

МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ИМЕНИ М.В. ЛОМОНОСОВА

Факультет Биоинженерии и Биоинформатики

Хачатурян Марина Александровна

**Отчет по качеству РСА расшифровки структуры
Об-метилгуанин метилтрансферазы из генома археи
*Sulfolobus solfataricus***

Москва 2016

СОДЕРЖАНИЕ

| | |
|--|-----------|
| АННОТАЦИЯ | 3 |
| ВВЕДЕНИЕ | 4 |
| РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ | 5 |
| 1 Основные параметры модели | 5 |
| 2 Значения индикаторов качества модели | 6 |
| 3 Маргинальные остатки | 10 |
| 3.1 MET 1 | 11 |
| 3.2 GLU 44 | 12 |
| 3.3 CYS 29 | 12 |
| 3.4 GLU 33 | 13 |
| 3.5 ARG 61 | 14 |
| 4 Сравнение модели из PDB с моделью из PDB_redo | 15 |
| ЗАКЛЮЧЕНИЕ | 16 |
| СПИСОК ЛИТЕРАТУРЫ | 17 |

АННОТАЦИЯ

В данной работе исследовалась структура выбранного белка, расшифрованная методом РСА. Для этого были использованы сервисы и программы MolProbity, EDS, WhatCheck, PROCHECK и PyMol. С помощью них была исследована как структура целиком, так и отдельные маргинальные по одному или нескольким факторам остатки.

Для более детального рассмотрения были выбраны пять наиболее разнообразных по маргинальности остатков. Одна из групп таких остатков – CYS29, ASP30 и GLU33, обнаруженная в этой работе, образует интересную, предположительно ошибочную, конформацию.

После детального изучения маргинальных остатков для улучшения модели был использован сервис PDB_redo. Этому сервису не удалось однозначно улучшить качество модели структуры белка.

ВВЕДЕНИЕ

Об-метилгуанин метилтрансфераза участвует в репарации алкилированного гуанина, перенося метильную группу с позиции Об гуанина на остаток цистеина фермента. Ферменты, участвующие в алкилировании ДНК и репарации алкилирования консервативны от прокариот до высших эукариот.

Было показано, что гомологичные репарирующие алкилтрансферазы из высших эукариот выделяются в большом количестве в раковых опухолях в ответ на алкилирующие ДНК лекарства химиотерапии, поэтому они стали потенциальными мишенями противоопухолевых лекарств [1].

Для понимания механизма распознавания алкилированной ДНК необходима 3D структура этого фермента и его комплекса с ДНК. Этот фермент контактирует с малой бороздкой ДНК с помощью конформации спираль-поворот-спираль, расположенной на С-конце, которая стабилизируется аргининовыми пальцами. После распознавания алкильная группа переносится в активном центре на цистеин.

Как правило, перенос алкильной группы вызывает сильную нестабильность фолдинга фермента, поэтому получить 3D структуру удастся только для случая метилирования и бензилирования [2]. Анфолдинг и деградация фермента интересны с точки зрения биологии и механики процесса, предположительно они могут запускать следующие репарирующие ДНК пути и облегчает диссоциацию комплекса [2][3][4].

Такие ферменты встречаются у Бактерий, Архей и Эукариот. Особенное значение они имеют для термофильных бактерий, для которых алкилирование ДНК – серьезное повреждение.

Кристаллические структуры были получены для дикого типа Об-метилгуанин метилтрансферазы (SsOGT) – 4ZYЕ; для мутанта С119L – 4ZYН; для формы с метилированным цистеином (*SsOGT^m*) – 4ZYG; и для ее комплекса с метилированной ДНК (*C119A :: dsDNA^m*) – 4ZYD [5]. Далее будет изучена подробнее расшифровка структуры дикого типа фермента (4ZYЕ).



Рис. 1: Структура Об-метилгуанин метилтрансферазы из *Sulfolobus tokodaii* (PDB 1WRJ). Получено с помощью PyMol.

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

1 Основные параметры модели

Структура нативного белка Об-метилгуанин метилтрансферазы (SsOGT) была расшифрована в 2015 году и в том же году структура была подтверждена и вышла статья [5]. Авторы расшифровки: Miggiano R., Rossi F. и Rizzi, M.

Фазовая проблема решалась методом молекулярного замещения с помощью программы PHASER [6] из пакета PHENIX [7] на основе структуры Об-метилгуанин метилтрансферазы из *Sulfolobus tokodaii* (PDB 1WRJ) (рис. 1). Измеренные структурные факторы лежат в диапазоне $h = [0; 43]$, $k = [0; 44]$, $l = [-41; 41]$. При этом общее количество измеренных рефлексов равняется 21535, из которых 20775 использовалось для оптимизации модели и 760 – для контроля по данным из файла структурных факторов. По данным из EDS всего рефлексов 21920. Разрешение структуры 1.85Å при полноте данных 100.0%. Диапазон разрешений от 56.03Å до 1.85Å. Это очень хорошее разрешение с учетом такой высокой полноты данных. Ячейка кристаллической решетки образована направляющими $a = 94.73\text{Å}$, $b = 94.73\text{Å}$ и $c = 76.71\text{Å}$ и углами $\alpha = 90.00$, $\beta = 90.00$ и $\gamma = 120.00$. Некристаллографической симметрии у этого белка нет.

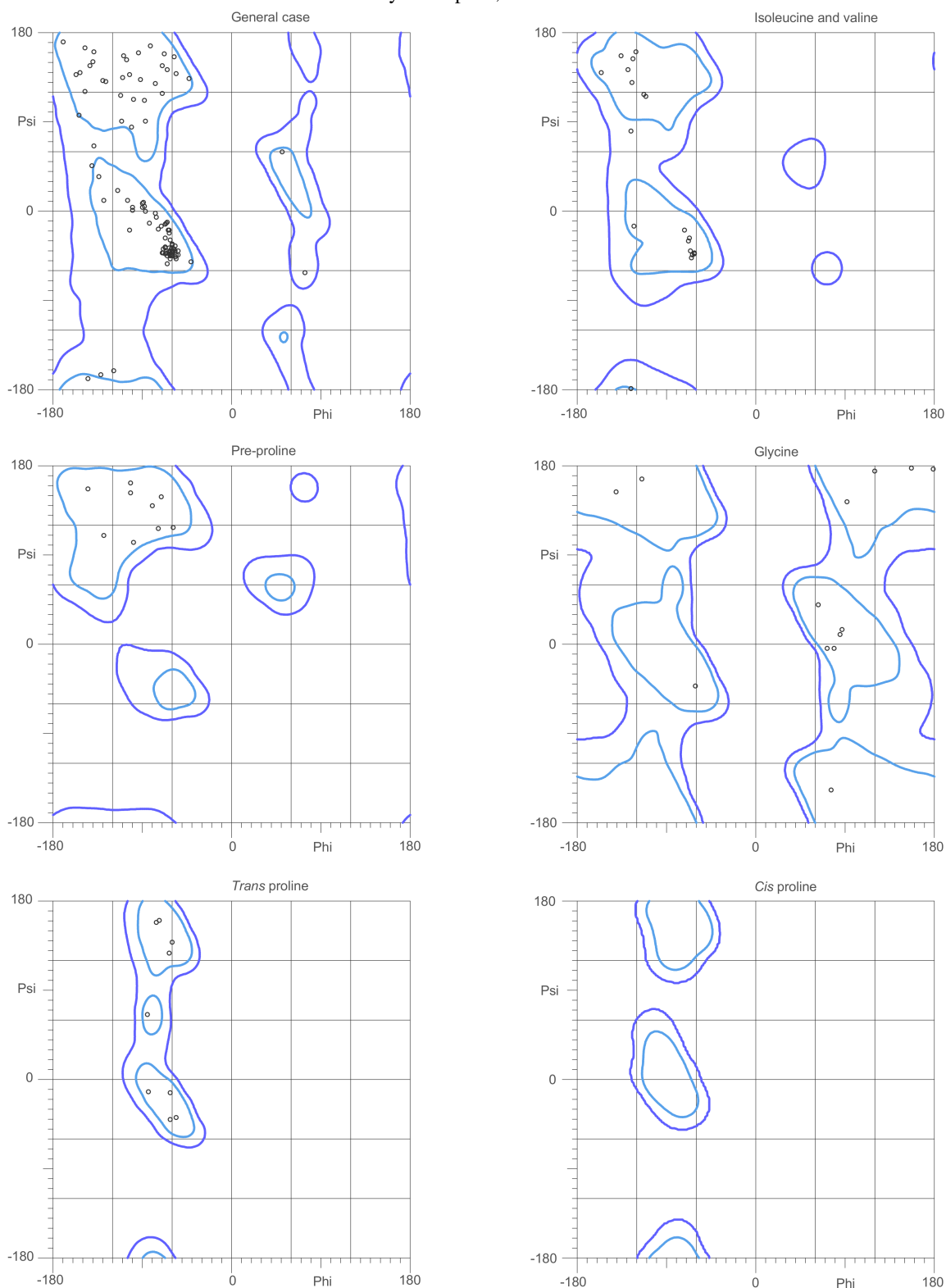


Рис. 2: Карты Рамачандрана для структуры 4ZYU. Получено с помощью MolProbity.

2 Значения индикаторов качества модели

Качество предсказанной модели можно определять, исходя из разных параметров. Одними из таких являются $R_{\text{фактор}}$, R_{free} , ротамеры и др.

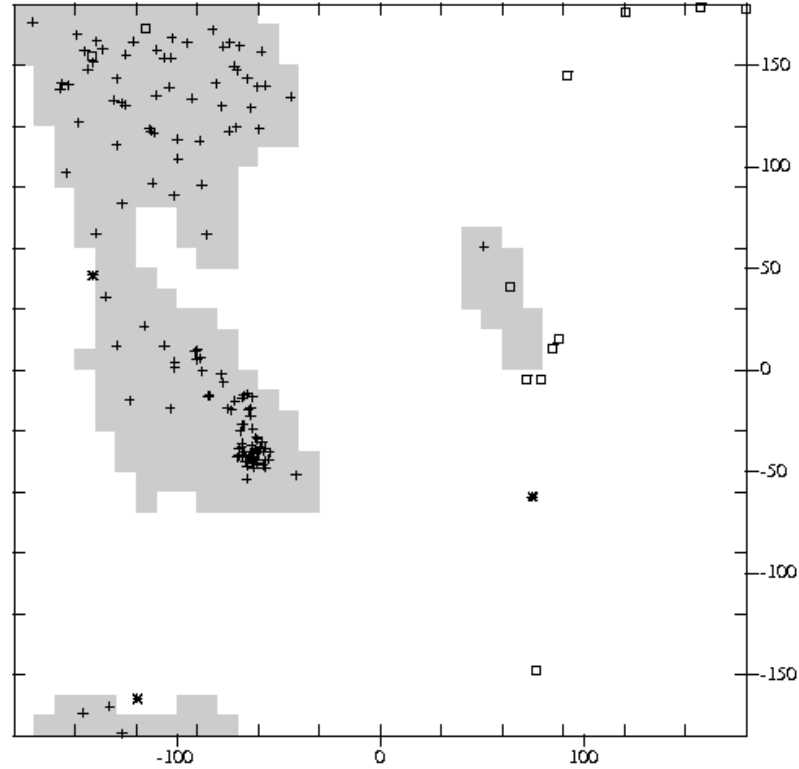


Рис. 3: Карта Рамачандрана для структуры 4ZYU. Получено с помощью EDS.

R-фактор показывает, насколько структура соответствует экспериментальным данным. Он вычисляется по следующей формуле:

$$R_{\text{фактор}} = \frac{\sum_{h,k,l} |F_{hkl}^{\text{calc}} - F_{hkl}^{\text{obs}}|}{\sum_{h,k,l} F_{hkl}^{\text{obs}}} * 100\%.$$

Этот параметр минимизируется в процессе оптимизации модели. Модель считают хорошей, если R-фактор не превышает 25%. В случае модели 4ZYU $R_{\text{фактор}} = 16.1\% < 25\%$, что свидетельствует о хорошем соответствии.

Параметр R_{free} – это контроль качества параметра $R_{\text{фактор}}$, то есть он является контролем от “переподгонки“. Для этого всю выборку рефлексов разделяют на 95%, используемых для оптимизации модели, и 5% контроля. Для модели 4ZYU $R_{\text{free}} = 18.2\%$, что считается хорошим значением, так как оно не превышает 25% и не более чем на 10% отличается от значения $R_{\text{фактор}}$. Еще одним важным параметром является число маргинальных остатков, то есть тех, которые находятся в энергетически невыгодных положениях по углам ϕ и ψ , что определяется по картам Рамачандрана. Для разных аминокислотных остатков допустимые зоны значения углов различаются. Сервис MolProbity это учитывает, поэтому для структуры 4ZYU не находит полных маргиналов (рис. 2). Менее точный сервис EDS находит три полных маргинала: ASP19, CYS31 и LEU115,

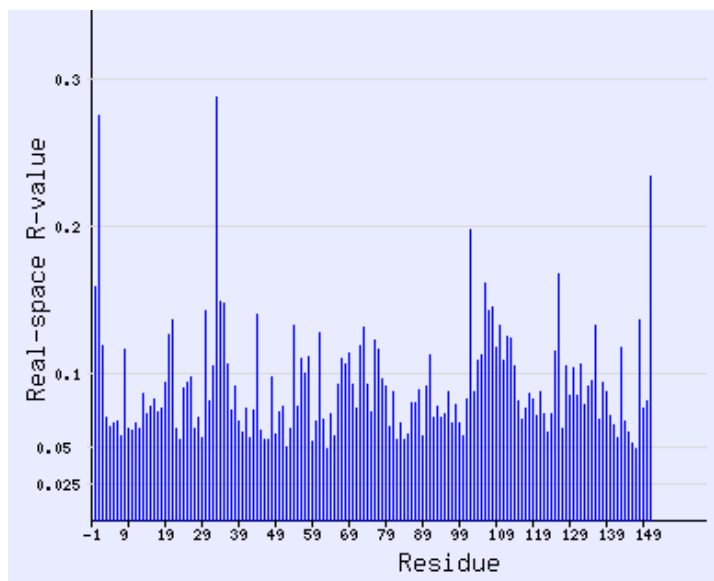


Рис. 4: Значения RSR для остатков структуры 4ZYE. Получено с помощью EDS.

так как использует только усредненную карту Рамачандрана для всех остатков, кроме глицина (рис. 3). Таким образом, по EDS процент полностью маргинальных остатков составляет 2.2%, а по MolProbity – 0.0%.

Среди остатков так же выделяют процент тех, которые лежат на карте Рамачандрана в хороших областях. В хороших структурах желательно, чтобы процент таких остатков не опускался ниже 98%, однако по данным MolProbity в структуре 4ZYE процент таких остатков составляет 96% (144/150), что тоже не так плохо, учитывая общее небольшое число остатков.

Ротамеры – это допустимые положения боковых цепей остатков. В хороших структурах их процент не опускается ниже 98%. В структуре 4ZYE таких остатков 129 (97.99%), 2 остатка (1.50%) имеют совсем недопустимые углы. Еще 2 не отнесены ни в какую из этих групп сервисом MolProbity. Остальные 17 остатков – это глицины, которые не имеют боковых цепей.

RSR – пространственный $R_{\text{фактор}}$ – параметр, позволяющий определить маргинальные остатки по несоответствию боковой цепи "экспериментальной" электронной плотности. Это еще один параметр соответствия модели эксперименту. Для модели 4ZYE RSR равен 0.095, что лежит в пределах нормы и является признаком хорошей модели. Значения RSR для каждого из остатков представлены на рис. 4. Видно, почти все остатки вписаны хорошо, а из четырех, явно выбивающихся, два расположены на концах цепи. Маргинальными же оказались: MET1, GLU33, ARG102 и GLU151. Так как RSR зависит от разрешения модели, существует еще и относительная оценка RSR – RSR-Z-score. Можно говорить о том, что остаток плохо вписан, если $Z > 2$. Для структуры 4ZYE с разрешением 1.85Å значения Z для всех остатков представлены на рис. 5.

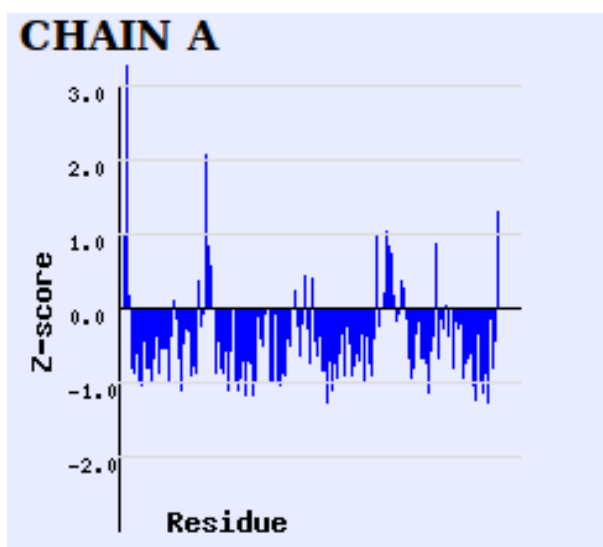


Рис. 5: Значения RSR-Z-score для остатков структуры 4ZYЕ. Получено с помощью EDS.

Из маргинальных остатков остались только MET1 и GLU33. Еще одним фактором оценки качества структуры является комфортность окружения каждого аминокислотного остатка. Информацию об этом можно получить с помощью программы WhatCheck. Для каждой боковой цепи выдается значение Z-score комфортности окружения. Если $Z < -5$, то с такими остатками что-то не так. Среднее значение по структуре довольно хорошее – $-0,739$, однако довольно много остатков имеют $Z < -5$. Они приведены в таблице 1.

Таблица 1: Значения Z-score комфортности окружения для "плохих" остатков структуры 4ZYЕ. Получено с помощью WhatCheck.

| Остаток | MET1 | LYS8 | ARG61 | LYS67 | ASN126 | LYS148 | GLU151 |
|---------|--------|--------|--------|--------|--------|--------|--------|
| Z-score | -5.630 | -5.118 | -6.984 | -6.269 | -5.307 | -6.039 | -6.209 |

И еще один значимый фактор – $B_{\text{фактор}}$ – отвечает за температурные колебания атомов. Он учитывает среднеквадратичное отклонение от положения равновесия для каждого атома. В норме $B_{\text{фактор}}$ лежит в диапазоне от 10 до 30. Как правило, более "свободные" атомы располагаются на поверхности глобулы рис. 6. Структура 4ZYЕ не исключение из этого правила.

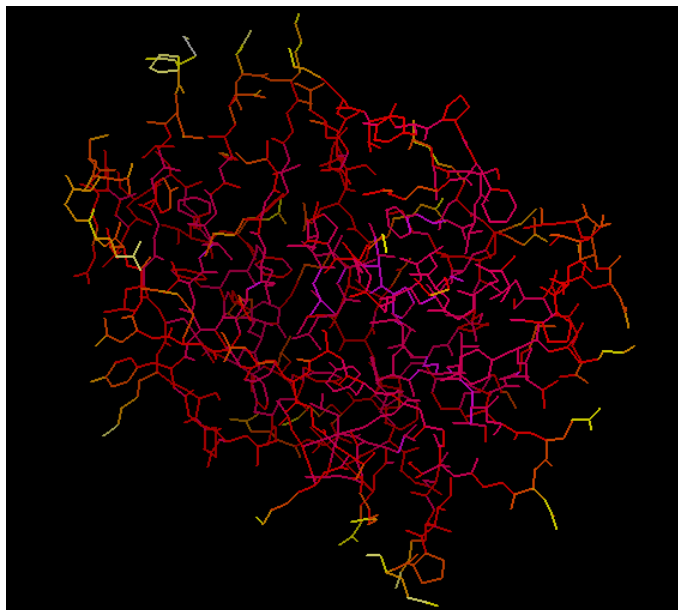


Рис. 6: Структура 4ZYE, раскрашенная по $B_{\text{фактор}}$. Получено с помощью MolProbity.

3 Маргинальные остатки

Таблица 2: Сводная таблица маргинальных остатков структуры 4ZYE.

| № | Остаток | Причина маргинальности |
|-----|---------|--|
| 1 | MET | Комфортность окружения; Пространственный $R_{\text{фактор}}$ |
| 8 | LYS | Комфортность окружения |
| 29 | CYS | Угол связи |
| 30 | ASP | Наложение Ван-дер-Ваальсовых радиусов атомов |
| 33 | GLU | Пространственный $R_{\text{фактор}}$ |
| 44 | GLU | Наложение Ван-дер-Ваальсовых радиусов атомов; Ротамер |
| 48 | LYS | Наложение Ван-дер-Ваальсовых радиусов атомов |
| 58 | ILE | Ротамер |
| 61 | ARG | Комфортность окружения |
| 67 | LYS | Комфортность окружения |
| 77 | LYS | Наложение Ван-дер-Ваальсовых радиусов атомов |
| 95 | ASP | Наложение Ван-дер-Ваальсовых радиусов атомов |
| 126 | ASN | Комфортность окружения |
| 148 | LYS | Комфортность окружения |
| 151 | GLU | Комфортность окружения |

Некоторые из маргинальных остатков можно рассмотреть подробнее. Они выбраны на основе указанных выше факторов с помощью сервисов EDS, WhatCheck и MolProbity. В сводной таблице сервиса MolProbity маргинальность остатка по тому или иному параметру обозначается красным цветом. Если же остаток находится на границе допустимой

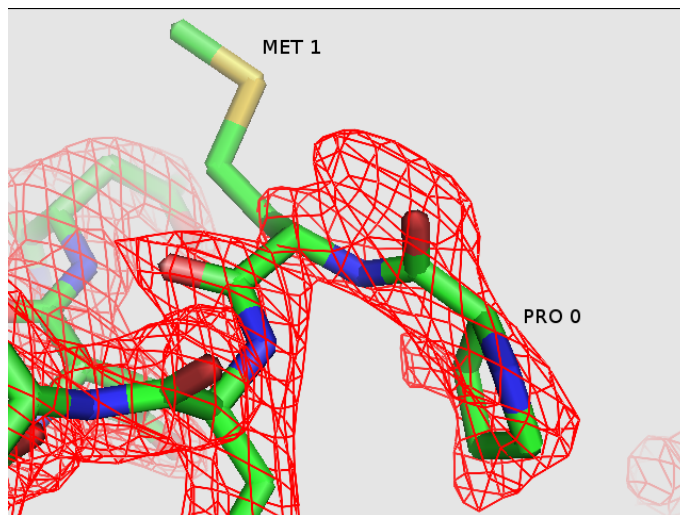


Рис. 7: Остаток MET 1, вписанный в электронную плотность на уровне $\sigma = 1$. Получено с помощью PyMol.

области, он помечен бледнокрасным. Далее будут рассматриваться только маргинальные остатки. Согласно MolProbity, маргинальных остатков по перекрытию атомов 5, плохих ротамеров – 2 и 1 по углу связи. По комфортности окружения в структуре 7 маргиналов (таблица 1), по RSR – 2(4). Всего в сумме – 15 остатков (таблица 2).

Из приведенных остатков наиболее интересными являются MET1, GLU44 (маргиналы сразу по двум параметрам), CYS29 (редкий тип маргинальности), GLU33 (пример плохого RSR не на концах цепи) и ARG61 (пример некомфортного окружения). Все случаи наложения не между атомами белка, поэтому не так интересны.

3.1 MET 1

Остаток MET1 расположен на поверхности глобулы следом за PRO0. Кроме низкого значения комфортности окружения и пространственного $R_{\text{фактор}}$, у него и у PRO0 еще высокое значение $B_{\text{фактор}} - B = 62.190$ и $B = 61.630$ соответственно, согласно сервису EDS.

Такое низкое значение Z-score комфортности окружения, видимо, вызвано тем, что боковая цепь метионина сильно экспонирована, что являет сильно энергетически невыгодным положением, так как она полностью состоит из неполярных атомов. Такое расположение не удастся объяснить молекулами соседних ячеек кристалла, остатки которых могли бы взаимодействовать с MET1, однако даже на расстоянии в 5\AA других остатков в кристалле нет.

Пространственный $R_{\text{фактор}}$, то есть соответствие модели электронной плотности, равен 0.275 (Z-score = 3.7271). Это хорошо видно на визуализации структуры (рис. 7), где даже на уровне $\sigma = 1$ нет никакой электронной плотности около всего остатка. То есть такое расположение маловероятно *in vivo* и совсем не соответствует электронной

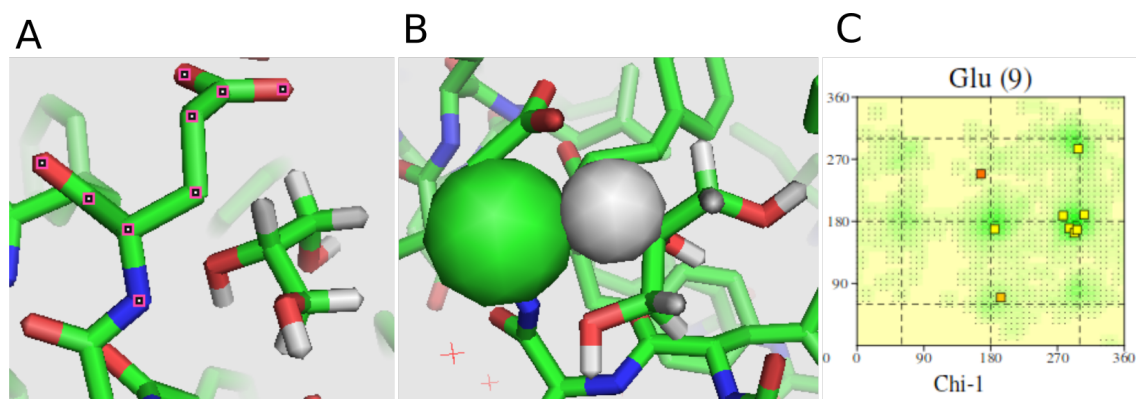


Рис. 8: Остаток GLU 44 в структуре 4ZYE. Получено с помощью PyMol и PROCHECK.

плотности, из чего можно сделать вывод, что это скорее не особенность белка, а ошибка в расшифровке структуры.

3.2 GLU 44

Остаток GLU44 (рис. 8 А) маргинален по двум параметрам: плохой ротамер и перекрытие Ван-дер-Ваальсовых радиусов атомов. Эти два параметра часто связаны между собой, так как неудачное определение углов боковой цепи остатка может повлечь за собой и неправильное расположение атомов относительно соседних групп. В данном случае перекрытие происходит между водородом Н2 лиганда и водородом НВ3 GLU44. На рис. 8 В показаны Ван-дер-Ваальсовы сферы соответствующего углерода и водорода лиганда, видно, что даже они пересекаются, а расстояние между центрами наложенных водородов равно 0.71\AA , согласно сервису MolProbity, что не может быть правдой.

Как ротамер этот остаток находится близко, но все же в недопустимой зоне рис. 8 С (красный квадратик) по углу $\chi-1$ по данным MolProbity этот угол составляет 167° . Заметим, что этот угол непосредственно связан с атомом СВ, водород при котором участвует в перекрытии.

Этот остаток не противоречит электронной плотности, однако она практически полностью отсутствует в этом месте на уровнях значимости больше 1.

Описанное выше говорит о том, что скорее всего маргинальность этого остатка не особенность белка, а ошибка в расшифровке структуры.

3.3 CYS 29

Остаток CYS29 был отмечен сервисом MolProbity как маргинальный по углу связи. Этот параметр, как и длина связи, очень жесткий, что говорит о серьезном нарушении.

Согласно полученной авторами структуре CYS29 образует дисульфидную связь с CYS31. Выглядит эта конструкция очень неустойчиво и не несет понятной биоло-

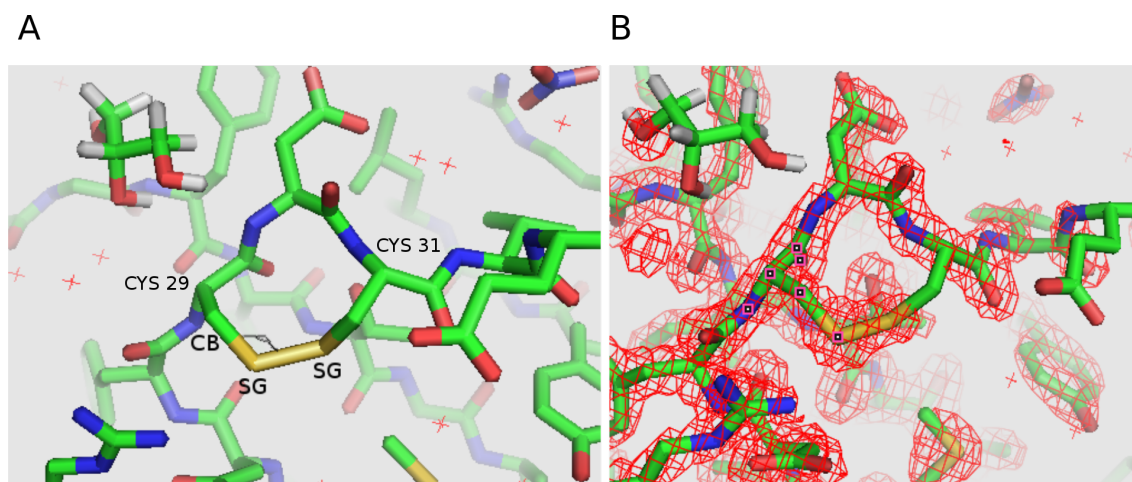


Рис. 9: Остаток CYS 29 и дисульфидная связь –CYS29–CYS31– в структуре 4ZYE. Получено с помощью PyMol.

гической роли. В белке дисульфидные связи редко бывают случайными и образуют элементы четвертичной структуры или стабилизируют третичную структуру и ее элементы.

В модели эта связь предсказалась с грубыми нарушениями именно угла с вершиной в атоме серы (рис. 9 А), угол составил 117.947° , что говорит не в пользу этого варианта. Остаток ASP30, участвующий в этом "кольце" тоже маргинальный (таблица 2), причиной чего может так же быть выбор авторов в пользу этого "кольца".

Тем не менее указанная конформация атомов очень хорошо ложится в электронную плотность (рис. 9 В), $\sigma = 2$. И еще одним аргументом могло быть то, что в белке часто все цистеины разбиваются на пары, образующих дисульфидные мостики, или все без одного, если их нечетное. В этом белке всего три цистеина: CYS29 и CYS30 в "кольце" и CYS119 в активном центре участвует в переносе алкильной группы, то есть CYS29 и CYS30 больше не с кем образовывать дисульфидный мостик.

Аргументы против такой конформации кажутся более весомыми, поэтому, вероятно, маргинальность этого остатка не особенность белка, а ошибка в расшифровке структуры.

3.4 GLU 33

Остаток GLU33 отмечен сервисом EDS как маргинальный по RSR. Действительно, он не соответствует электронной плотности даже на уровне $\sigma = 1$ (рис. 10), а $RSR = 0,287$ ($Z\text{-score} = 2.066$). Этот остаток отрицательно заряжен, поэтому действительно должен быть обращен от поверхности глобулы в раствор, а не внутрь глобулы.

Однако заметим, что этот остаток находится в непосредственной близости к "кольцу", образованному дисульфидным мостиком –CYS29–CYS31–, описанному вы-

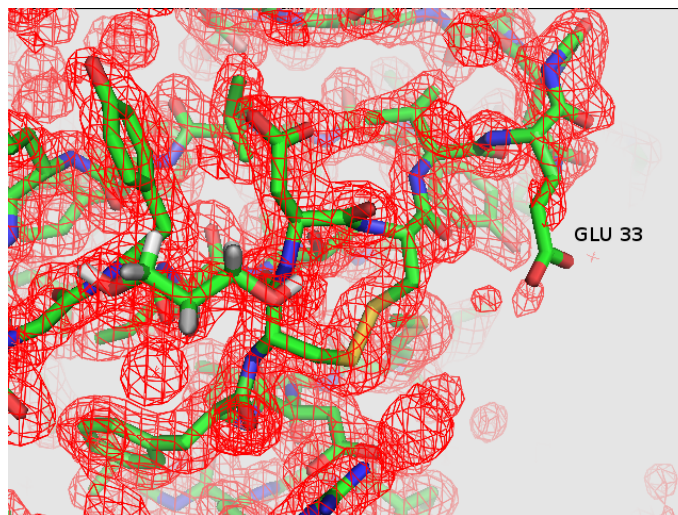


Рис. 10: Остаток GLU 33 в структуре 4ZYE. Получено с помощью PyMol.

ше, и обращен к нему. Если бы его конформацию определили бы несколько в другую сторону, что было бы допустимо, так как электронная плотности там нет, CYS31 и CYS30 могли бы расположиться иначе.

GLU33 находится слишком близко к "кольцу" цистеинов, чтобы быть особенностью этого белка, поэтому его маргинальность, предположительно, не особенность белка, а ошибка в расшифровке структуры.

3.5 ARG 61

Остаток ARG61 располагается на поверхности глобулы. Он положительно заряжен и образует водородные связи с водой. Рядом с ним располагается только один гидрофобный остаток – MET80, который не может создать некомфортное окружение для аргинина на поверхности. В этом месте нет контакта с молекулами белка других ячеек кристалла. При этом у него очень низкое значение $Z = -6.984$ по данным WhatCheck.

По рис. 11 видно, что этот остаток не очень хорошо вписывается в электронную плотность, $\sigma = 1.5$. Однако других несоответствий обнаружить не удалось.

Не очень понятно, что именно программа посчитала как некомфортное окружение, поэтому этот остаток, предположительно, не ошибка в расшифровке структуры, а особенность белка.

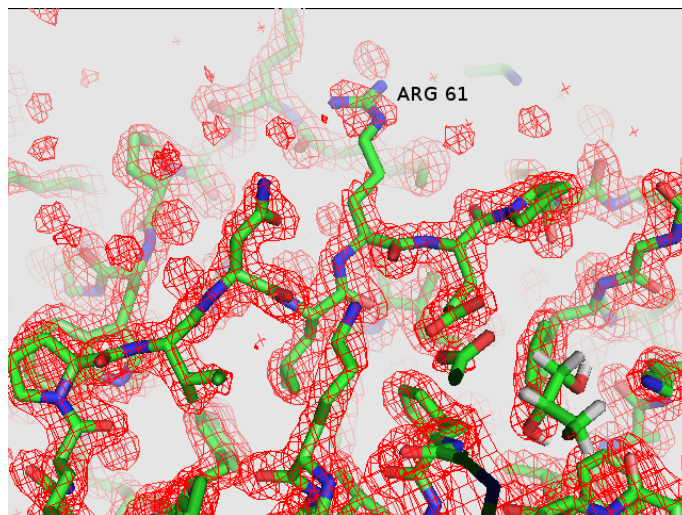


Рис. 11: Остаток ARG 61 в структуре 4ZYE. Получено с помощью PyMol.

4 Сравнение модели из PDB с моделью из PDB_redo

Сервис PDB_redo по тем же экспериментальным данным построил новую оптимизированную модель структуры 4ZYE. Для сравнения модели 4ZYE и 4ZYE_redo был использован сервис MolProbity. Наглядно результат представлен в таблице 3.

Таблица 3: Сравнительная таблица качества структуры 4ZYE и 4ZYE_redo.

| Параметр | 4ZYE | | 4ZYE_redo | |
|--|--------|--------|-----------|--------|
| Clashscore (перекрывание атомов) | 4.82 | | 0.8 | |
| Плохие ротамеры | 2 | 1.50% | 3 | 2.26% |
| Предпочтительные ротамеры | 129 | 96.99% | 125 | 93.98% |
| Outliers на карте Рамачандрана | 0 | 0.00% | 0 | 0.00% |
| Предпочтительные на карте Рамачандрана | 144 | 96.00% | 143 | 95.33% |
| Плохие связи | 0/1252 | 0.00% | 7/1252 | 0.56 |
| Плохие углы | 2/1681 | 0.12% | 2/1681 | 0.12% |

Почти все параметры, кроме перекрывания атомов, которые сервис изменил, ухудшились. Появились маргинальные остатки по длине связи, что является достаточно жестким параметром. Тем не менее сервис MolProbity считает свой собственный комплексный из указанных параметров скор, который для модели 4ZYE составил 1.66, что хуже, чем 1.34 для модели 4ZYE_redo. Вероятно, большой вклад вносит избавление от перекрываний атомов.

В целом нельзя сказать, что сервису удалось улучшить модель структуры.

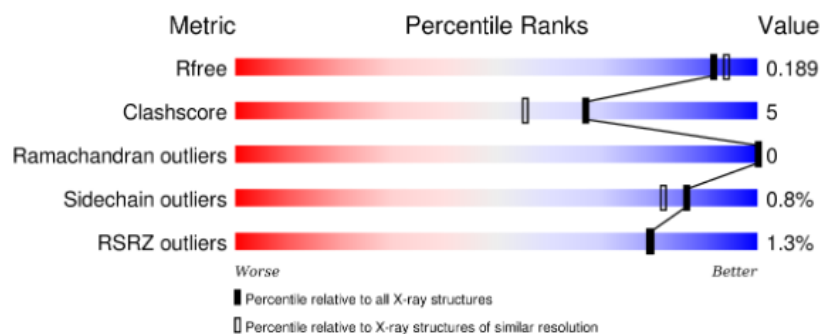


Рис. 12: Сравнение структуры 4ZYE со средними значениями для структур такого разрешения. Получено с помощью PDB.

ЗАКЛЮЧЕНИЕ

Структура Об-метилгуанин метилтрансферазы (PDB ID 4ZYE) расшифрована очень хорошо. Об этом можно судить по графику, полученному с помощью сервиса PDB, (рис. 12). При очень неплохом разрешении (1.85\AA) эта структура по всем параметрам, кроме R_{free} , оказывается лучше среднего по структурам такого разрешения. Тем не менее нельзя сказать, что структура переоптимизирована, так как разница между R_{free} и $R_{фактор}$ составляет $18.2\% - 16.1\% = 2.1\% < 5\%$.

СПИСОК ЛИТЕРАТУРЫ

- [1] Ami Sabharwal и Mark R Middleton. “Exploiting the role of O 6-methylguanine-DNA-methyltransferase (MGMT) in cancer therapy”. В: *Current opinion in pharmacology* 6.4 (2006), с. 355—363.
- [2] Douglas S Daniels и др. “Active and alkylated human AGT structures: a novel zinc site, inhibitor and extrahelical base binding”. В: *The EMBO journal* 19.7 (2000), с. 1719—1730.
- [3] Sreenivas KANUGULA, Karina GOODTZOVA и E Anthony PEGG. “Probing of conformational changes in human O6-alkylguanine-DNA alkyl transferase protein in its alkylated and DNA-bound states by limited proteolysis”. В: *Biochemical Journal* 329.3 (1998), с. 545—550.
- [4] Joseph J Rasimas и др. “Active-site alkylation destabilizes human O6-alkylguanine DNA alkyltransferase”. В: *Protein science* 13.1 (2004), с. 301—305.
- [5] Giuseppe Perugino и др. “Structure-function relationships governing activity and stability of a DNA alkylation damage repair thermostable protein”. В: *Nucleic acids research* 43.18 (2015), с. 8801—8816.
- [6] Airlie J McCoy и др. “Phaser crystallographic software”. В: *Journal of applied crystallography* 40.4 (2007), с. 658—674.
- [7] Paul D Adams и др. “PHENIX: a comprehensive Python-based system for macromolecular structure solution”. В: *Acta Crystallographica Section D: Biological Crystallography* 66.2 (2010), с. 213—221.