

ФАКУЛЬТЕТ БИОИНЖЕНЕРИИ И БИОИНФОРМАТИКИ

**Качество расшифровки структуры белка УДФ-N-ацетилглюкозамин
ацилтрансферазы (PDB ID: 2AQ9) методом рентгеноструктурного
анализа**

*Отчет по структурной биоинформатике студентки 4-го курса Зареченской
Анастасии Сергеевны*

Преподаватели:

А. В. Алексеевский

С. А. Спиринов

В. Ю. Лунин

Аннотация

В данной работе была произведена оценка качества расшифровки структуры белка УДФ-N-ацетилглюкозамин ацилтрансферазы (PDB ID: 2AQ9) методом рентгеноструктурного анализа с помощью информации на сайте PDB, а также сервисов MolProbity, EDS и других. Производилось как оценивание параметров качества структуры в целом, так и для отдельных остатков. Особое внимание было уделено поиску и анализу маргинальных остатков по следующим основным критериям: расположение на карте Рамачандрана, значение RSR, ротамеры. Итогом работы является вывод о качестве расшифровки данной структуры.

Введение

Бактериальные инфекции — основная причина смертности на протяжении всей истории человечества. Никакие катастрофы не уносили столько жизней, сколько эпидемии чумы, холеры и туберкулеза. Даже во время войны люди чаще гибли от кишечных инфекций и других бактериальными заболеваниями, чем на поле боя. Коренным образом ситуация изменилась только в XX веке. Гениальное открытие антибактериальной активности гриба *Penicillium notatum* Александром Флемингом [1] и разработка методов промышленного получения пенициллина Хоурдом Флори и Эрнстом Чейни [2] дали человечеству в руки мощное оружие для борьбы с бактериями. Обнаруженный вскоре Зелманом Ваксманом стрептомицин [3] открыл «золотой век» поиска новых антибиотиков. За довольно короткий период было найдено множество разнообразных молекул с антибактериальной активностью. В конце 1970х эффективность поиска новых антибиотиков значительно снизилась, а бактерии, наоборот, начали демонстрировать стремительный рост распространения устойчивости к лекарствам. Прошло более семидесяти лет после массового внедрения антибиотиков, а люди, даже в развитых странах, продолжают умирать от бактериальных инфекций. В связи с этим поиск новых антибиотиков остается важной практической задачей.

Рассмотренный в данном отчете белок УДФ-N-ацетилглюкозамин ацилтрансфераза (далее – LpxA) представляет собой важный объект для изучения. LpxA является катализатором первой стадии пути биосинтеза липида А – эндотоксина, закоривающего липополисахаридный слой в наружной мембране грамотрицательных бактерий. Наличие данного белка является необходимым условием для поддержания жизнедеятельности бактериальной клетки. Кроме того, известно, что белки, принимающие участие в биосинтезе липида А, уникальны для бактерий, что позволяет минимизировать проявления побочных эффектов у эукариот при проведении антибактериальной терапии [4]. Вышеперечисленное позволяет рассмотреть белок LpxA в качестве потенциальной мишени для антибиотиков. Кристаллическая структура комплекса белка LpxA с пептидом 920, качеству расшифровки которой посвящен отчет, получена с целью открыть новые подходы к дизайну и синтезу ингибиторов LpxA – эффективных антибактериальных препаратов.

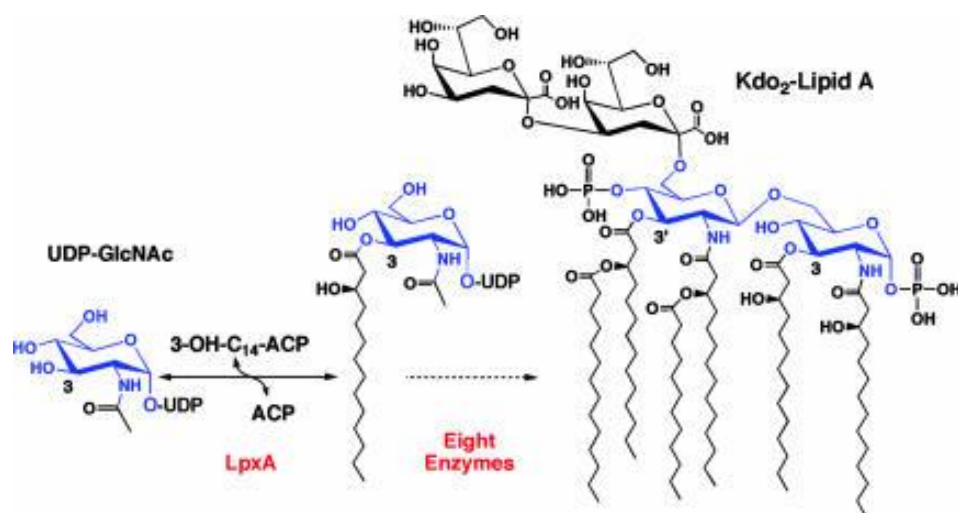


Рисунок 1. Функция LpxA в процессе биосинтеза липида А. Белок переносит R-3-гидроксиацильную часть с R-3-гидроксиацил - АЦП на УДФ-N-Ацетилглюкозамин ($K_p = 0.01$).

Стоит отметить, что ранее кристаллическая структура данного белка была представлена в отсутствие субстратов или ингибиторов (PDB ID: 1LXA,

разрешение 2.6 Å) [5]. Эта модель позволила заключить, что белок LpxA – гомотример. Структуру отличает наличие 24 полных и шести частичных повторов гексапептидов. Три непрерывных повторения гексапептидов (18 аминокислотных остатков) складываются в одну катушку левозакрученной β -спирали. Такая укладка является характерной для многих бактериальных ацил- и ацетилтрансфераз.

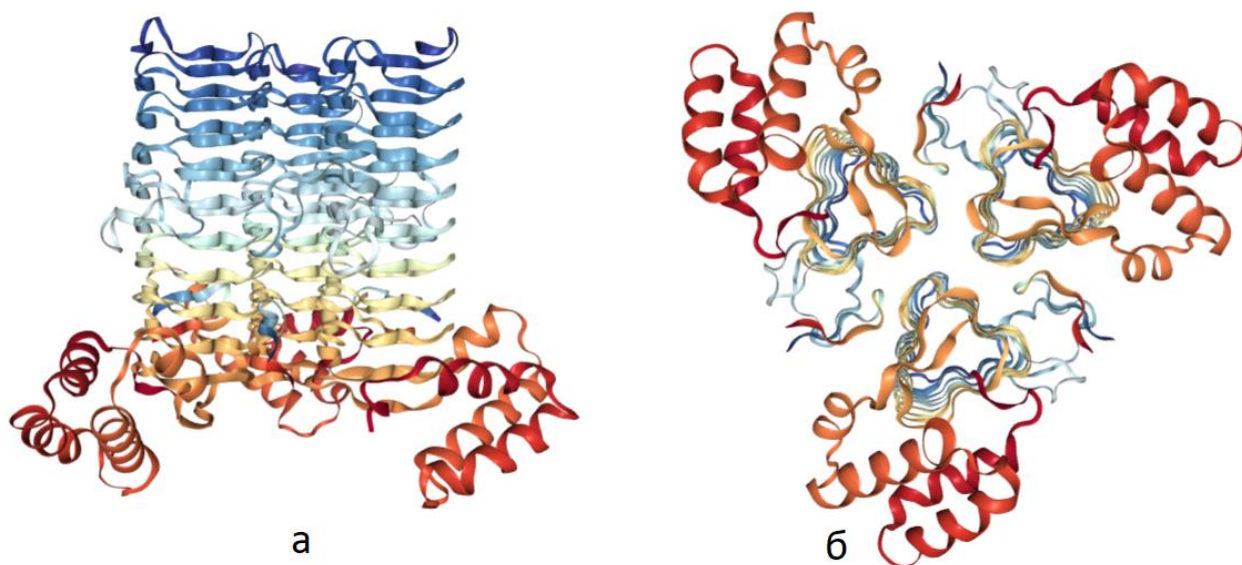


Рисунок 2. Пространственная структура белка lpxA а – вид сбоку; б – вид снизу

В 2006 году была представлена структура 2AQ9, комплекс lpxA с пентадекапептидом 920 - ингибитором lpxA ($K_i = 50$ nM), который в процессе реакции конкурирует с ацил-АЦП. Пептид 920 был обнаружен ранее с помощью метода фагового дисплея.

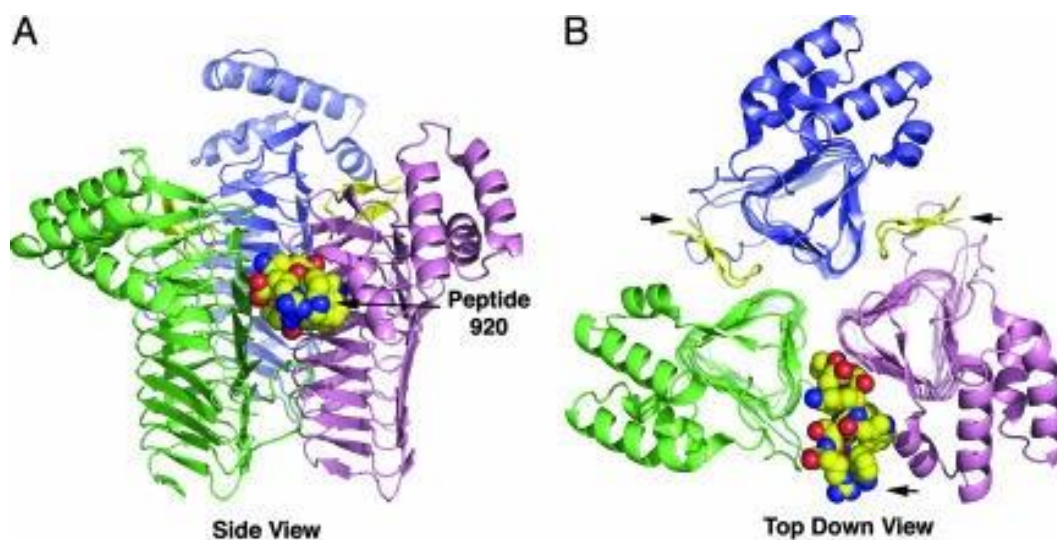


Рисунок 3. Кристаллическая структура комплекса LrxA с пептидом 920 с разрешением 1,8 Å. (A) Вид сбоку. Субъединицы гомотримера показаны фиолетовым зеленым и синим цветами. Пептид 920 выделен желтым. (B) Вид комплекса снизу.

Результаты и обсуждение

Общая информация о модели

Структура 2AQ9 была выложена на сайт PDB в 2006 году группой ученых из университета Дьюка, Северная Каролина, США [4]. Структура была определена с разрешением 1,8 Å. В ходе эксперимента было получено 262329 структурных факторов. Для решения фазовой проблемы был использован метод молекулярного замещения с помощью программы AMoRE. В качестве гомолога использовалась структура 1LXA. Асимметрическая единица – одна субъединица белка - состоит из 262 а.о. Биологическая единица включает в себя 3 идентичные субъединицы белка (гомотример). Всего в биологической единице 2270 атомов (с учетом молекул воды и других гетероатомов). В таблице 1 представлена краткая характеристика структуры с основной информацией об эксперименте и составе кристалла.

Таблица 1. Основные параметры структуры 2AQ9

| Параметр | Значение |
|--|---|
| Год | 2006 |
| Авторы | Williams, A.H., Immormino, R.M., Gewirth, D.T., Raetz, C.R. |
| Метод решения фазовой проблемы | Молекулярное замещение |
| Количество измеренных рефлексов | 28 264 |
| Полнота набора структурных факторов | 91.3% |
| Диапазон расширений структурных факторов | 18.00 – 1.80 Å |
| Разрешение | 1,8 Å |
| Тип симметрии | P 2 ₁ 3 |
| Параметры кристаллической ячейки | a = 96.734 Å, b = 96.734 Å, c = 96.734 Å |

Значения индикаторов качества модели в целом.

В данном разделе рассмотрены основные индикаторы качества модели: R фактор, R-free, а также критерии, характеризующие геометрические параметры структуры, такие как карта Рамачандрана, перекрывание Ван-дер-Ваальсовых радиусов атомов, длины и углы связей, ротамеры. Значения индикаторов R и R-free, а также количество маргинальных остатков по некоторым критериям представлены в таблице 2.

Таблица 2. Значения основных индикаторов качества структуры 2AQ9.

| Параметр | Значение | |
|--|---|----------------------------|
| R-фактор | 0.194 (0.226 по DCC) | |
| R-free | 0.211 (0.237 по DCC) | |
| Доля структурных факторов, отобранных для расчета R-free | 8.86% | |
| Доля маргинальных остатков по некоторым критериям | | |
| Критерий | Количество маргиналов (по PDB full report) | Доля маргиналов (%) |
| Углы связей | 1 из 2973 | 0,034 |
| Длины связей | 1 из 2198 | 0,046 |
| RSR | 6 из 277 | 2,2 |
| Карта Рамачандрана | 0 | 0 |
| Ротамеры | 0 | 0 |
| Перекрывания | 20 из 4837 | 0,41 |

На рисунке 4 представлена характеристика структуры 2AQ9 по отношению ко всем структурам из PDB (полученных методом рентгеноструктурного анализа), а также по отношению к структурам близкого разрешения. Можно сделать вывод, что 2AQ9 расшифрована лучше, чем структуры такого же разрешения и все структуры из PDB.

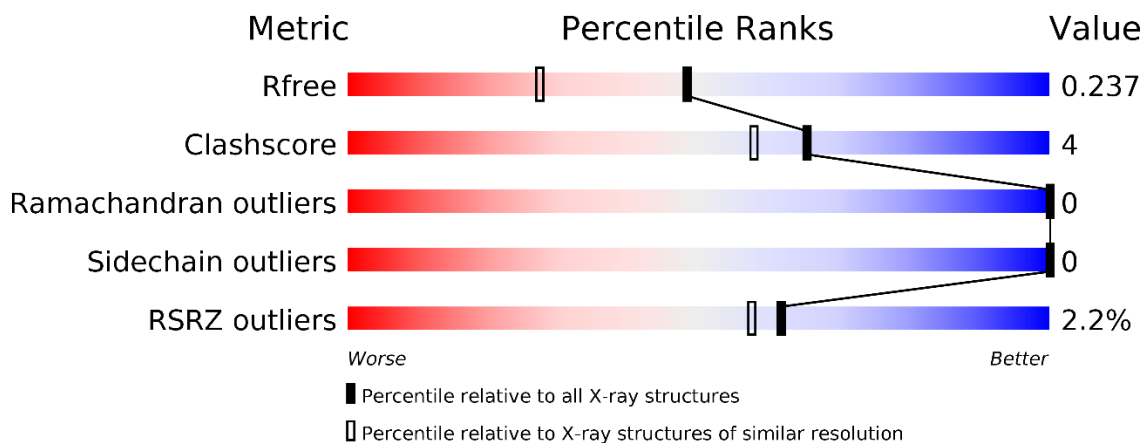


Рисунок 4. Характеристика расшифровки структуры 2AQ9. Справа указаны значения основных индикаторов качества: R-free, критерий перекрытия Ван-дер-Ваальсовых радиусов атомов (Clashscore), доля маргиналов по карте Рамачандрана (Ramachandran outliers), по положению боковой цепи (Sidechain outliers) и по RSR (RSRZ outliers). Шкала отражает распределение значений этих параметров по сравнению со всеми структурами PDB (черные отметки) или со структурами схожего разрешения (белые отметки). Градиент цвета показывает, насколько параметры данной структуры лучше (синий) или хуже (красный) средних значений.

Summary statistics

| | | | | |
|-------------------|---|-----------|---|--------------------------------|
| All-Atom Contacts | Clashscore, all atoms: | 5.04 | 96 th percentile* (N=837, 1.80Å ± 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 | 221 | 98.66% | Goal: >98% |
| | Ramachandran outliers | 0 | 0.00% | Goal: <0.05% |
| | Ramachandran favored | 261 | 96.67% | Goal: >98% |
| | MolProbity score [^] | 1.54 | 92 nd percentile* (N=11444, 1.80Å ± 0.25Å) | |
| | Cβ deviations >0.25Å | 0 | 0.00% | Goal: 0 |
| | Bad bonds: | 11 / 2294 | 0.48% | Goal: 0% |
| Bad angles: | 2 / 3104 | 0.06% | Goal: <0.1% | |
| Peptide Omegas | Cis Prolines: | 0 / 10 | 0.00% | Expected: ≤1 per chain, or ≤5% |
| | Cis nonProlines: | 1 / 273 | 0.37% | Goal: <0.05% |

Рисунок 5. Выдача сервиса MolProbity [6]. Результат выдачи показывает хорошие показатели оценок качества данной структуры.

Рассмотрим основные индикаторы качества расшифровки данной структуры подробнее

- Р фактор** Этот параметр описывает, насколько хорошо полученная модель соответствует экспериментальным данным. По результатам авторов структуры R фактор составляет 0,194 (при оптимизации DCC в PDB R фактор составил 0,226). Хорошими значениями для этого параметра считаются показатели R меньше 0,25, поэтому можно сказать, что данная структура достаточно хорошо соответствует экспериментальным данным.

- **R-free** R-free является внутренним контролем, позволяющим оценить, не была ли модель переоптимизирована. По результатам авторов R-free равен 0,211 (DCC - 0,237). Нормальными значениями считаются R-free меньше 0,2. Однако R-free для этой модели не сильно превышает порог, а разница между R и R-free составляет 8%, что меньше 10%, поэтому можно считать, что структура не была переоптимизирована.
- **Clashscore** Этот параметр отражает число стерических перекрытий (Ван-дер-Ваальсовы радиусы соседних атомов перекрываются более чем на 0,4 Å) в расчете на 1000 атомов. При анализе с помощью сервиса MolProbity, который искусственно вносит в структуру атомы водорода, было получено значение clashscore 5,04. Это значение является 96-ым перцентилем (100-ый перцентиль соответствует наилучшему качеству по данному индикатору среди моделей с близким разрешением, 0-ой перцентиль - наихудшему). Таким образом, хоть в структуре и присутствует некоторое количество перекрытий между атомами, в целом модель проявляет себя хорошо.
- **Ramachandran outliers** На рисунке 4 представлена карта Рамачандрана для всего белка, построенная программой MolProbity. Она позволяет оценить, насколько хорошо конформация основной цепи белка в структуре соответствует предпочитаемому положению, основываясь на значениях торсионных углов ϕ и ψ . Можно заметить, что в запрещенной области не оказалось ни одного остатка. Несколько остатков находятся в разрешенной, но не предпочитаемой области

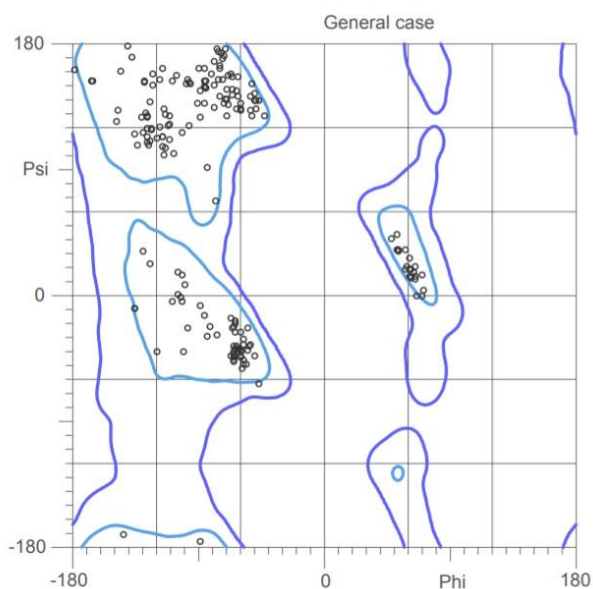


Рисунок 6. Карта Рамачандрана для целого белка IpxA (для асимметрической единицы из структуры 2AQ9). Построена сервисом MolProbity.

Также были получены карты для отдельных остатков (Gly, Pro, Val и Ile), но все эти аминокислоты в данной структуре находятся в предпочитаемой области, поэтому карты для них не приведены в работе. Далее карта Рамачандрана была построена при помощи сервиса EDS [7].

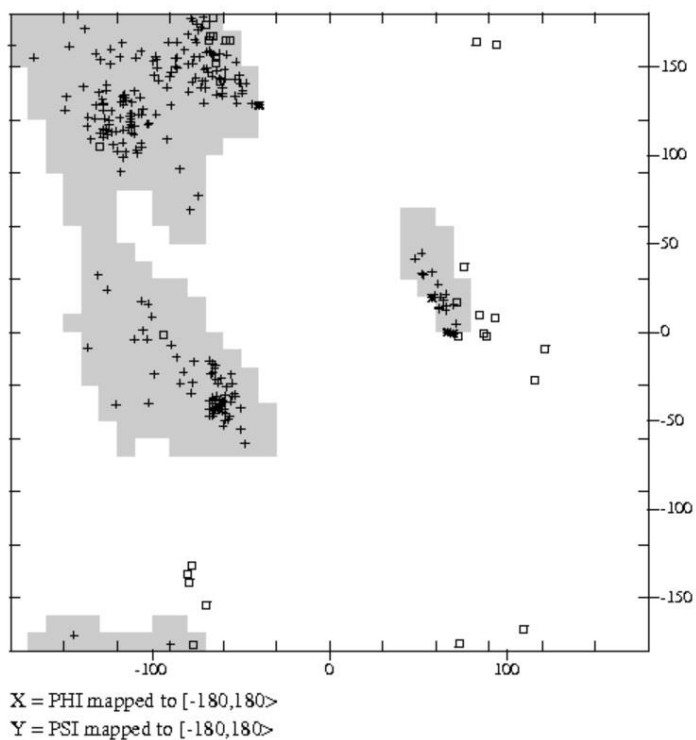


Рисунок 6. Карта Рамачандрана для целого белка IpxA (для асимметрической единицы из структуры 2AQ9). Построена сервисом EDS.

EDS выявила 4 маргинальных остатка (1,7%).

- Sidechain outliers** (ротамеры) Данный параметр учитывает конформацию боковых цепей аминокислот. Ротамерами называются боковые цепи в предпочитаемых для данного остатка положениях, которые описываются наборами допустимых значений торсионных углов χ . По оценке MolProbity таких остатков должно быть меньше 0,3%. В данной структуре ротамеров не обнаружено. Таким образом, в модели 2AQ9 все остатки находятся в предпочитаемой для них конформации.
- RSR outliers** Параметр RSR характеризует, насколько хорошо каждый аминокислотный остаток соответствует экспериментальным данным, то есть, насколько хорошо остаток в модели вписывается в электронную плотность. Этот индикатор рассчитывается для отдельных остатков. Обычно оценивается Z-score (RSRZ) - насколько RSR для остатка в данной модели отличается от среднего значения RSR по структурам из PDB близкого разрешения. Если Z-score больше 2, считается, что остаток маргинальный по RSR и не очень хорошо описывается экспериментальными данными. По данному индикатору структура 2AQ9 имеет нормальное качество - 2,2% маргинальных остатков.

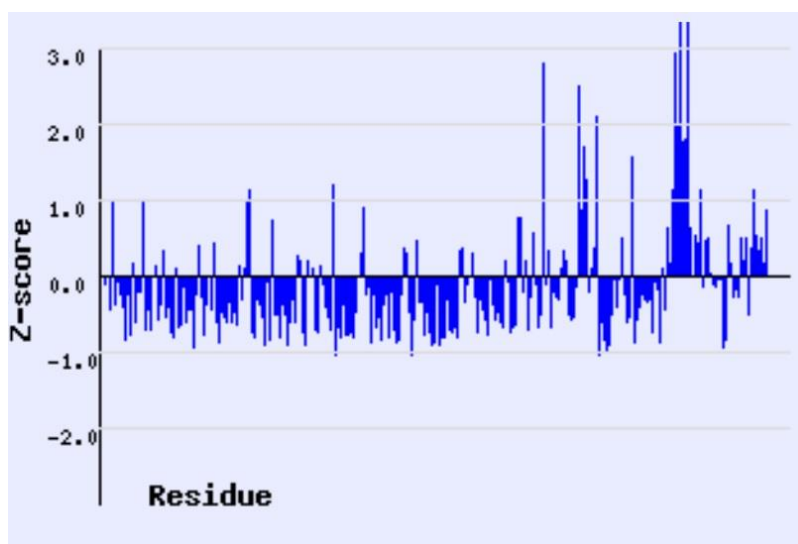


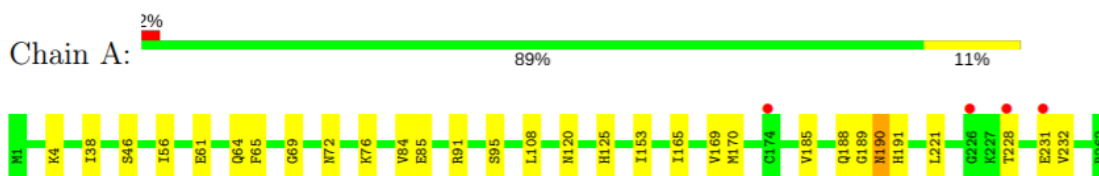
Рисунок 6. RSRZ остатков (сервис EDS) Высокие пики, где Z-score > 2 соответствуют маргинальным остаткам по RSR.

- **Bond angles or lengths** Эти параметры позволяют оценить, не нарушены ли в структуре длины химических связей или валентные углы. По данным PDB full report в модели 2AQ9 длина 1 (0,046%) связи признана неоптимальной и всего лишь 1 (0,034%) из углов связей также неоптимален. По результатам MolProbity найдено 2 неоптимальных угла (0,06%). К тому же, MolProbity отмечает 11 (0,48%) связей недопустимой длины.

Некоторые маргинальные остатки

Обратимся к подробному отчету о валидации структуры 2AQ9, который можно найти на сайте PDB в поле Full report. В отчете имеются схемы, представленные на рис. 7 в качестве примера, которые показывают, какие остатки обладают отклонениями в своей геометрии по различным показателям и позволяют оценить маргинальность каждого остатка по совокупности показателей. Зеленый цвет остатка в такой схеме означает отсутствие маргинальности по всем показателям, желтый – маргинальность по одному, оранжевый – по двум, красный – по трем и более критериям. Серыми выделены остатки, не представленные в структуре.

- Molecule 1: Acyl-[acyl-carrier-protein]-UDP-N-acetylglucosamine O-acyltransferase



- Molecule 2: peptide inhibitor

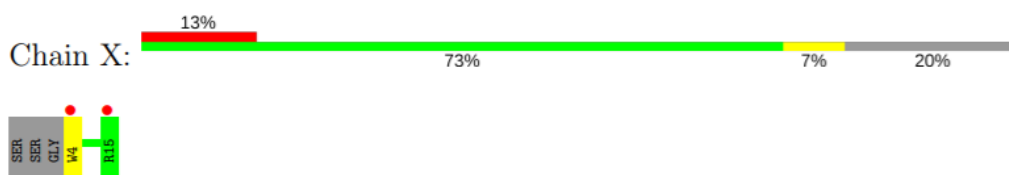


Рисунок 7. Карта маргинальности остатков в структуре 2AQ9

В таблице 3 приведен список десяти маргинальных остатков с указанием критериев, по которым остаток оказывается маргинальным. Далее эти остатки разобраны более подробно. (Для рассмотрения также были взяты маргиналы из выдачи EDS)

Таблица 3. Некоторые маргинальные остатки из структуры 2AQ9.

| | Остаток | Критерий маргинальности |
|----|----------------|---|
| 1 | Glu 231A | RSR (0,367), RSRZ (3,585) |
| 2 | Thr 228A | RSR (0,264), RSRZ (4,143) |
| 3 | Gln 188A | RSR (0,289), RSRZ (2,509) |
| 4 | Gly 226A | RSR (0,243), RSRZ (2,943) |
| 5 | Glu 69A/Val84A | Перекрытие Ван-дер-Ваальсовых радиусов (Clash overlap = 0,5 Å) |
| 6 | Phe 65A/Ser95A | Перекрытие Ван-дер-Ваальсовых радиусов (Clash overlap = 0,41 Å) |
| 7 | Asn 23A | Карта Рамачандрана (EDS) |
| 8 | Ile119A | Карта Рамачандрана (EDS) |
| 9 | Phe 150A | Карта Рамачандрана (EDS) |
| 10 | Phe 162A | Карта Рамачандрана (EDS) |

Анализ некоторых маргинальных остатков
Маргиналы по признаку RSR, RSRZ

1

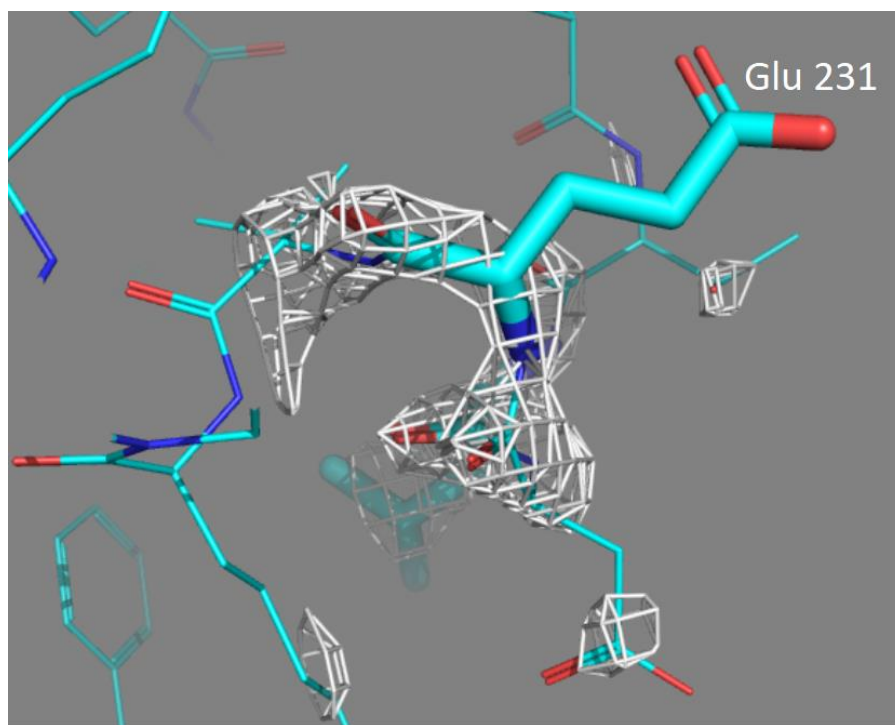


Рисунок 8. Остаток Glu 231 и его окружение. Электронная плотность показана на уровне подрезки $1,5\sigma$.

2

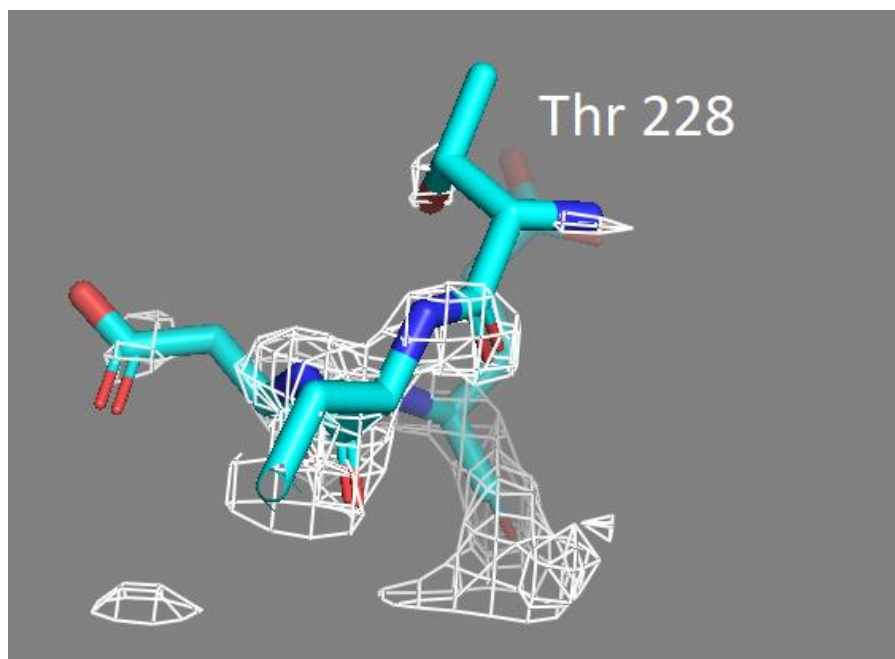


Рисунок 9. Остаток Thr 228 и его окружение. Электронная плотность показана на уровне подрезки $1,5\sigma$.

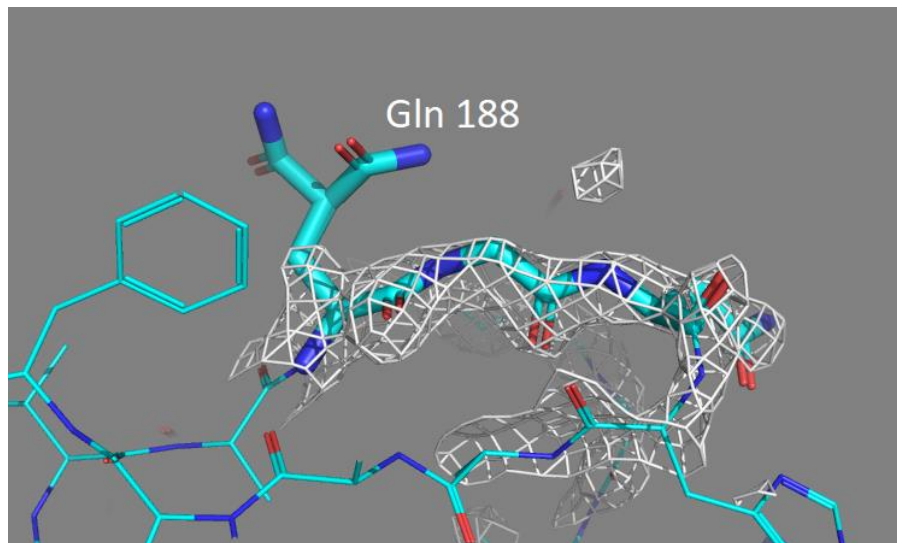


Рисунок 10. Остаток Gln 188 и его окружение. Электронная плотность показана на уровне подрезки 1,5 σ .

На рисунках 8, 9 и 10 заметно, что рассмотренные остатки не вписываются в электронную плотность. Причиной маргинальности является скорее всего ошибка расшифровки.

Маргиналы по расположению на карте Рамачандрана

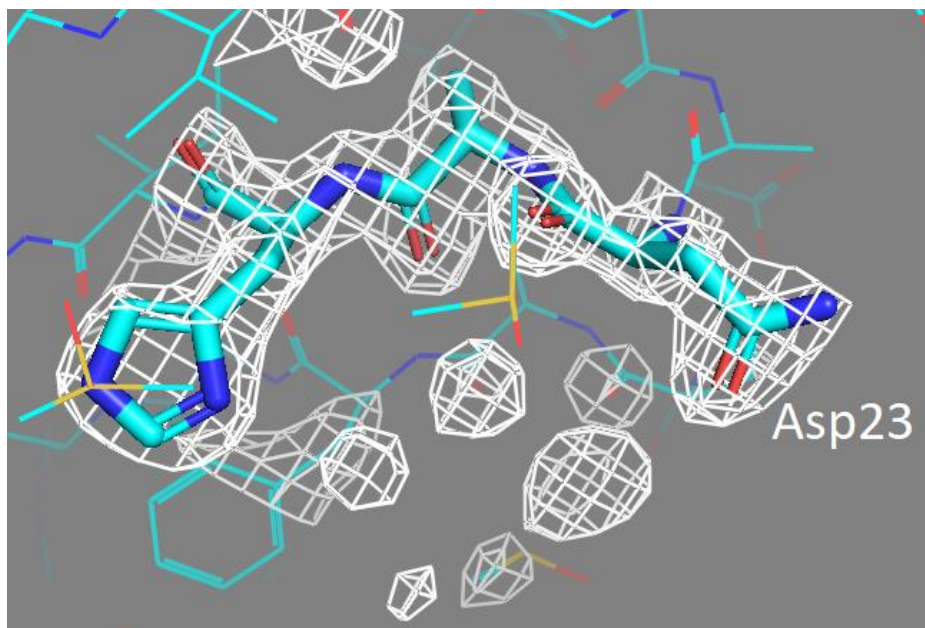


Рисунок 11. Остаток Asp23 и его окружение. Электронная плотность показана на уровне подрезки 1,5 σ .

Маргинальный признак – положение в недопустимой области карты Рамачандрана (сервис EDS). Такое возможно, если остаток расположен близко с другими остатками белка, например, при контакте между субъединицами, или ионами металла, входящих в структуру белка в качестве кофакторов. Скорее всего данный случай не является ошибкой расшифровки, так как остаток хорошо описывается электронной плотностью.

5

Маргиналы по наличию перекрытия Ван-дер-Ваальсовых радиусов

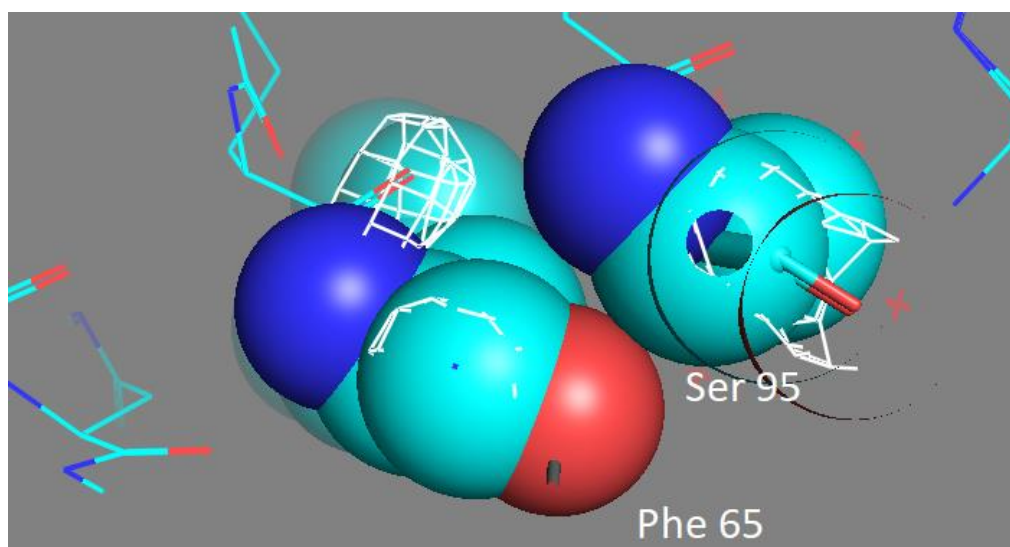


Рисунок 12. Перекрытие остатков Ser 95 и Phe 65

Пример перекрытия атомов - атомы Н и О из Ser 95 и Phe65. Между этими атомами должна образовываться водородная связь, но длина этой связи получается 2,2 Å, что несколько меньше характерной для водородной связи длины. Из-за того, что атомы расположены так близко, наблюдается небольшое перекрытие их Ван-дер-Ваальсовых радиусов на 0,41 Å (плохими считаются значения больше 0,4 Å).

Стоит отметить, что оба остатка вписываются в электронную плотность, и данный участок структуры расшифрован нормально.

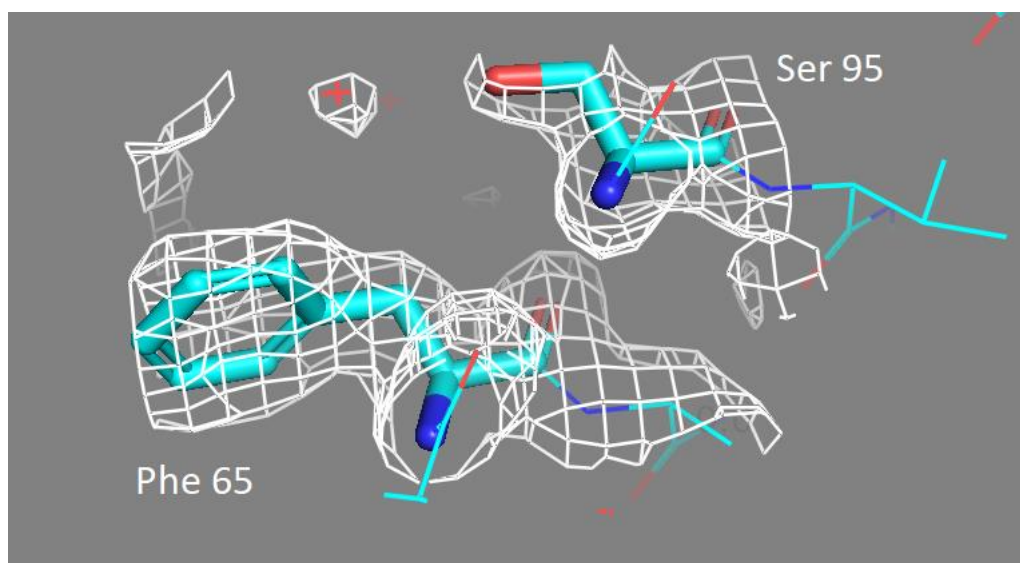


Рисунок 13. Остатки Ser95 и Phe 65. Электронная плотность показана на уровне подрезки 1σ .

Сравнение со структурой из PDB_redo

Таблица 4. Сравнение основных индикаторов качества и количества маргинальных остатков (по результатам MolProbity) для исходной PDB-структуры 2AQ9 и для аналогичной структуры из PDB_redo.

| Критерий | PDB | PDB_redo |
|--------------------|------------------------------------|-------------------------------------|
| R-фактор | 0.194 (0.226 по DCC) | 0.1784 |
| R-free | 0.211 (0.237 по DCC) | 0.2190 |
| Карта Рамачандрана | 0 | 0 |
| Валентные углы | 0 | 2 |
| Длины связей | 44 | 11 |
| Ротамеры | 0 | 1 |
| Clashscore | 5,04 (96 ^{ой} перцентиль) | 2,39 (99 ^{ый} перцентиль) |
| MolProbity Score | 1,54(92 ^{ой} перцентиль) | 1,02 (100 ^{ый} перцентиль) |

| | | | | |
|-------------------|---|-----------|---|---|
| All-Atom Contacts | Clashscore, all atoms: | 2.39 | 99 th percentile* (N=792, 1.68Å ± 0.25Å) | |
| | Clashscore is the number of serious steric overlaps (> 0.4 Å) per 1000 atoms. | | | |
| Protein Geometry | Poor rotamers | 1 | 0.44% | Goal: <0.3% |
| | Favored rotamers | 222 | 97.37% | Goal: >98% |
| | Ramachandran outliers | 0 | 0.00% | Goal: <0.05% |
| | Ramachandran favored | 265 | 98.15% | Goal: >98% |
| | MolProbity score [^] | 1.02 | | 100 th percentile* (N=8853, 1.68Å ± 0.25Å) |
| | Cβ deviations >0.25Å | 0 | 0.00% | Goal: 0 |
| | Bad bonds: | 44 / 2310 | 1.90% | Goal: 0% |
| | Bad angles: | 0 / 3122 | 0.00% | Goal: <0.1% |
| Peptide Omegas | Cis Prolines: | 0 / 10 | 0.00% | Expected: ≤1 per chain, or ≤5% |
| | Cis nonProlines: | 1 / 273 | 0.37% | Goal: <0.05% |

Рисунок14. Выдача сервиса MolProbity для структуры из PDB_redo

Из таблицы 4 видно, что структура улучшилась. Некоторые параметры стали хуже, но незначительно. При наложении структуры совпадают (практически идентичны)

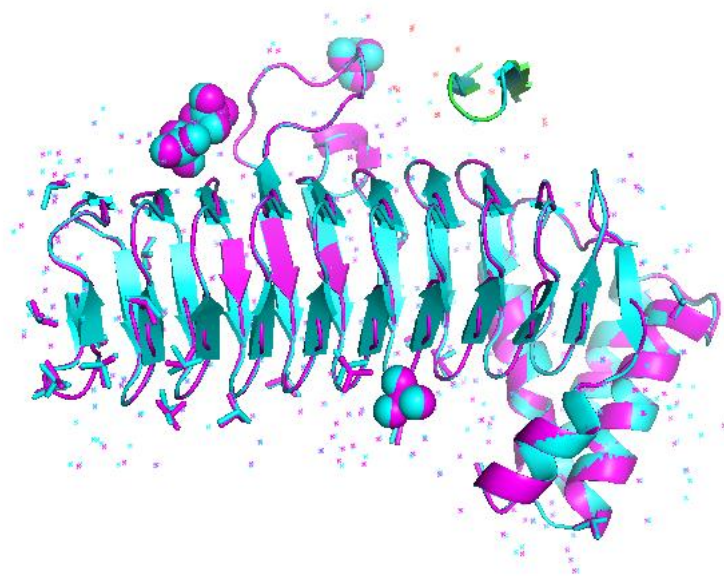


Рисунок 15. Наложение исходной структуры 2AQ9 (cyan) и структуры PDB_redo (magenta).

Заключение

Качество расшифровки структуры 2AQ9 белка LpxA из организма E.Coli очень хорошее. Индикаторы качества выше, чем у PDB-структур с близким разрешением. Анализ PDB_redo показал, что расшифровка структуры очень близка к оптимальной. В структуре есть маргинальные остатки, но рассмотренные нами являются следствием скорее ошибок расшифровки, чем функциональными особенностями.

Ссылки

1. Fleming A. On the Antibacterial Action of Cultures of a Penicillium, with Special Reference to Their Use in the Isolation of B. Influenzae. // British Journal of Experimental Pathology. 1929. Т. 10. № 3. — С. 226-236
2. Clark R.W. The life of Ernst Chain : penicillin and beyond. — New York: St. Martin's Press, 1985.
3. Comroe J.H. Pay Dirt - Story of Streptomycin .1. From Waksman to Waksman // American Review of Respiratory Disease. 1978. Т. 117. № 4. — С. 773-781.
4. Williams, A.H., Immormino, R.M., Gewirth, D.T., Raetz, C.R. Structure of UDP-N-acetylglucosamine acyltransferase with a bound antibacterial pentadecapeptide. (2006) Proc.Natl.Acad.Sci.Usa 103: 10877-10882
5. Raetz C. R. H., Roderick S. L. Science. 1995;270:997–1000.
6. <http://molprobit.biochem.duke.edu/index.php>
7. <http://eds.bmc.uu.se/eds/>