

Оценка качества расшифровки структуры периплазматического металлсвязывающего белка из *Candidatus Liberibacter asiaticus* (PDB ID: 4cl2)

Кучеренко Варвара

Аннотация

В данном отчете проводилась оценка качества расшифровки структуры периплазматического металлсвязывающего белка, полученной методом рентгеноструктурного анализа (РСА). Для файла с PDB ID = 4CL2 были рассмотрены показатели качества модели в целом, найдены и проанализированы некоторые маргинальные аминокислотные остатков.

Введение

«Болезнь желтого дракона» (Citrus Huanglongbing, HLB) – поражающее цитрусовые заболевание, вызываемое бактериями *Candidatus Liberibacter spp. Candidatus Liberibacter asiaticus* (CLA) считается одним из самых агрессивных видов-возбудителей HLB. Система Znu грамотрицательной бактерии CLA чрезвычайно важна для выживания и патогенности микроорганизма. Эта система обеспечивает транспорт ионов металлов через мембрану бактерии [1]. Белки системы принадлежат к суперсемейству ABC-транспортеров (подгруппа A-I), куда относятся транспортеры цинка, марганца и железа [2]. ABC-транспортеры состоят из трех компонентов: экстрацитоплазматические рецепторы – *растворимые связывающие белок* (SBP), находящийся в периплазме грамотрицательных бактерий или связанный с цитоплазматической мембраной грамположительных бактерий, *трансмембранная пермеаза* и *нуклеотид-связывающий белок* (АТФаза) [3].

Общая структура SBP семейства A-I состоит из пары N- и C-концевых доменов с укладкой $(\alpha/\beta)_4$ - сэндвич, соединенных длинной основной α -спиралью. Сайт связывания металлов находится на стыке доменов [1]. Для цинк-связывающих SBP известны кристаллические структуры как белка, связавшего ион металла, так и свободного [4,5,6] – смена состояний происходит благодаря конформационным перестройкам C-концевого домена, при этом не осуществляется каких-либо значимых сдвигов связующей α -спирали или доменов друг относительно друга [7]. В случае Mn/Fe-связывающих SBP, наоборот, демонстрируется «жесткое» перемещение C-концевого домена при смене состояний. В итоге образуется «открытая» конформация, доступная ионам металла [4].

У CLA имеются всего два гомолога системы Znu, только одна из которых предположительно контролирует транспорт марганца, необходимого для жизнедеятельности бактерий [1]. Структура 4CL2 соответствует SBP из этой системы.

Результаты

Общая информация о модели

Структура периплазматического металлсвязывающего белка бактерии *Candidatus Liberibacter asiaticus* (PDB ID: 4CL2) была опубликована на сайте PDBе 28 января 2015 года группой ученых (Sharma N., Selvakumar P., Bhose S., Ghosh D.K., Kumar P., Sharma A.K.) [8].

Параметры модели указаны в таблице 1 ниже:

Таблица 1. Информация о модели	
Состав комплекса	Мономер цепь A: 275 а.о. (из 294 а.о.) Лиганды и окружение: 6 молекул сульфата (SO ₄), 2 молекулы ацетата, 2 молекулы глицерина и один ион марганца (Mn ²⁺)
Метод решения фазовой проблемы	Молекулярное замещение
Число измеренных рефлексов	57802
Разрешение	1.63Å
Полнота набора структурных факторов	95.4%
Диапазон разрешений структурных факторов	1.53Å -29.27Å
Параметры кристаллографической ячейки	Длины направляющих векторов: a = 94.060 Å, b = 94.060 Å, c = 94.510 Å; Углы между векторами: α = 90.00°, β = 90.00°, γ = 120.00°
Наличие некристаллографической симметрии в ассиметрической ячейке	Нет
Тип кристаллической симметрии	P 3 ₂ 2 1

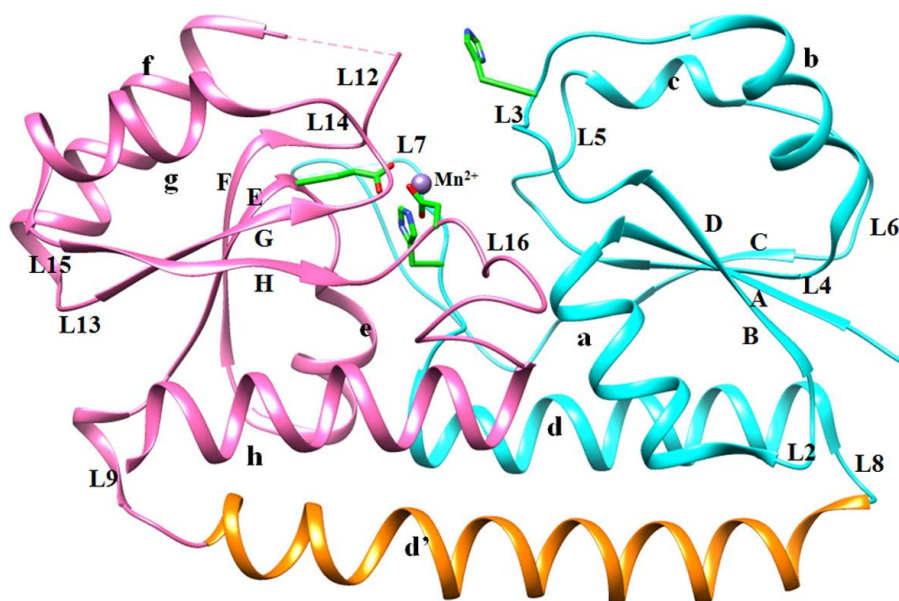


Рис. 1. Структура 4CL2. β-листы обозначены буквами A-H, α-спирали обозначены буквами a-h, петли - как L2-L16. Остатки, координирующие ион металла, отображены стиками [1]

Значения индикаторов качества модели в целом

К этим индикаторам относятся R-фактор, характеризующий соответствие построенной модели экспериментальным данным, и R_free, отвечающий за контроль переоптимизации модели.

R-фактор рассчитывается по следующей формуле:

$$R = \frac{\sum |F_{obs} - F_{calc}|}{\sum F_{obs}}$$

Где F_obs – экспериментальные модули факторов, F_calc – посчитанные по модели модули факторов. Суммирование производится по всем векторам (h,k,l). Соответственно, чем меньше значение R-фактора, тем лучше считается модель (отличным считается показатель меньше 0.25).

R_free рассчитывается аналогичным образом, за исключением того, что для вычислений берется небольшое количество случайно выбранных рефлексов, не участвовавших в оптимизации. Для идеальной модели значение R_free должно быть равно R-фактору, но на практике для хорошей модели допустима разница в 10%.

Таблица 2. Значения R-фактора и R-free структуры 4CL2

R-фактор	0.178 (depositor), 0.167 (EDS)
R_free	0.201 (depositor), 0.194 (EDS)
Доля структурных факторов, отобранных для вычисления R_free	5.06% (2927 рефлексов)

Значения R-факторов достаточно малы, чтобы назвать модель хорошей. Разница между R_free и R-фактором составляет 0.023 и 0.027 (DCC и EDS соответственно), что удовлетворяет критериям хорошей модели.

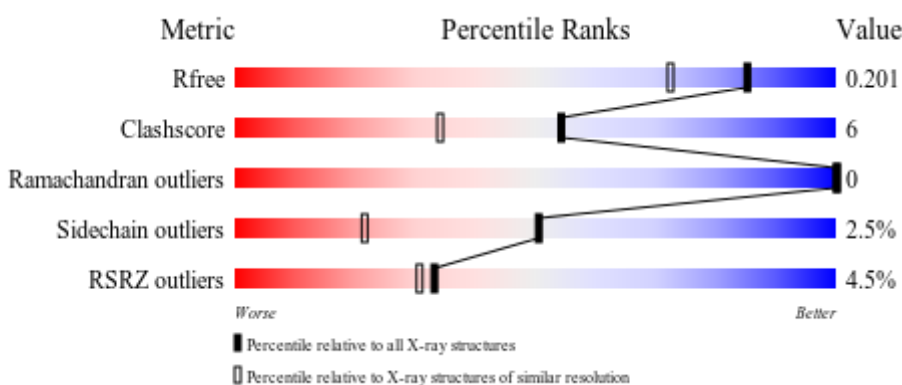


Рис. 2. Качество модели 4CL2 в сравнении с другими моделями из PDB. Градиент цвета показывает, насколько параметры данной структуры лучше (синий) или хуже (красный) средних значений (белый). Черные прямоугольники отражают качество относительно всех структур, белые — относительно структур схожего разрешения. Clashscore – число пар атомов, сильно близко расположенных друг к другу, приходящихся на 1000 атомов (вычисляется сервером MolProbity с учетом добавленных атомов водорода). Ramachandran outliers – процент аминокислотных остатков с необычными торсионными углами ϕ и ψ (вычисляется сервером MolProbity). Sidechain outliers – процент аминокислотных остатков с необычной конформацией боковой цепи (вычисляется сервером MolProbity). RSRZ outliers –

процент аминокислотных остатков с $RSRZ > 2$, где $RSRZ$ – это Z-score, являющийся нормализацией *real-space R-value* (RSR), который представляет собой меру качества соответствия атомной модели остатка «экспериментальной» электронной плотности вокруг него (вычисляется сервером EDS). [8]

Видно, что среди всех PDB-структур 4CL2 имеет хорошие показатели по R-фактору и Ramachandran outliers, тогда как по остальным показателям модель средняя. По сравнению со структурами схожего разрешения модель является посредственной.

Для более подробного анализа параметров качества был использован сервис MolProbity, предназначенный для валидации структур макромолекул [9]. С его помощью перед анализом в структуру можно добавить атомы водорода, оптимизировать сети водородных связей и произвести инверсии боковых цепей аспарагина, глутамина и гистидина, если это необходимо.

Исследуемая структура состоит из 2764 атомов (2419 атомов белковой цепи). Сервер MolProbity добавил 2364 атома водорода к полипептидной цепи для анализа показателя Clashscore. В структуру 4CL2 всего было добавлено 2200 атомов водорода, 84 из которых подверглись «подгонке» для оптимизации водородных связей. Также было найдено 4 неправильно ориентированных остатка (Asn53, His63, Asn208, Asn237). Данные анализа приведены в таблице 3.

Таблица 3. Индикаторы качества модели структуры 4CL2 (MolProbity)

All-Atom Contacts	Clashscore, all atoms:	5.92		91 st percentile* (N=709, 1.63Å ± 0.25Å)
	Clashscore is the number of serious steric overlaps (> 0.4 Å) per 1000 atoms.			
Protein Geometry	Poor rotamers	5	2.04%	Goal: <0.3%
	Favored rotamers	234	95.51%	Goal: >98%
	Ramachandran outliers	0	0.00%	Goal: <0.05%
	Ramachandran favored	251	97.29%	Goal: >98%
	MolProbity score [^]	1.71		74 th percentile* (N=7243, 1.63Å ± 0.25Å)
	Cβ deviations >0.25Å	0	0.00%	Goal: 0
	Bad bonds:	2 / 2238	0.09%	Goal: 0%
Bad angles:	3 / 3044	0.10%	Goal: <0.1%	
Peptide Omegas	Cis Prolines:	0 / 12	0.00%	Expected: ≤1 per chain, or ≤5%

Немного пояснений к параметрам в таблице:

ClashScore - это число недопустимых наложений атомов на 1000. Наложением является пересечение Ван-дер-Ваальсовых радиусов более, чем на 0.4 Å. *Poor rotamers* - это число остатков с маргинальными по отклонению от ротамеров боковыми цепями. *Ramachandran outliers* - полные маргиналы по карте Рамачандрана те, которые лежат вне допустимой области. *Cβ deviations >0.25Å* - число Cβ-атомов с неприемлемым отклонением от ожидаемого положения. *Bad bonds* - число ковалентных связей остова, существенно отклоняющихся от теоретических. *Bad angles* - число валентных углов остова, существенно отклоняющихся от теоретических. *Ramachandran favored* - число и процент остатков в предпочитаемой области. *MolProbity score* - интегральная оценка структуры по данным этого сервиса. *Favored rotamers* – число остатков с боковыми цепями не маргинальным по отклонению от ротамеров.

Для данной структуры ClashScore равно 5.92, что является 91-тым перцентилем. Это говорит о хорошем качестве модели по сравнению с другими, близкими по разрешению (в диапазоне $1.63\text{\AA} \pm 0.25\text{\AA}$).

Судя по таблице, модель содержит некоторые отклонения от теории.

Карта Рамачандрана позволяет оценить, насколько конформация остова модели соответствует предпочитаемым торсионным углам ϕ и ψ .

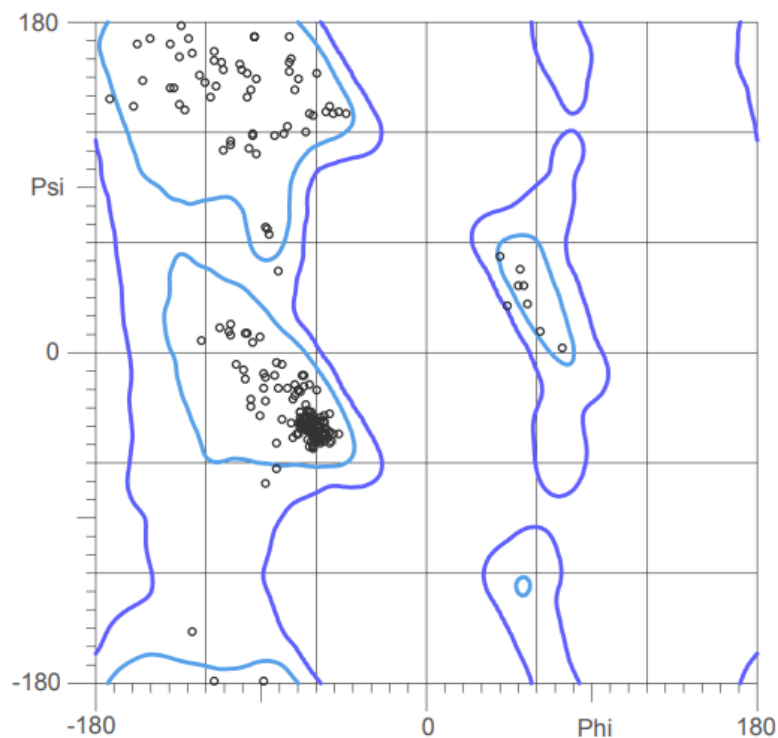


Рис. 3. Карта Рамачандрана для структуры 4CL2, построенная сервисом MolProbity. Голубым контуром обозначена предпочтительные области, синим – допустимые. Аминокислотные остатки изображены точками.

97.2% от всех остатков (273/281) лежат в предпочтительной области, остальные аминокислотные остатки – в допустимой области. Таким образом, в модели маргинальные остатки по торсионным углам отсутствуют.

Т.к. остатки изолейцина, глицина и валина часто принимают нестандартные значения торсионных углов, для них были построены отдельные карты. Также отдельные карты были построены по изомерам пролина и по пре-пролину.

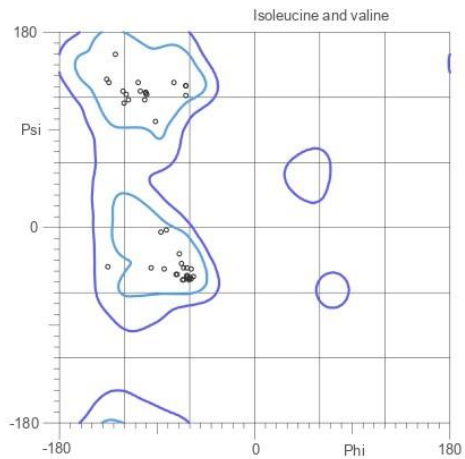


Рис. 4. Карта Раманчандрана для остатков изолейцина и валина структуры 4CL2, построенная сервисом MolProbity. Голубым контуром обозначена предпочтительные области, синим – допустимые. Аминокислотные остатки изображены точками.

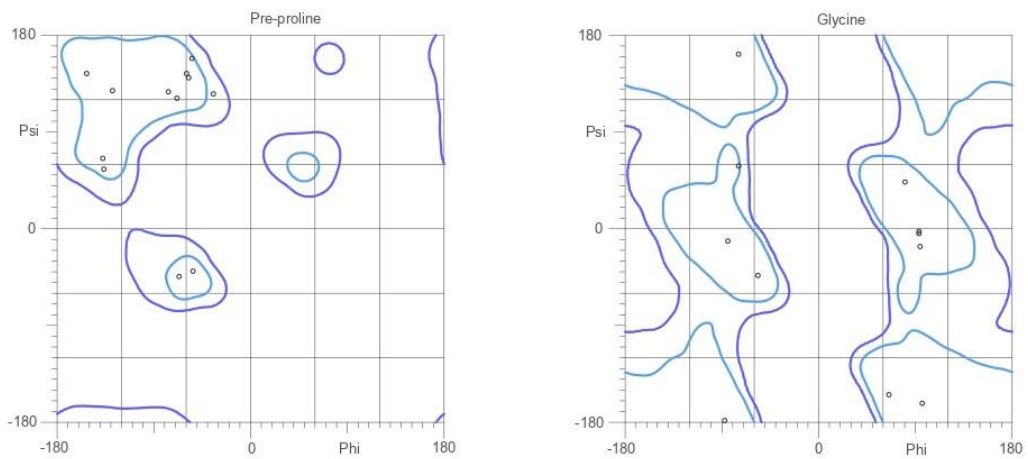


Рис. 5. Карта Раманчандрана для остатков пре-пролина (слева) и глицина (справа) структуры 4CL2, построенная сервисом MolProbity. Голубым контуром обозначена предпочтительные области, синим – допустимые. Аминокислотные остатки изображены точками.

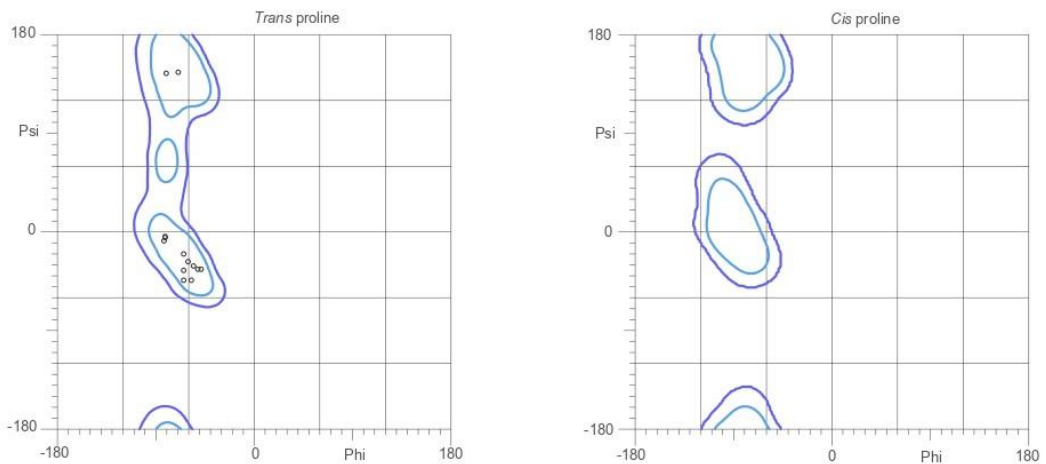


Рис. 6. Карта Раманчандрана для остатков транс-пролина (слева) и цис-пролина (справа) структуры 4CL2, построенная сервисом MolProbity. Голубым контуром обозначена предпочтительные области, синим – допустимые. Аминокислотные остатки изображены точками.

Несмотря на склонность к «нестандартным» торсионным углам у рассмотренных на отдельных картах остатках, все они оказались в допустимой области. Цис-изомеров пролина найдено вовсе не было, что логично, т.к. транс-пролин гораздо чаще образует пептидные связи (в случае цис-конформации получается стерическое отталкивание α -атомов).

Для оценки конформации боковых цепей аминокислотных остатков служат ротамеры – наборы предпочитаемых положений боковых цепей для данного остатка, заданные допустимыми значениями торсионных углов χ . Согласно критерию оценки MolProbity у хорошей модели доля «плохих» ротамеров должна быть менее 0.3%. В структуре 4cl2 95.51% (234/245) ротамеров являются «хорошими» и только 2.45% (5/245) – «плохими». 2.45% > 0.3%, что не удовлетворяет критериям хорошей модели.

Анализ маргинальных остатков

Информация о маргинальных остатках представлена в таблице 4.

Таблица 4. Маргинальные остатки (согласно MolProbity)

Критерий отбора	Число аминокислотных остатков	Процент аминокислотных остатков (%)
Хиральность	0	0
Планарность	0	0
Торсионные углы остова белка (карта Рамачандрана)	0	0
Торсионные углы боковых цепей остатков белка (ротамеры)	5	2.45
Соответствие модели «экспериментальным» данным (RSRZ>2)	12	4.4
Длина связи	2	0.09
Углы связи	3	0.01
Clashscore	25	1.06

На рисунке 7 визуализировано качество моделирования аминокислотных остатков на полипептидной цепи 4cl2.

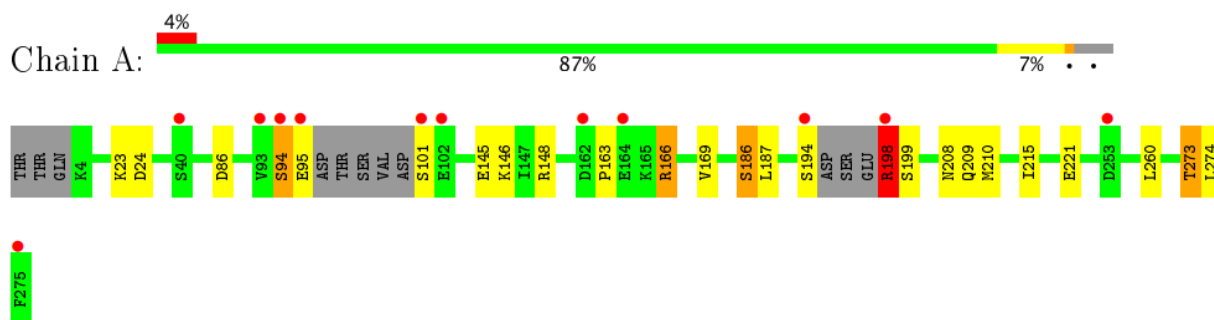


Рис.7. Вверху: Схема соотношения в процентах аминокислотных остатков структуры 4cl2, окрашенных по принципу их соответствия геометрическим критериям качества модели. Зеленым цветом обозначены немаргинальные остатки, желтым – маргиналы по одному критерию, оранжевым – по двум, красным – по трем и более, серым окрашены участки, которые присутствовали в кристалле, но не вошли в модель. Внизу: Схема расположения аминокислотных остатков разного качества на полипептидной цепи. Принцип окраски тот же. Красными точками отмечены остатки с плохо соответствующими электронными плотностями ($RSRZ > 2$). Для обозначения идущих подряд нескольких «зеленых» остатков (два и более) используются зеленые полосы (внизу).

«Серая» область на схеме выше небольшая, однако в модели присутствует множество маргиналов в связи с перекрытиями Вандерваальсовых радиусов атомов, а также многочисленными несоответствиями остатков экспериментальной электронной плотности, что снижает качество модели.

В таблице 5 приведены примеры маргинальных остатков 4cl2 с указанием критерия отбора.

Таблица 5. Маргинальные остатки структуры 4cl2

Аминокислотный остаток	Критерий маргинальности
ARG198	Перекрытие радиусов атомов, неблагоприятная конформация боковой цепи, плохое соответствие экспериментальной электронной плотности, длина связи
PRO163	Перекрытие радиусов атомов
SER94	Перекрытие радиусов атомов, плохое соответствие экспериментальной электронной плотности, неблагоприятная конформация боковой цепи
ASN53	Подозрение на необходимость инверсии
SER 186	Неблагоприятная конформация боковой цепи

Проанализируем подробнее маргинальные остатки из таблицы 5.

ARG 198

Это наиболее проблемный остаток: конформация его боковой цепи неблагоприятна (углы χ_1 , χ_2 , χ_3 , χ_4 составляют 315.9, 199.2, 190, 343.3), наблюдается перекрытие Ван-дер-Ваальсовых радиусов атомов (CG с HN11, 0.59 Å) в рамках одного аминокислотного остатка, также показатель RSRZ равен 2.3, что больше 2.

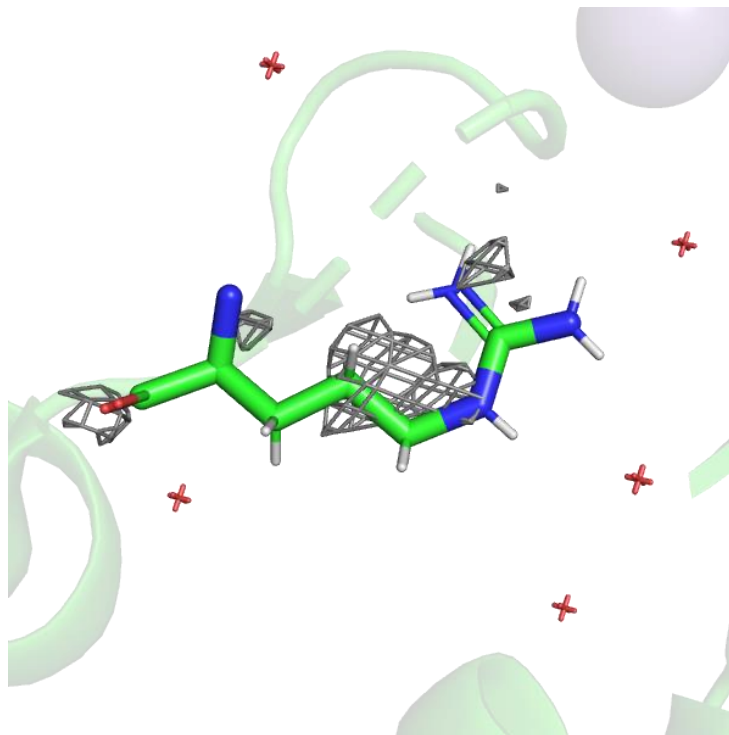


Рис.8. Arg 198 с изображением электронной плотности на уровне подрезки 1.5 σ

На рисунке 8 изображен данный остаток и электронная плотность вокруг него на уровне подрезки 1.5 σ . Видно, что электронная плотность на данном уровне подрезки не соответствует остову, более того, она концентрируется в области атомов углерода с небольшой электроотрицательностью, что говорит о неверной расшифровке. Такое плохое качество расшифровки возможно обусловлено расположением аминокислотного остатка в петле – наиболее подвижном элементе структуры. Петли ввиду своей большой подвижности плохо кристаллизуются, поэтому положение остатков в петлях ассиметрических ячеек может различаться. Остаток находится недалеко от центра связывания иона марганца, однако функционально он не значим. Контакт с соседними структурами не имеется.

PRO163

Ван-дер-Ваальсов радиус HA атома этого остатка перекрывается на 0.583 Å и 0.578 Å с радиусом HD3 и CD ARG 166 соответственно.

Такое перекрытие возможно из-за положения Pro 163 и Arg 166, которые являются частью петли, соединяющей С-концевой домен с линкерной α -спиралью. Регион перехода спирали в домен характеризуется резким перегибом и большой мобильностью.

Кроме того, при визуализации электронной плотности, выяснилось, что на уровне подрезки 1.5σ кольцо пролина не вписывается в график. Даже на уровне подрезки 1σ кольцо пролина покрывается не полностью.

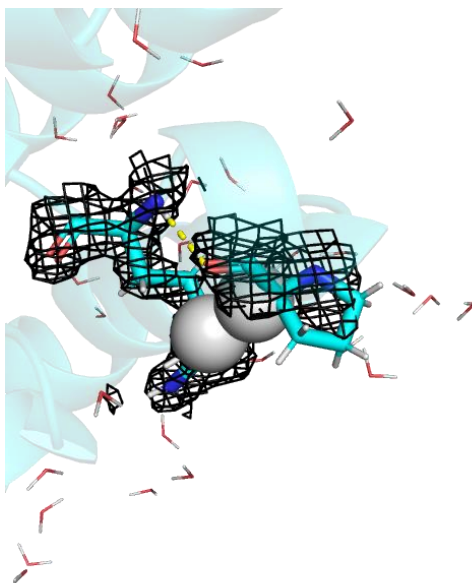


Рис.9. Pro 163 и Arg 166 с изображением электронной плотности на уровне подрезки 1.5σ . Атомы, Ван-дер-Ваальсовы радиусы которых перекрываются, изображены сферами для наглядности.

Возможно проблема слишком близкого расположения атомов была бы решена путем смещения кольца пролина, т.е. произошла ошибка расшифровки. Однако почти такое смещение приведет к изменениям торсионных углов, возможно до неблагоприятных величин.

SER94

Этот остаток маргинален сразу по нескольким критериям. Ван-дер-Ваальсов радиус HG атома этого остатка перекрывается на 0.473 \AA с радиусом атома N Ser 101, также конформация боковой цепи Ser 94 неблагоприятна ($\chi = 331.7$), а показатель RSRZ = 6.8. Маргинален снова расположен в районе петли (в N-концевом домене) у поверхности белка. Его боковая цепь обращена внутрь глобулы белка, однако функционально значимую роль он не играет.

На уровне подрезки 1.5σ у Ser94 электронная плотность «охватывает» только часть остова. При уменьшении параметра σ покрывается только остов, боковая цепь остается нетронутой. Хотя расстояние между HG и N соседнего остатка Ser 101 составляет 2.6 \AA , скорее всего здесь имеется водородная связь, тогда как искажение электронной плотности возникло в связи с повышенной подвижностью петли. Маргинальность по перекрытию атомов и конформации боковой цепи связана с ошибочной расшифровкой.

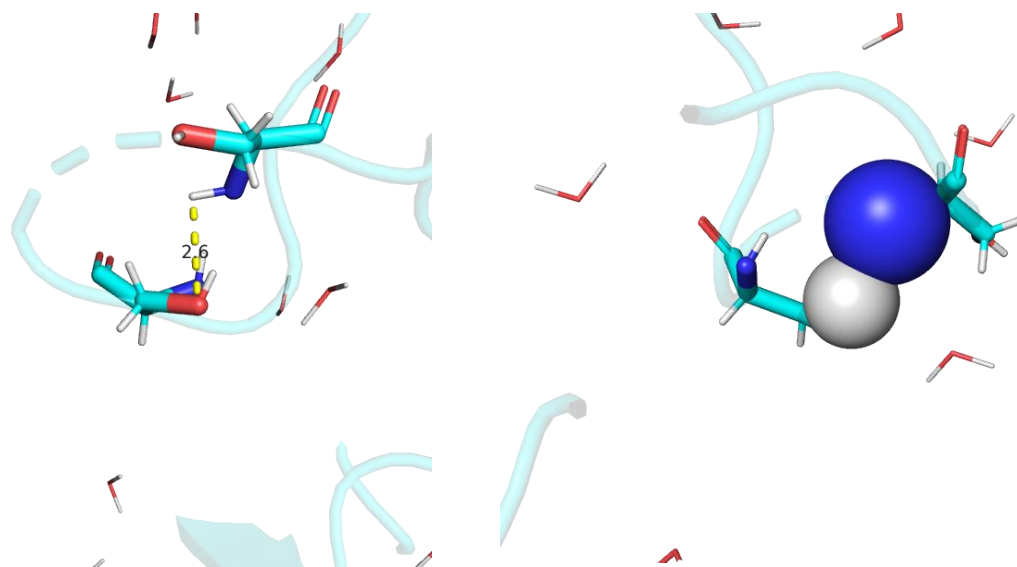


Рис.10. Слева: Расстояние между полярными атомами Ser 94 и Ser 101. Справа: Атомы Ser 94 и Ser 101, ван-дер-ваальсовы радиусы которых перекрываются, изображены сферами для наглядности.

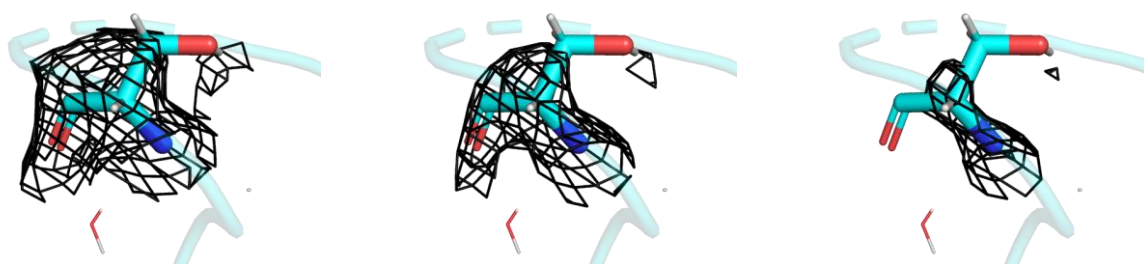


Рис.11. Слева: Ser 94 с изображением электронной плотности на уровне подрезки 0.5 σ . Центр: Ser 94 с изображением электронной плотности на уровне подрезки 1 σ . Справа: Ser 94 с изображением электронной плотности на уровне подрезки 1.5 σ .

ASN 53

Рассмотрим аминокислотный остаток с подозрением на необходимость инверсии боковой цепи. Он довольно хорошо вписан в «сетку» электронной плотности (на уровне подрезки 1.5 σ). С помощью сервиса MolProbity был получен PDB-файл с инвертированной боковой цепью для сравнения. Как и ожидалось, после инверсии остаток тоже хорошо «вписывается» в электронную плотность. Диаметр области электронной плотности, в которую вписан кислород, немного меньше, чем диаметр области с вписанным азотом. Кислород больше азота, следовательно, должен создавать вокруг себя чуть большее электронное облако, поэтому я бы сказала, что предпочтительнее инвертировать боковую цепь. Однако, т.к. разница между размерами электронных облаков совсем небольшая, можно сказать, что оба варианта приемлемы.

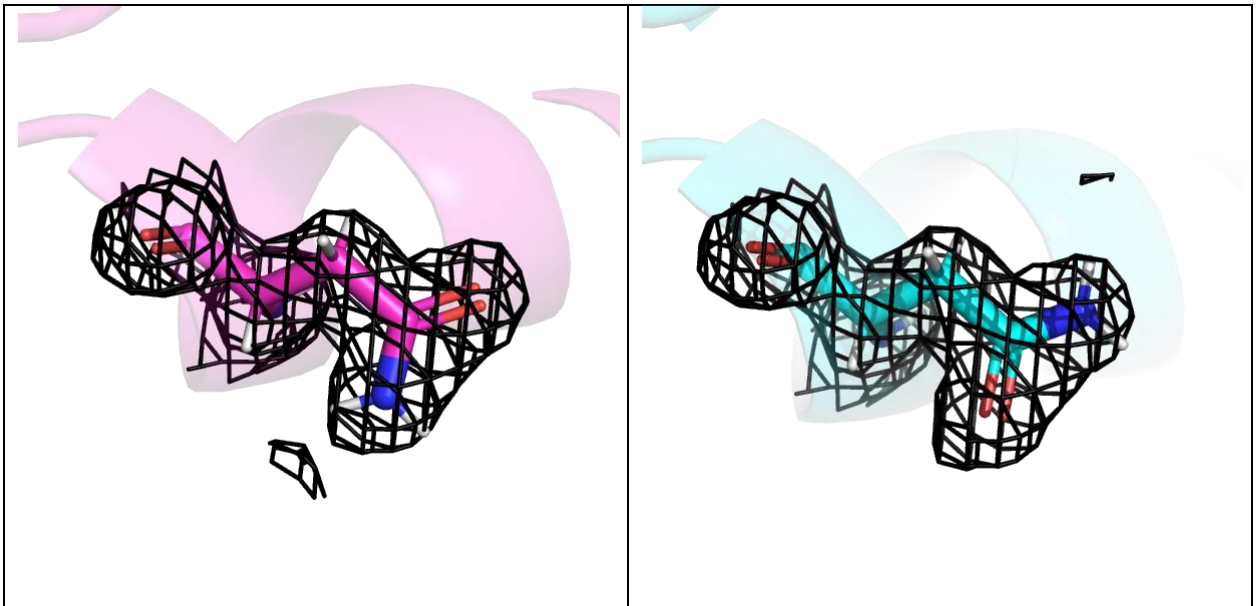


Рис.12. Слева: Asn 53 с изображением электронной плотности на уровне подрезки 1.5 σ до инверсии. Справа: После инверсии.

SER 186

У данного остатка наблюдается неблагоприятная конформация боковой цепи (χ : 223.1). Остаток входит в β -тяж. Серин связан водородной связью с фенилаланином из соседнего β -тяжа.

Судя по изображению электронной плотности, остаток достаточно хорошо «вписывается» в электронное облако. Видимо, ошибки расшифровки здесь нет – скорее всего, это особенность структуры, связанная с гидрофобным окружением остатка.



Рис.13. Слева: Водородные связи, образованные Ser 186. Центр: Ser 186 с изображением электронной плотности на уровне подрезки 1.5 σ . Справа: Ser 186 и его окружение.

Сравнение модели из PDB с моделью из PDB_redo

База данных PDB_redo содержит оптимизированные структуры из PDB с картами электронной плотности, описанием внесенных изменений и оценкой качества модели [10].

В версии PDB_redo для структуры 4cl2 были изменены 5 ротамеров. В итоге 16 остатков стали соответствовать электронной плотности хуже. Изменились также показатели качества (см. таблицу 6)

Таблица 6. Индикаторы качества модели в PDB и PDB_redo

Параметр	PDB	PDB_redo
R-фактор	0.1651	0.1740
R_free	0.1926	0.2010
RMSZ длин связей	0.856	0.754
RMSZ углов связей	0.873	0.818

Далее оптимизированная структура была проанализирована с помощью MolProbity.

Таблица 7. Индикаторы качества модели структуры 4cl2 после оптимизации PDB_redo (MolProbity)

All-Atom Contacts	Clashscore, all atoms:	5.69	92 nd percentile* (N=709, 1.63Å ± 0.25Å)
	Clashscore is the number of serious steric overlaps (> 0.4 Å) per 1000 atoms.		
Protein Geometry	Poor rotamers	5	2.04% Goal: <0.3%
	Favored rotamers	234	95.51% Goal: >98%
	Ramachandran outliers	0	0.00% Goal: <0.05%
	Ramachandran favored	251	97.29% Goal: >98%
	MolProbity score [^]	1.70	76 th percentile* (N=7243, 1.63Å ± 0.25Å)
	Cβ deviations >0.25Å	0	0.00% Goal: 0
	Bad bonds:	2 / 2239	0.09% Goal: 0%
Bad angles:	2 / 3046	0.07% Goal: <0.1%	
Peptide Omegas	Cis Prolines:	0 / 12	0.00% Expected: ≤1 per chain, or ≤5%

Судя по таблице 7, показатели изменились не сильно: немного улучшился Clashscore (с 5.92 до 5.69), уменьшилось количество отклонений от нормальных углов связей (с 0.1% до 0.07%), а также ухудшился показатель MolProbity score (с 1.71 до 1.70). В остальном показатели не поменялись.

В итоге оптимизация не дала значительных улучшений и в данном случае оказалась бесполезной. Возможно PDB_redo более эффективен для моделей более низкого качества. По-видимому, для дальнейшей оптимизации модели необходимо вручную проверять сомнительные остатки.

Заключение

По результатам оценки качества расшифровки структуры периплазматического металлсвязывающего белка (PDB ID: 4CL2) можно сказать, что модель среднего качества. Показатели R-фактора и R_{free} довольно хорошие, также модель не содержит остатков из запрещенной области карты Рамачандрана. Структура содержит достаточно много плохо совпадающих с экспериментальной электронной плотностью маргинальных аминокислотных остатков, а также несколько неротамерных (предположительно, как минимум один из них является особенностью структуры). Кроме того, присутствует некоторое количество маргиналов по перекрытию Ван-дер-Ваальсовых радиусов. Большая часть проблем связана с наличием множества подвижных петель в структуре.

Список литературы

- [1] Sharma N, Selvakumar P, Bhose S, Ghosh DK, Kumar P, Sharma AK, 2015. "Crystal structure of a periplasmic solute binding protein in metal-free, intermediate and metal-bound states from *Candidatus Liberibacter asiaticus*". *J. Struct. Biol.* 189(3), 94-184.
- [2] Higgins, C.F., 2001. "ABC transporters: physiology, structure and mechanism" an overview. *Res. Microbiol.* 152, 205–210.
- [3] https://en.wikipedia.org/wiki/ATP-binding_cassette_transporter
- [4] Counago, R.M., Ween, M.P., Begg, S.L., Bajaj, M., Zuegg, J., O'Mara, M.L., Cooper, M.A., McEwan, A.G., Paton, J.C., Kobe, B., 2014. Imperfect coordination chemistry facilitates metal ion release in the Psa permease. *Nat. Chem. Biol.* 10, 35–41.
- [5] Lee, Y.-H., Dorwart, M.R., Hazlett, K.R.O., Deka, R.K., Norgard, M.V., Radolf, J.D., Hasemann, C.A., 2002. The crystal structure of Zn (II)-free *Treponema pallidum* TroA, a periplasmic metal-binding protein, reveals a closed conformation. *J. Bacteriol.* 184, 2300–2304.
- [6] Yatsunyk, L.A., Easton, J.A., Kim, L.R., Sugarbaker, S.A., Bennett, B., Breece, R.M., Vorontsov, I.I., Tierney, D.L., Crowder, M.W., Rosenzweig, A.C., 2008. Structure and metal binding properties of ZnuA, a periplasmic zinc transporter from *Escherichia coli*. *J. Biol. Inorg. Chem.* 13, 271–288.
- [7] Chandra, B.R., Yogavel, M., Sharma, A., 2007. Structural analysis of ABC-family periplasmic zinc binding protein provides new insights into mechanism of ligand uptake and release. *J. Mol. Biol.* 367, 970–982
- [8] http://www.ebi.ac.uk/pdbe/entry-files/download/4cl2_full_validation.pdf
- [9] <http://molprobity.biochem.duke.edu>
- [10] <https://pdb-redo.eu>