

Отчёт по качеству PCA расшифровки структуры пероксидазы-каталазы *Mycobacterium tuberculosis*

студентки 402 группы Быковой Дарьи

Аннотация

В работе проведён анализ качества расшифровки структуры пероксидазы-каталазы *Mycobacterium tuberculosis*. Рассмотрены общие характеристики модели, а также отдельные маргинальные по некоторым показателям остатки (выбросы на карте Рамачандрана, остатки с большим показателям RSR и т.д.). Проведено сравнение структуры из PDB со структурой из PDB_redo.

Введение

Для своего отчёта я выбрала структуру пероксидазы-каталазы *Mycobacterium tuberculosis* с pdb id 1SJ2. Данный фермент способен катализировать окисление разнообразных субстратов и, в частности, вовлечён в активацию популярного противотуберкулёзного препарата изониазида, то есть путём окисления превращает изониазид в действующее вещество [1]. Белок кодируется геном katG, и мутации в некоторых его сайтах приводят к устойчивости бактерии к изониазиду [2]. Структура, обсуждаемая в данной работе, была получена в 2004 году и стала первой для данного белка. Получение 3D структуры позволило лучше разобраться в механизме катализируемой реакции, а также предсказать сайт связывания изониазида и открыло возможности для интерпретации мутаций, приводящих к устойчивости к антибиотику.

Результаты и обсуждение

Общая информация о модели

Структура получена методом рентгеноструктурного анализа в 2004 году (авторы Bertrand et al., ссылка на статью дана во введении) и имеет разрешение 2,41 Å, полнота при этом составляет 90,3 %. Наименьшее разрешение набора полученных в ходе расшифровки гармоник (resolution range high) равно 2,41 Å, наибольшее (resolution range low) - 23.71 Å. Для решения фазовой проблемы авторы использовали метод молекулярного замещения, вспомогательная структура имеет pdb id 1MWV (структура того же фермента из другой бактерии). Число измеренных рефлексов - 61822. Группа симметрии кристалла - P 42 21 2. Параметры кристаллографической ячейки приведены в таблице 1.

Таблица 1. Параметры кристаллографической ячейки

Длина (Å)	Угол (°)
a = 150.33	$\alpha = 90$
b = 150.33	$\beta = 90$
c = 154.281	$\gamma = 90$

Ассиметрическая единица и функциональная форма для данного белка совпадают, белок представляет собой димер, в котором субъединицы расположены зеркально относительно друг друга.

Значения индикаторов качества модели

R-фактор модели равен 0,211, R-free - 0,268. Таким образом, R-free ненамного больше, чем R-фактор, значит, модель не переоптимизирована.

При помощи сервиса MolProbity [3] в структуру добавлены водороды и проанализирована геометрия всех остатков.

Clashscore, то есть число недопустимых наложений атомов на 1000 для данной структуры, составляет 17,02, что меньше, чем в 74% структур, и, следовательно, вполне допустимо. В структуре найдено 27 маргинальных остатков, то есть таких, конформация которых находится в недопустимых областях на картах Рамачандрана, что составляет 1,89 % от всех остатков. Таким образом, маргиналов значительно больше, чем рекомендуется (должно быть примерно 0,05 % при достаточно хорошем разрешении), некоторые из маргинальных остатков подробно рассмотрены далее. Вместе с тем, 92,24% остатков находятся в предпочтительных областях на картах Рамачандрана, что меньше, чем ожидалось бы в идеальном случае (98%). Кроме того, найден 1 C β -атом с неприемлемым отклонением от ожидаемого положения, а также 0,21% ковалентных связей и 0,17% валентных углов существенно отклоняются от теории. 32 остатка (2,83%) имеют маргинальные по отклонению от ротамеров боковые цепи, в то время как в идеальном случае таких остатков должно быть не более 0,3%. Однако, несмотря на перечисленные недостатки, MolProbity score для данной структуры составляет 2,55, что лучше, чем у 66% структур с примерно таким же разрешением.

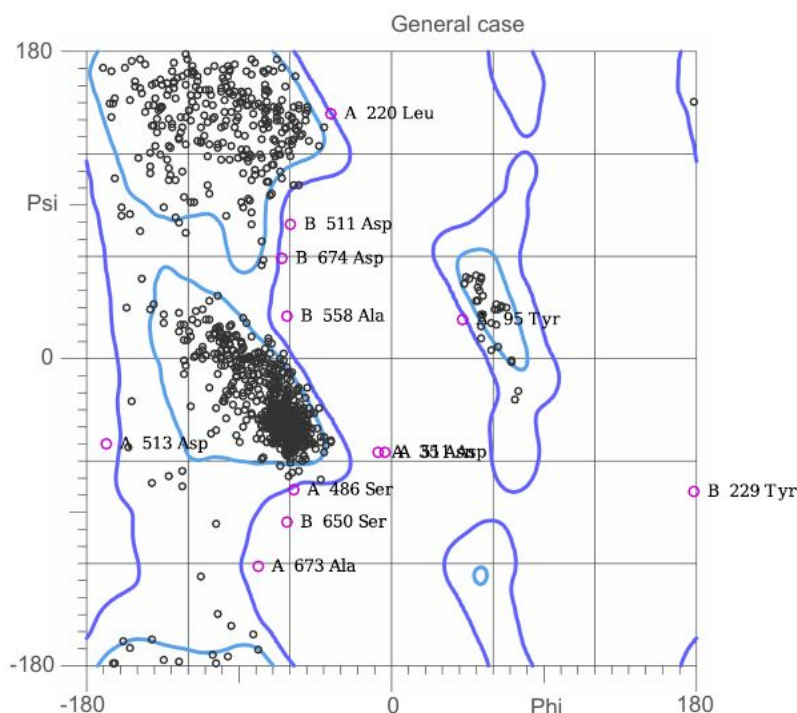


Рис. 1. Остатки на карте Рамачандрана. Изображение получено при помощи сервиса MolProbity

Если обратиться к оценке качества структуры самим PDB (рис.1), то видно, что 1sj2 по параметрам Clashscore и Ramachandran outliers значительно проигрывает другим структурам с тем же разрешением, имеет немного худшее, чем в среднем, значение Rfree и чуть лучшие показатели Sidechain outliers и RSRZ outliers. Заметим, что оценка Clashscore относительно других структур в самом PDB сильно отличается от таковой в MolProbity. Вероятно, оценка PDB более универсальна.

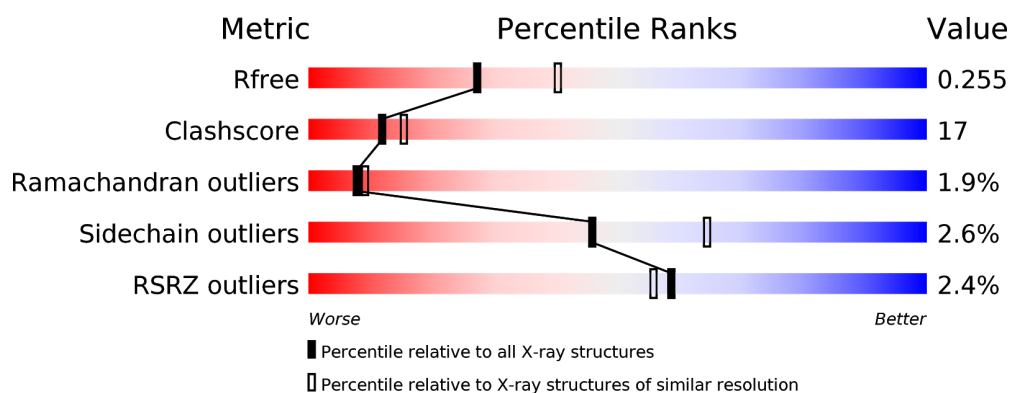


Рис. 2. Оценка качества структуры, приведённая на странице PDB.

Маргинальные остатки

При помощи серверов WHAT IF [4], а также Electron Density Server (EDS) [5] были выявлены остатки, маргинальные по некоторым параметрам. EDS для каждого остатка выдаёт RSR, то есть величину, показывающую, насколько хорошо данный остаток вписан в электронную плотность. Чем больше RSR, тем менее похоже, что остаток вписан правильно. Значение RSR больше 20% говорит о серьёзном расхождении. Из набора программ WHAT IF были использованы следующие: Coarse Packing Quality Control, Anomalous bond lengths, Anomalous bond angles, Highly likely peptide flips. Сервис Coarse Packing Quality Control оценивает комфортность окружения для каждого остатка, сервисы Anomalous bond lengths и Anomalous bond angles находят остатки с подозрительными (значительно отклоняющимися от нормальных) длинами связей или углами между связями, сервис Highly likely peptide flips выдаёт список остатков, которые, вероятно, неправильно повернуты в структуре (транс- заменён на цис- или наоборот или неправильно ориентирован остов).

Итак, при помощи описанных программ были получены списки маргинальных остатков. Кроме того, была использована информация от MolProbity об остатках-выбросах на карте Рамачандрана. Далее, нас более всего интересовали (1) остатки, маргинальные сразу по нескольким показателям и (2) остатки с экстремальными значениями показателей (очень большой RSR, очень низкий Z-score при оценке окружения и т.д.). В таблице 2 приведены некоторые маргинальные остатки с показателями, по которым они отбирались.

Таблица 2. Некоторые маргинальные остатки

Остаток	Показатель качества
GLY 305 B	Возможно, пептидная цепь перевёрнута, окружение, RSR
TYR 229 A	Длина и угол связи, окружение
GLU 651 B	Возможно, пептидная цепь перевёрнута, окружение, RSR
TYR 95 A	Карта Рамачандрана, окружение
ASP 511 B	Окружение, RSR, карта Рамачандрана
PRO 501 B	Карта Рамачандрана, возможно, надо заменить транс на цис
LYS 433 A	Окружение, RSR
ASN 535 B	RSR
ALA 532 B	Карта Рамачандрана, RSR, окружение
HIS 25 B	Окружение, RSR

Ниже подробно рассмотрены 5 маргинальных остатков.

1) GLY 305 B

Глицин вместе с окружением 5 Å и картой электронной плотности на уровне подрезки 2 изображён на рис. 3 (выделен розовым). Как мне кажется, данный остаток неплохо вписывается в электронную плотность, хотя значение RSR для него довольно высокое - 0,239 и видно, что остов несколько выбивается из облака ЭП. Я бы не стала переворачивать его, как это предлагает программа, потому что в таком случае остаток будет хуже вписываться в ЭП. Что касается окружения, то Z-score, выданный Coarse Packing Quality Control, равен -3,178, то есть довольно низок. Я думаю, что в данном случае это связано с положением остатка в белке: он находится в петле на периферии и водородную связь может образовать только с водой, которая, кстати, есть поблизости (на расстоянии 3,1 Å от N). Итак, я считаю, что положение данного остатка восстановлено приемлемо.

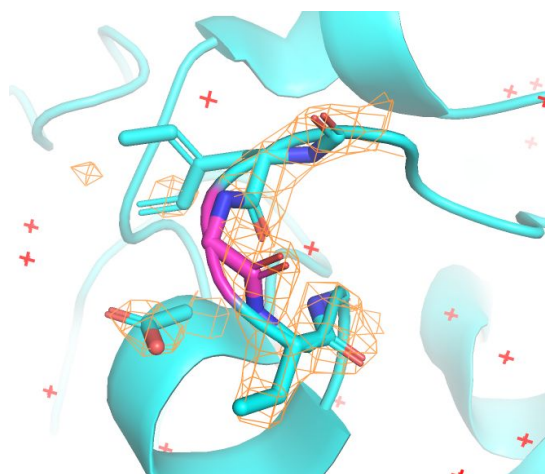


Рис. 3. GLY 305 B и его окружение.

2) TYR 229 A

Данный остаток со своим окружением показан на рис. 4 (уровень подрезки mesh - 2). На первый взгляд кажется, что происходит нечто странное: с ароматическим кольцом тирозина ковалентно связаны метионин и триптофан. Однако оказывается, что такая конфигурация нормальна для каталитического сайта данной каталазы [6]. Z-score тирозина в данном окружении -2,288, но, вероятно, он низкий из-за описанных необычных взаимодействий. Кроме того, мы наблюдаем стэкинг-взаимодействие с PHE 252 A. Программы Anomalous bond lengths и Anomalous bond angles выявляют подозрительно малую длину связи между атомом N тирозина и C предыдущего остатка (1,05 Å, Z-score = -13,8) и подозрительно маленький угол, образованный атомами остова O-C-N (O и C относятся к предыдущему остатку ILE, угол равен 112,02°, Z-score = -6,9). Судя по карте электронной плотности, атомы остова тирозина и его неароматическая часть действительно вписаны достаточно произвольно, так что здесь, возможно, стоило бы внести корректировки.

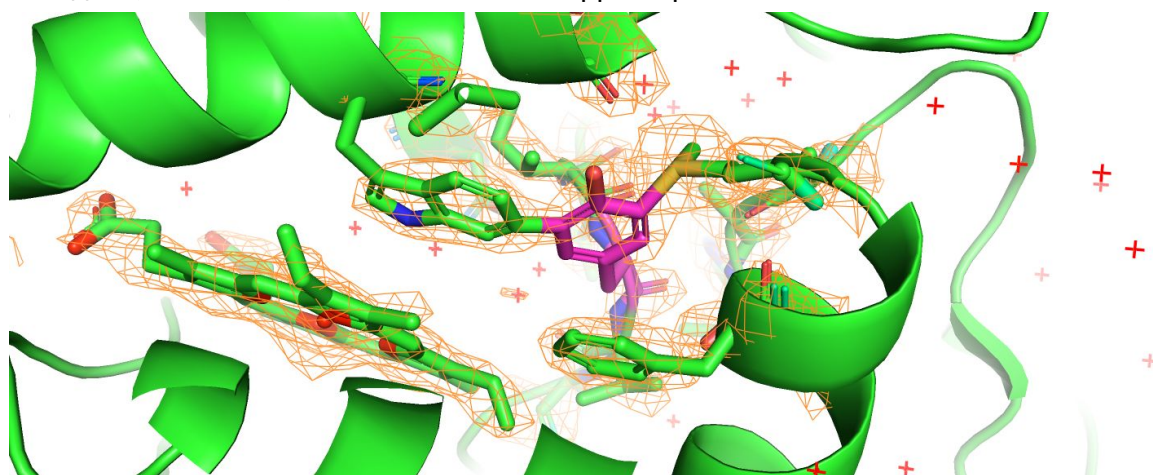


Рис. 4. TYR 229 A и его окружение.

3) ASN 535 B

Остаток выбран всего по одному параметру: его RSR составляет 0,39, он входит в топ-5 самых непокрытых ЭП остатков. Однако, как можно видеть из рис. 5, он просто очень подвижный.

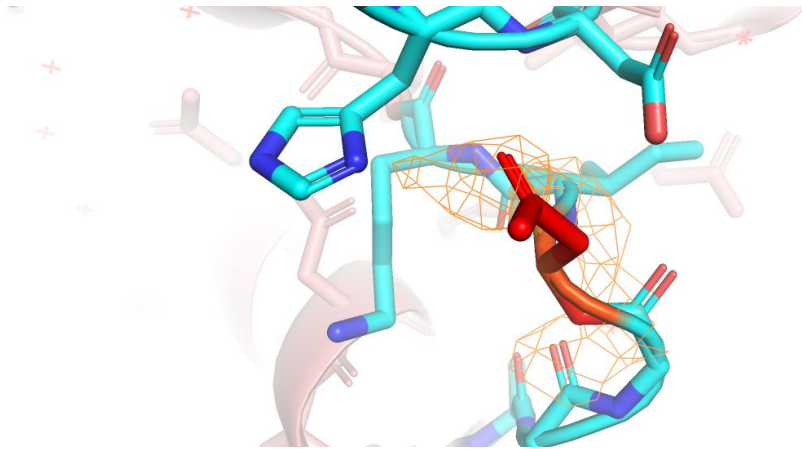


Рис. 5. ASN 535 В и его окружение. Атомы остатка покрашены по В-фактору. Электронная плотность показана на уровне подрезки 1.

4) ASP 511 В

RSR данного остатка составляет 0,362, что близко к максимуму среди всех остатков белка. Z-score окружения равен -4,18. Кроме того, ASP 511 попадает в неоптимальную область карты Рамачандрана (-60,3, 79,3). Из рис. 6 видно, что, даже на уровне подрезки ЭП 1, остаток плохо вписывается в ЭП. Видимо, это снова связано с высокой подвижностью данного остатка. Действительно, если покрасить его атомы по В-фактору, то окажется, что у атомов карбоксильной группы он весьма высокий. Что касается окружения, отметим, что поблизости располагаются ещё два аспартата. Впрочем, если посмотреть чуть шире, то можно найти недалеко лизин и два аргинина, которые должны стабилизировать отрицательный заряд.

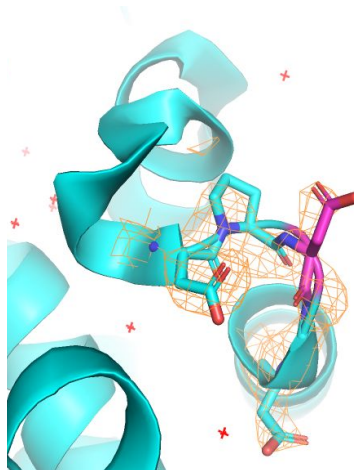


Рис. 6. ASP 511 В и его окружение.

5) PRO 501 В

Программой Highly likely peptide flips предлагается изменить конформацию данного остатка с транс- на цис-. Кроме того, данный остаток помечен как несоответствующий карте Рамачандрана. Мне кажется, транс-конформация более соответствует ЭП, чем цис (рис. 7).

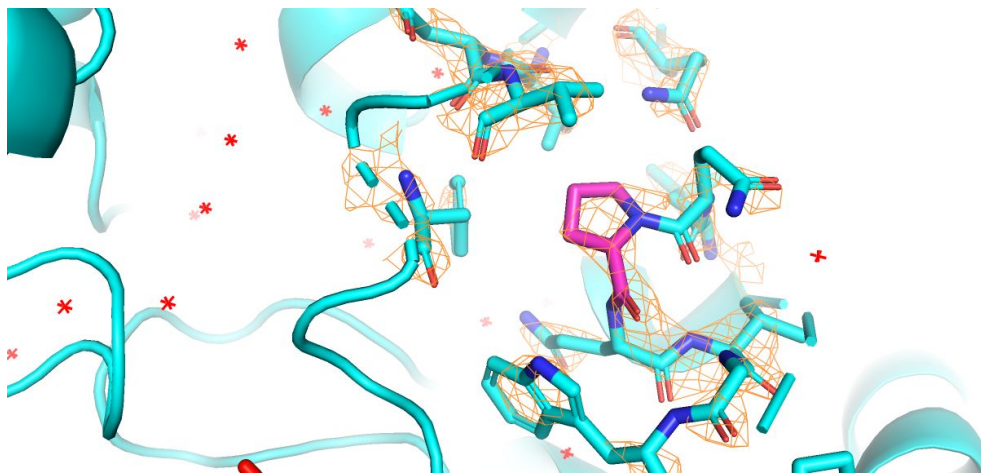


Рис. 7. PRO 501 B и его окружение.

Сравнение модели из PDB с моделью из PDB_redo

PDB_redo [7] удалось снизить Rfree до 22,7% (пересчитанный по изначальным данным Rfree составил 29%), а также уменьшить число маргинальных остатков. На рис. 8 можно видеть, куда переместились подозрительные остатки на карте Рамачандрана (оранжевые точки соответствуют структуре PDB_redo, голубые - PDB).

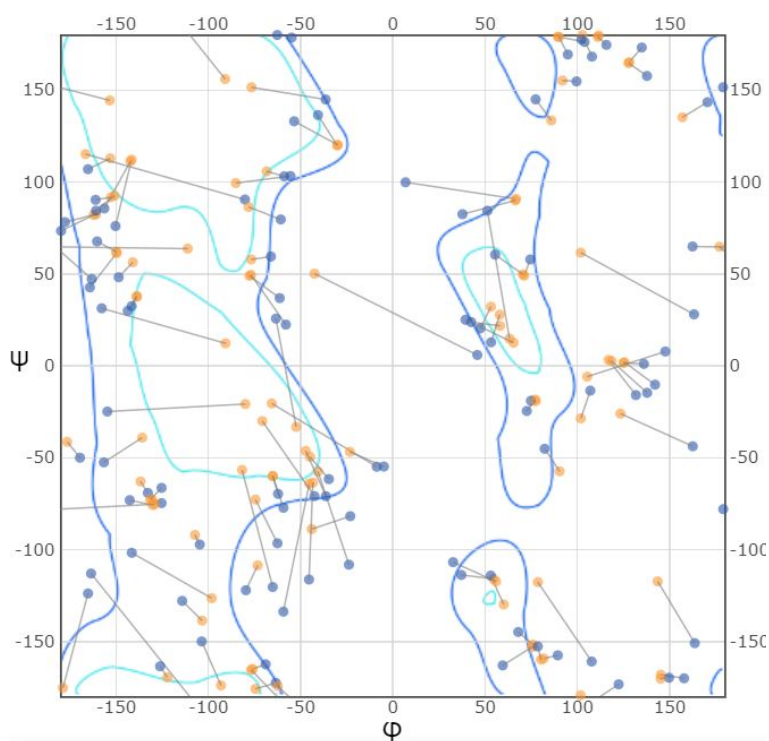


Рис. 8. Положение остатков на карте Рамачандрана в PDB (голубой) и PDB_redo (оранжевый)

Согласно отчёту PDB_redo, для 12 остатков были выбраны другие ротамеры, было удалено 46 молекул воды, для 76 остатков было улучшено соответствие электронной плотности.

Хотя изменения для некоторых остатков значительные, при наложении структур глобальных перестроек не наблюдается. Рассмотрим несколько маргинальных остатков и проследим, как изменилось их положение в структуре PDB_redo.

GLY 305 В остался на месте. TYR 229 А слегка повернулся, и новое положение более выгодно для стэкинг-взаимодействия (рис. 9). Положение атомов остова при этом практически не изменилось.



Рис. 9. TYR 229 А. Структура из PDB_redo покрашена бирюзовым, изначальная структура - розовым.

ASP 511 В немного изменил своё положение, но в ЭП вписывается по-прежнему плохо (рис.11). Ещё одно подтверждение, здесь дело не в плохой расшифровке, а в подвижности остатка.

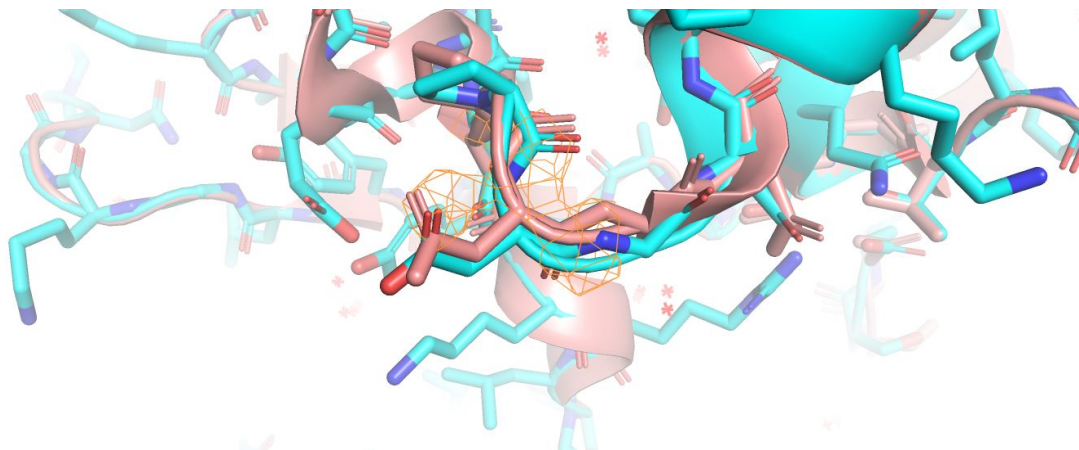


Рис. 11. ASP 511 В. Структура из PDB_redo покрашена бирюзовым, изначальная структура - розовым. Уровень подрезки ЭП - 1.

Конформация PRO 501 В изменена не была. Любопытно, как PDB_redo преобразовал положение ASN 535 В (рис. 12). Теперь заметно, что аспарагин может образовывать водородную связь с азотом в кольце гистидина. Мы видим, что облачко электронной плотности чуть вытянуто в этом направлении, так что, скорее всего, предложенная конформация более удачна.

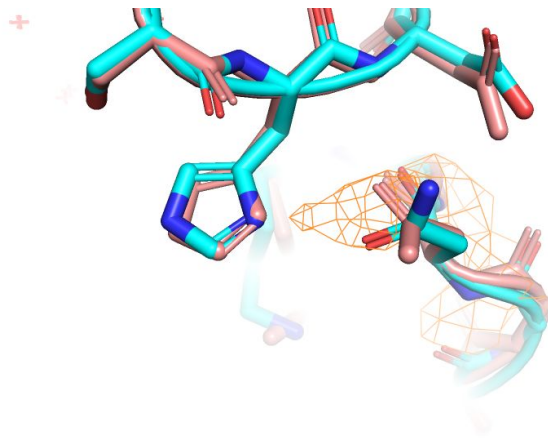


Рис. 12. ASN 535 B. Структура из PDB_redo покрашена бирюзовым, изначальная структура - розовым. Уровень подрезки ЭП - 1.

Заключение

Структура имеет ряд недостатков, неучтённых мной в первой версии данного отчёта. Во-первых, это аномально большое число остатков из запрещённых областей карты Рамачандрана. Во-вторых, высокий Clashscore согласно PDB. Обе эти проблемы можно частично решить автоматически, даже не переделывая эксперимент. Возможно, 16 лет назад, когда авторы публиковали данную структуру, стандарты качества были другими и возможностей для бенчмаркинга было меньше. Большинство рассмотренных проблемных остатков имеют достаточно высокую подвижность и, как следствие, достаточно размытое облако электронной плотности. Следовательно, вписывание остатков в ЭП в таких случаях происходило достаточно произвольно и по какой-то причине авторы упустили возможность в таких спорных случаях прибегнуть к знаниям о геометрии остатков. Вероятно, для них компромисс между вписыванием как можно точнее в ЭП и биологией разрешился в пользу лучшего соответствия полученной ЭП (о чём говорят показатели качества PDB: RSRZ outliers лучше, чем в среднем). В местах, где сигнал в эксперименте был не очень внятным, такой подход привёл к ошибкам. Коль скоро сейчас есть ресурс PDB_redo, для извлечения биологических фактов в данном случае лучше использовать структуру из него.

Список литературы

- 1) Bertrand et al. Crystal Structure of Mycobacterium tuberculosis Catalase-Peroxidase. doi: 10.1074/jbc.M402382200. September 10, 2004. The Journal of Biological Chemistry.
- 2) Walker et al. Whole-genome sequencing for prediction of Mycobacterium tuberculosis drug susceptibility and resistance: a retrospective cohort study. doi: 10.1016/S1473-3099(15)00062-6. June 23, 2015. Lancet Infect Dis.
- 3) MolProbity, <http://molprobity.biochem.duke.edu/index.php>
- 4) WHAT IF, <https://swift.cmbi.umcn.nl/servers/html/validate.html>
- 5) EDS, <http://eds.bmc.uu.se/eds/index.html>
- 6) Zhao et al. Specific Function of the Met-Tyr-Trp Adduct Radical and Residues Arg-418 and Asp-137 in the Atypical Catalase Reaction of Catalase-Peroxidase

KatG. doi: 10.1074/jbc.M112.401208. October 26, 2012. The Journal of Biological Chemistry.

7) PDB_redo, <https://pdb-redo.eu/>