

**МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
имени М.В.ЛОМОНОСОВА**

ФАКУЛЬТЕТ БИОИНЖЕНЕРИИ И БИОИНФОРМАТИКИ

Отчет о качестве РСА структуры белка PDB ID: 4RI1

Положинцева Артемия Ильича

Москва

2018 г.

Содержание

Аннотация	2
Введение	2
Результаты	3
Общая информация о модели	3
Значения индикаторов качества модели в целом	3
R-фактор	3
R-free	4
Перекрывание Ван-дер-Ваальсовых радиусов атомов (Clashscore)	4
Карта Рамачандрана	5
Ротамеры (Sidechain outliers)	5
RSR outliers	5
Анализ маргинальных остатков	5
Lys3 chain A	6
Leu91 chain B	6
Glu158 chain A	7
Lys170 chain A	7
Lys2 chain A	8
Сравнение модели из PDB с моделью из PDB-REDO [3]	9
Заключение	10
Список литературы	10

Аннотация

В работе был произведен анализ качества расшифровки структуры PDB ID: 4RI1. Описаны основные показатели качества модели и их соответствие общепринятым значениям. Рассмотрено несколько маргинальных аминокислотных остатков.

Введение

N-ацетилтрансфераза биосинтеза псевдоаминовой кислоты (PseH) катализирует третью стадию биосинтеза псевдоаминовой кислоты в *Helicobacter pylori*. Эта кислота необходима для работы жгутика бактерии и развития рака желудка. Фермент принадлежит к GCN5-родственному суперсемейству N-ацетилтрансфераз (GNAT).

Структура была получена методом рентгеноструктурного анализа в работе 2015 года [1]. Это первая расшифрованная структура суперсемейства GNAT с подобной специфичностью. PseH является гомодимером в кристалле, каждая субъединица которого

имеет завернутый бета-лист в центре, окруженный пятью альфа-спиралями и структурно весьма схож с другими представителями суперсемейства GNAT. Структура была кристаллизована и расшифрована в комплексе с ацетил-КоА, что позволило определить сайт связывания кофактора.

Результаты

Общая информация о модели

Таблица 1. Основные параметры структуры 4RI1

Параметр	Значение
Состав комплекса	В асимметричной ячейке - 3 субъединицы, 3 молекулы AcCoA
Год	2015
Авторы	Abu I. Ud-Din, Yu C. Liu, Anna Roujeinikova
Разрешение	2.3 Å
Метод разрешения фазовой проблемы	MIRAS (method of multiple isomorphous replacement coupled with anomalous scattering)
Количество измеренных рефлексов	54302 (нативный белок)
Полнота набора структурных факторов	96% (нативный белок)
Диапазон разрешений структурных факторов	30–2.3 Å
Тип симметрии кристалла	I2 ₁ 2 ₁
Параметры элементарной ячейки	a = 107.8 Å, b = 145.6 Å, c = 166.2 Å

Значения индикаторов качества модели в целом

Основными индикаторами качества модели считаются R фактор и R-free. Также качество структуры можно оценить по доле маргинальных остатков по разным критериям, таким как карта Рамачандрана, перекрывание Ван-дер-Ваальсовых радиусов атомов (clashscore), длин и углов связей, ротаметров.

R-фактор

R-фактор является оценкой того насколько расшифрованная структура способна предсказать наблюдаемые результаты. Хорошими значениями для этого параметра считаются показатели меньше 0.25; R-фактор рассматриваемой структуры равен 0.178, т.е. полученная

модель хорошо предсказывает результаты рентгеноструктурного эксперимента.

R-free

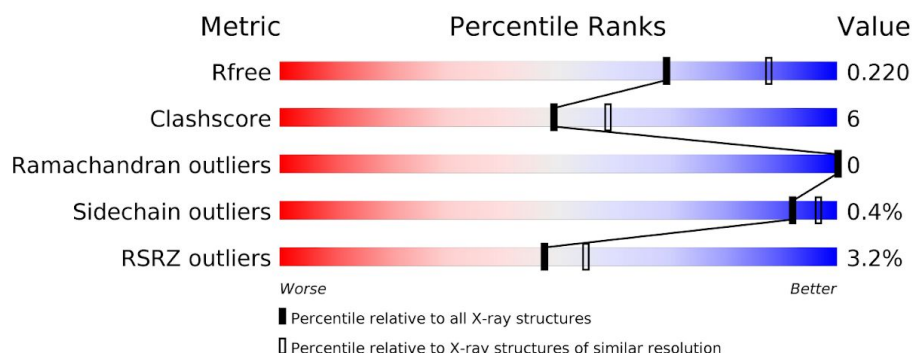


Рис. 1: Основные показатели качества модели (черным – относительно всех PCA моделей; белым – относительно PCA моделей близкого разрешения).

R-free вычисляется точно так же, как R-фактор, но по относительно небольшой заранее отобранной случайной выборке экспериментальных данных (рефлексов), не используемых при оптимизации модели (в данном случае 5% от полученных данных). Таким образом, R-free является контролем переоптимизации модели – правильно оптимизированная модель должна предсказывать одинаково хорошо все наблюдаемые результаты (а не только те, по которым она оптимизировалась). В данном случае R-free составил 0.220, хоть это значение и лучше, чем у большинства PCA структур подобного разрешения (Рис. 1), оно на 23.6% больше R-фактора, в то время как хорошими считаются значения менее 10%. Это означает, что модель явно переоптимизирована.

Перекрытие Ван-дер-Ваальсовых радиусов атомов (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. Значение clashscore для рассматриваемой структуры - 6.0. По данным MolProbity, это значение является 98-ым перцентилем, это означает, что структура по clashscore качественнее, чем 98 процентов структур.

Карта Рамачандрана

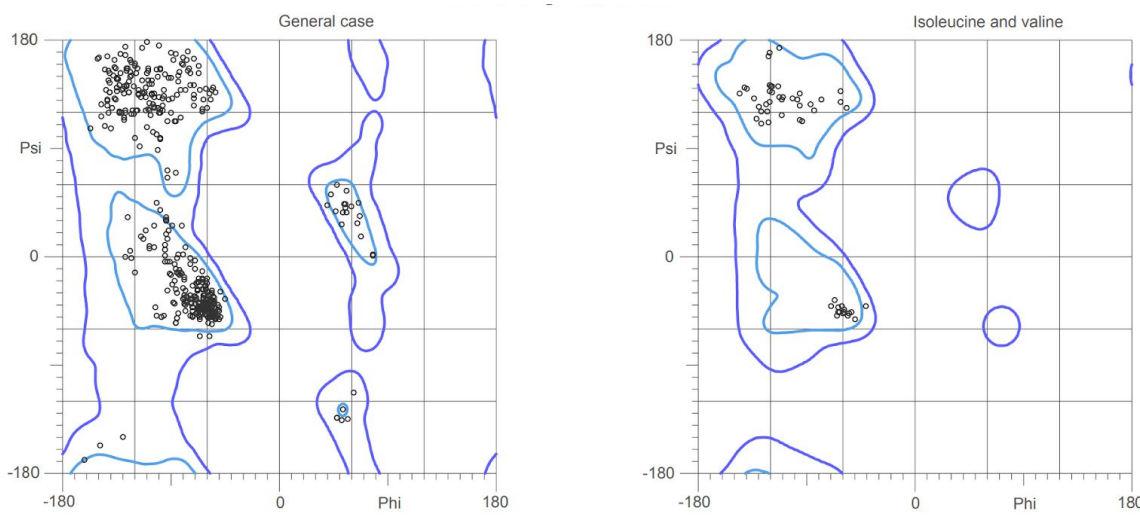


Рис. 2. Карты Рамачандрана для рассматриваемого белка

Карта Рамачандрана позволяет оценить, есть ли маргиналы по торсионным углам ϕ и ψ . На рисунке 2 представлена карта Рамачандрана для всего белка, построенная программой MolProbity [2]. В данном случае нет ни одного остатка в запрещенной области, а в не предпочтительных регионах находятся 17 остатков.

Ротамеры (Sidechain outliers)

Данный параметр учитывает конформацию боковых цепей аминокислот и позволяет найти маргинальные остатки по этому критерию. Ротамерами называются боковые цепи в предпочитаемых для данного остатка положениях, которые описываются наборами допустимых значений торсионных углов χ_i . В данной структуре маргинальных (запрещённых) остатков по положению боковых цепей - 6, что составляет 1.24 % от всех остатков. По оценке же MolProbity таких остатков должно быть меньше 0.3%, поэтому их больше оптимального.

RSR outliers

Параметр RSR характеризует, насколько хорошо остаток в модели вписывается в экспериментальную электронную плотность. 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 > 2, считается, что остаток маргинальный по RSR и не очень хорошо описывается экспериментальными данными. RSR оценивается для каждого остатка отдельно и может служить мерой локального качества. Для оценки качества структуры в целом можно посмотреть на долю маргинальных остатков по RSR: в данной структуре их 3.2 %. Это весьма средние значения.

Анализ маргинальных остатков

Данных RSR по остаткам получено не было, поскольку сервер EDS не обладает электронной картой белка 4tr1 и не смог выдать результатов. Маргинальных остатков по картам Рамачандрана также получено не было, поэтому они отсутствуют в таблице.

Таблица 2. Примеры маргинальных остатков

Остаток	Критерий, по которому остаток маргинальный
Lys3 chain A	Углы χ_i
Leu91 chain B	Clashscore (overlap 0.82Å!)
Glu158 chain A	Clashscore (overlap 0.56 Å)
Lys170 chain A	Clashscore (overlap 0.54 Å), Углы χ_i
Lys2 chain A	Clashscore (overlap 0.43 Å), Углы χ_i

Lys3 chain A

Остаток имеет необычные значения торсионных углов боковой цепи. В этом месте был плохой сигнал, что показывает электронная плотность, и структура не разрешилась (рис. 3).

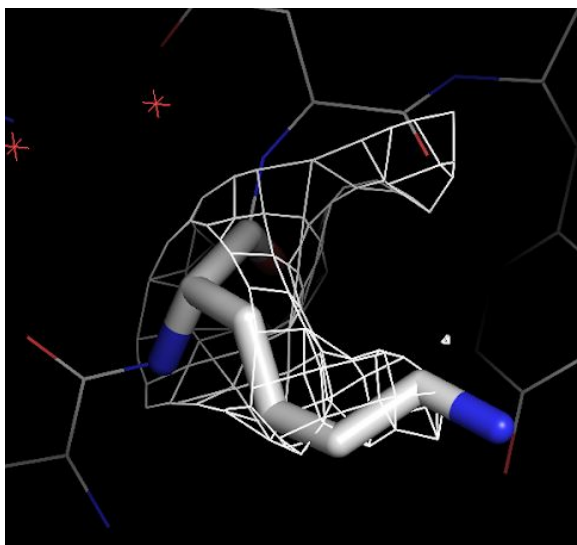


Рис. 3 Lys3 chain A с электронной плотностью ($Z = 1.4$)

Leu91 chain B

Очень большой clashscore 0.82Å: атомы водородов leu123 (желтая раскраска) и leu91 (сиреневая раскраска рис.4) сильно пересекаются.

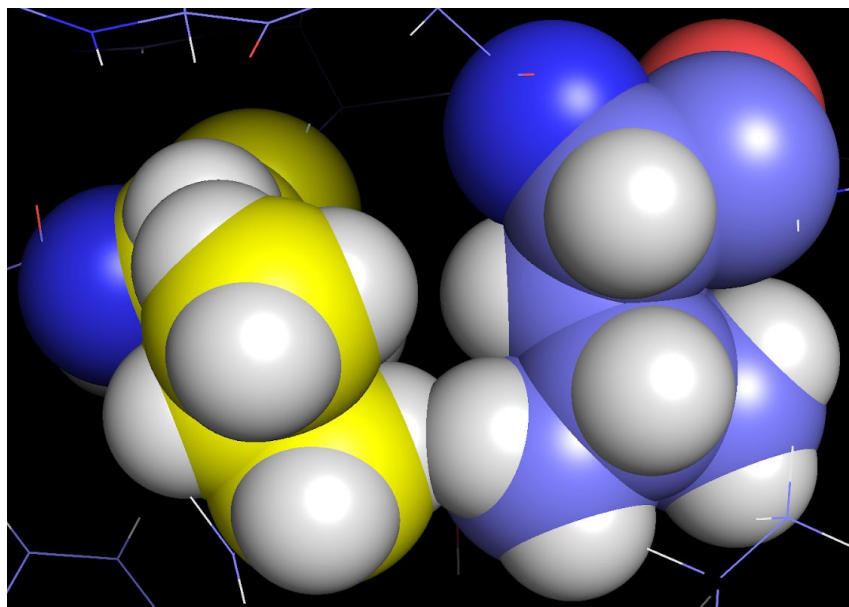


Рис. 4 leu91 chain B с электронной плотностью ($Z = 1.4$)

Glu158 chain A

В данном остатке наблюдается явная ошибка определения (дубликация боковой цепи), возникшая в силу отсутствия электронной плотности в данной области (рис. 5).

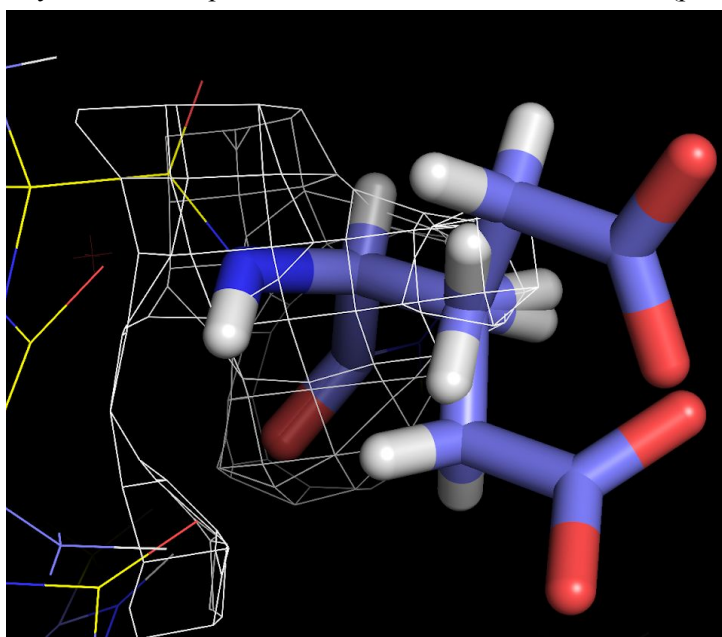


Рис. 5 Glu158 chain A с электронной плотностью ($Z = 1.4$)

Lys170 chain A

В данном случае была неправильно определена вода, что дало сильное пересечение с азотом и водородом lys170 (рис.6А). Также на рис. 6Б видно что вокруг боковой цепи данного лизина отсутствует электронная плотность, что объясняет ошибки в определении воды и неверные торсионные углы в боковой цепи.

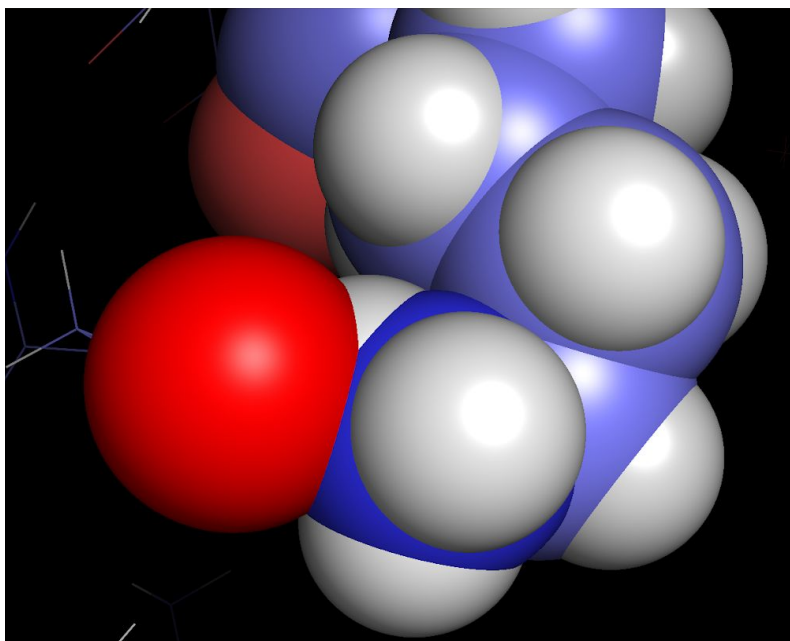


Рис. 6А боковая цепь Lys170 chain A с водой рядом

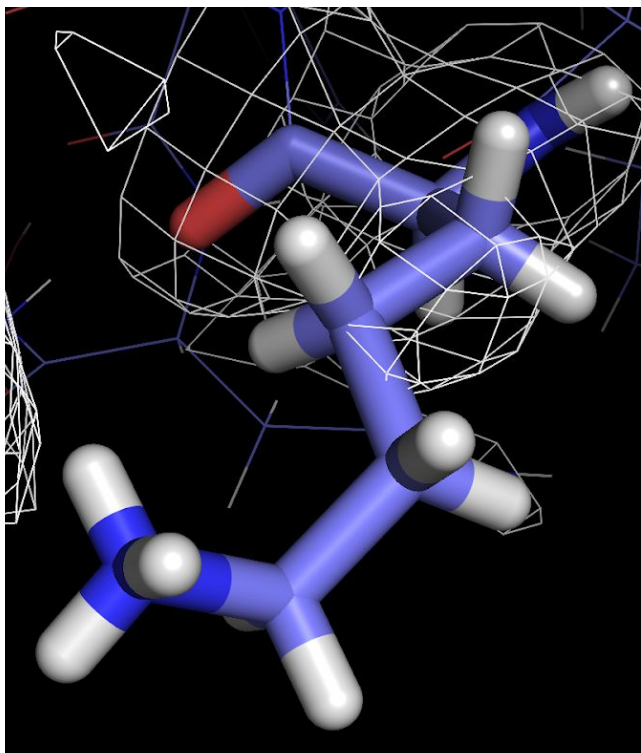


Рис. 6Б Lys170 chain A с электронной плотностью ($Z = 1.4$)

Lys2 chain A

В данном случае присутствует перекрывание собственных водородов HA и HD3 lys2 (рис.7А). Оно возникло из-за неверных торсионных углов боковой цепи, которые в свою очередь обусловлены недостаточной электронной плотностью в данной области(рис.7Б).

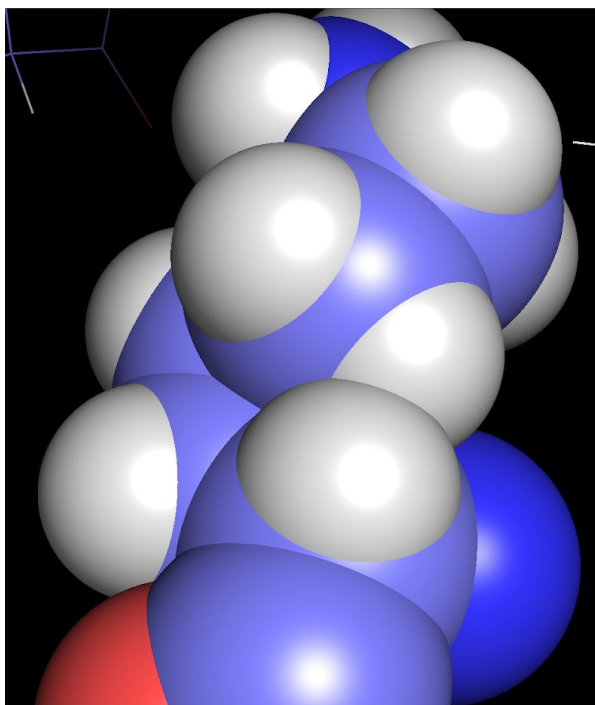


Рис. 7А Lys2 chain A самопересекающиеся водороды расположены по центру

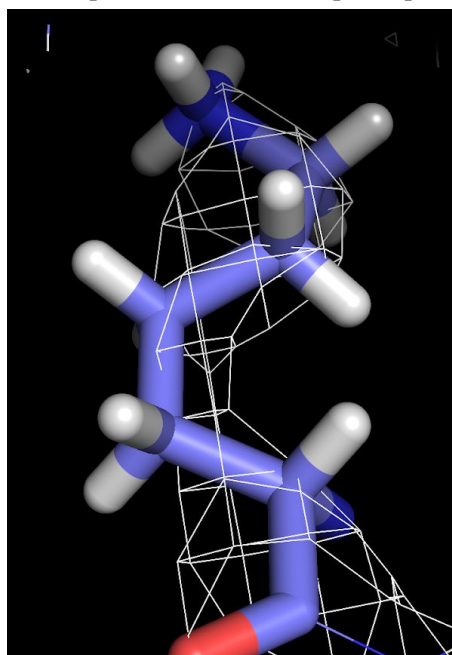


Рис. 7Б Lys2 chain A с электронной плотностью ($Z = 1.4$)

Сравнение модели из PDB с моделью из PDB-REDO [3]

Как видно из таблицы 3, в PDB-REDO удалось улучшить модель по многим параметрам, особенно стоит отметить снижение разницы между R и R-free до 13% (против исходных 23.6%), что означает что модель гораздо меньше переоптимизирована. Множество остатков было лучше вписано в электронную плотность и убраны некоторые молекулы воды, что оправдано, если судить по обсуждавшимся выше маргинальным остаткам.

Таблица 3.

Параметр	PDB	PDB-REDO
R-фактор	0,1964	0,1766
R-free	0,2212	0,1995
Residues fitting density better	0	120
Ramachandran plot appearance	40	86
Rotamer normality	71	95

Заключение

Структура белка 4RI1 расшифрована достаточно качественно по меркам средней структуры из PDB. Она имеет разрешение 2.3 Å и её показатели лучше, чем у большинства структур такого же разрешения. По данным из PDB_redo можно заключить, что структура могла быть оптимизирована гораздо лучше, особенно если посмотреть на R-free, ротаметры и вписанные остатки. Рассмотренные маргиналы в основном были получены из-за ошибок расшифровки.

Список литературы

1. Ud-Din, A.I., Liu, Y.C. and Roujeinikova, A., 2015. Crystal structure of Helicobacter pylori pseudaminic acid biosynthesis N-acetyltransferase PseH: implications for substrate specificity and catalysis. PLoS One, 10(3), p.e0115634.
2. <http://molprobit.biochem.duke.edu>
3. <https://pdb-redo.eu/db/4ri1>