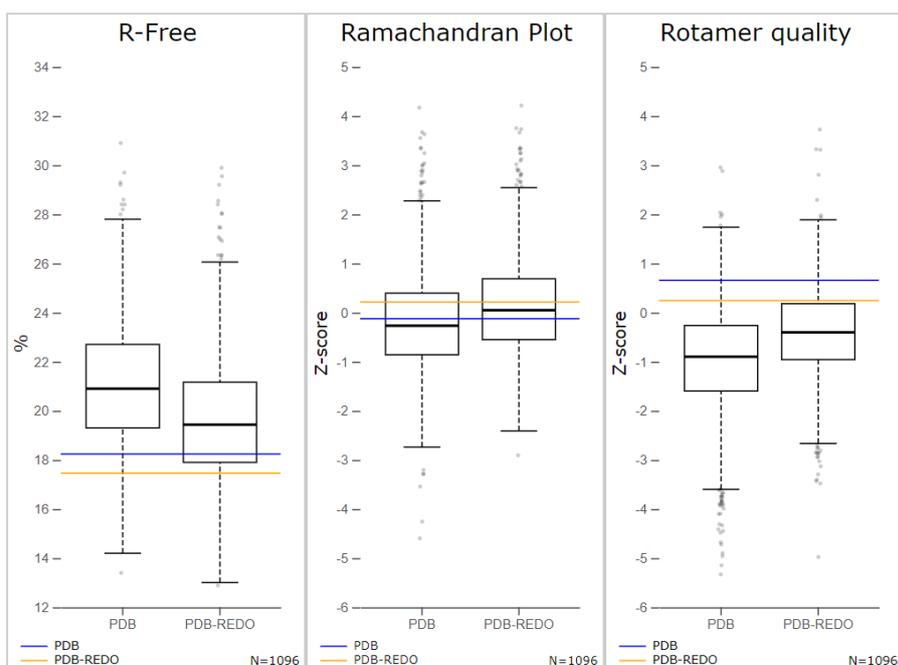


Рис. 14 Сравнение общих параметров структуры 5u7 до и после оптимизации.

Для анализа оптимизации был использован сервис MolProbity, для файла со структурой 5u7 построенной PDB\_redo. Результаты анализа приведены на рисунке 15.

All-Atom Contacts	Clashscore, all atoms:	2.06	99 <sup>th</sup> percentile* (N=829, 1.65Å ± 0.25Å)	
	Clashscore is the number of serious steric overlaps (> 0.4 Å) per 1000 atoms.			
Protein Geometry	Poor rotamers	1	0.39%	Goal: <0.3%
	Favored rotamers	248	97.64%	Goal: >98%
	Ramachandran outliers	0	0.00%	Goal: <0.05%
	Ramachandran favored	304	98.70%	Goal: >98%
	MolProbity score <sup>^</sup>	0.98	100 <sup>th</sup> percentile* (N=9033, 1.65Å ± 0.25Å)	
	Cβ deviations >0.25Å	0	0.00%	Goal: 0
	Bad bonds:	1 / 2466	0.04%	Goal: 0%
Bad angles:	1 / 3359	0.03%	Goal: <0.1%	
Peptide Omegas	Cis Prolines:	3 / 23	13.04%	Expected: ≤1 per chain, or ≤5%

Рисунок 15. Анализ оптимизированной структуры 5u7 в MolProbity.

Видно, что появились отклонения от нормальных длин связей и углов, увеличилось количество плохих ротамеров и соответственно уменьшилось число хороших. Ухудшился Clashscore, который показывает количество стерических перекрытий на 1000 атомов с 1.51 до 2.06. Никакого улучшения других параметров не наблюдается.

В результате, оптимизация не дала каких-то хороших результатов. Возможно качество модели уже было довольно велико, дальнейшее улучшение видимо требует ручной проверки всех сомнительных остатков. Использование модели PDB\_redo для структуры 5u7 вместо модели PDB не имеет смысла.

## Выводы

Структура белка PGT из протеобактерии *Burkholderia ambifaria* (PDB ID=5u7) имеет высокое разрешение и довольно качественную модель. Большинство индикаторов качества имеют значения выше среднего среди всех PDB структур с таким же

разрешением. Модель содержит немного маргинальных остатков по перекрытию Ван-дер-Ваальсовых радиусов и плохому соответствию экспериментальной электронной плотности. Однако многие остатки, которые присутствовали в кристалле не вошли в модель. Также в модели есть остаток, возможно нуждающийся в инверсии боковой цепи, и есть 4 остатка с несколькими изоформами.

## **Ссылки**

[1] <https://www.rcsb.org/structure/5uy7>

[2] [https://files.rcsb.org/pub/pdb/validation\\_reports/uy/5uy7/5uy7\\_full\\_validation.pdf](https://files.rcsb.org/pub/pdb/validation_reports/uy/5uy7/5uy7_full_validation.pdf) - Полный отчет по валидации структуры

[3] Crystal structure of a peptidoglycan glycosyltransferase suggests a model for processive glycan chain synthesis. Yanqiu Yuan, Dianah Barrett, Yi Zhang, Daniel Kahne, Piotr Sliz, and Suzanne Walker February 7, 2007. <https://www.pnas.org/content/104/13/5348>

[4] <http://molprobitry.biochem.duke.edu/>

[5] <https://pdb-redo.eu/db/5uy7>

[6] <https://www.ebi.ac.uk/pdbe/entry/pdb/5uy7/>