

Отчет по качеству РСА расшифровки структуры белка пероксидазы хрена (HRP, PDB ID = 1KZM)

Волынкина Инна

Аннотация

В данном отчете проводилась оценка качества расшифровки структуры белка пероксидазы хрена, полученной методом рентгеноструктурного анализа (РСА). Для файла с PDB ID = 1KZM было выполнено рассмотрение показателей качества модели в целом, поиск и анализ некоторых маргинальных аминокислотных остатков.

Введение

Пероксидаза хрена (Horseradish Peroxidase, HRP, EC: 1.11.1.7) – это сравнительно небольшой белок (всего лишь 38.825 кДа), выполняющий, однако, такую важную функцию, как катализ превращения молекулы пероксида водорода в молекулу воды. Данный процесс сопровождается окислением ароматического соединения, содержащего два фенольных кольца, что позволяет нейтрализовывать многие токсичные восстановители. Но в еще большей степени пероксидаза хрена интересна своим широким применением в биоинженерии в качестве маркера. Так, для анализа белковых полиакриламидных гелей по методу Вестерн-блота в качестве вторичных антител преимущественно используют конъюгаты с HRP, которая катализируя окисление молекул люминола в присутствии перекиси, способствует испусканию света. Также при изучении клеточных структур на уровне световой и электронной микроскопии часто применяется другой субстрат для пероксидазы: 3,3'-диаминобензидин (DAB), который в окисленном состоянии локально полимеризуется и выпадает в виде тёмно-коричневого электронно-плотного осадка. На рисунке 1 приведена схема катализируемой реакции.

Поговорим немного о структуре самого белка. HRP представляет собой единый полипептид, состоящий из 308 аминокислотных остатков. Ее вторичная структура представлена 17 спиралями, из которых 13 являются α -спиралями, а 4 – короткими 3_{10} -спиралями, и одним небольшим β -листом, образованным двумя антипараллельными β -тяжами из 3 аминокислотных

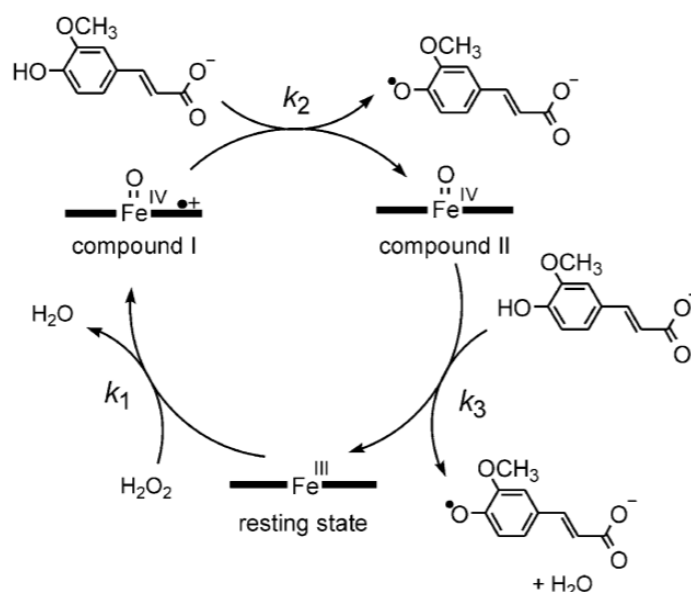


Рисунок 1. Схема каталитического цикла HRP на примере феруловой кислоты [1]

остатков каждый (рис. 2). Также расположение элементов вторичной структуры, помимо внутрибелковых гидрофобных взаимодействий, поддерживается четырьмя дисульфидными мостиками.

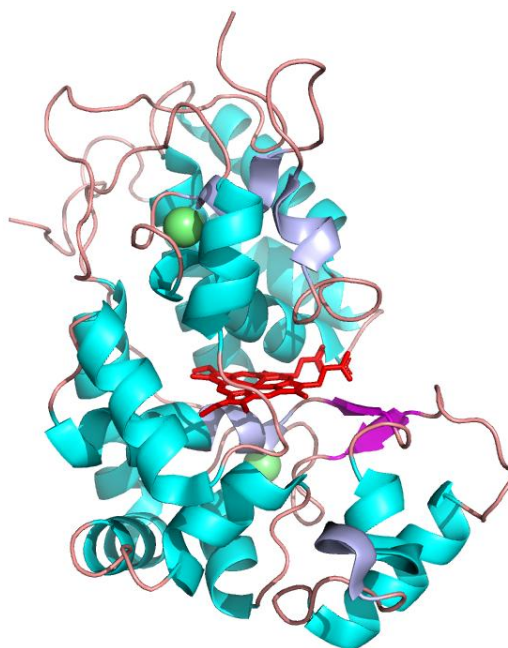


Рисунок 2. Структура HRP (1KZM) с раскраской по элементам вторичной структуры. Красным цветом обозначен гем, зеленым – ионы кальция, сиреневым – β -лист, фиолетовым – 3_{10} -спирали, бирюзовым – α -спирали, оранжевым – петли

Как известно из многочисленных рентгеноструктурных анализов белка HRP, в активном центре фермента протопорфирин IX (упрощенно называемый гемом) крепится к белковой глобуле через имидазольную группу His170,

правильная конформация которого обеспечивается водородной связью с Asp247. С другой стороны от плоскости протопорфирина IX располагаются боковые группы таких аминокислотных остатков, как Arg38, Phe41, His42 и Asn70. Они выполняют роль стабилизации переходных форм фермента и субстратов реакции (рис. 3). Однако для правильного функционирования фермента существенной является координация с двумя атомами кальция с дистальной и проксимальной сторон относительно плоскости протопорфирина IX соответственно. Каждый атом кальция образует по семь координационных связей с различными кислород-содержащими группами [1].

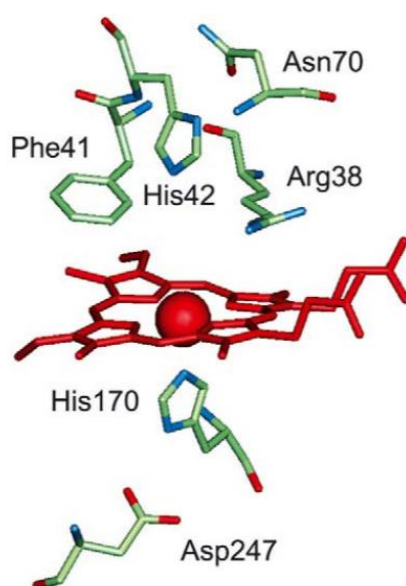


Рисунок 3. Строение активного центра фермента [1]

Структура, качество которой оценивалось в настоящей работе, была опубликована в 2002 году с целью исследовать влияние точечных замен в области кармана для гема на каталитический цикл [2]. В настоящее время механизм работы пероксидазы хрена изучен достаточно хорошо, поэтому в современных исследованиях HRP имеет преимущественно прикладное значение.

Результаты и обсуждение

Общая информация о модели.

Структура пероксидазы хрена С (HRP C) с PDB ID = 1KZM была опубликована на сайте PDBе 9 октября 2002 года группой ученых: Meno K., Jennings S., Smith A. T., Henriksen A. и Gajhede M..

Асимметрическая единица представлена мономером полипептидной цепи, состоящей из 308 аминокислотных остатков (полный белок состоит из 353 звеньев [3], однако в результате кристаллизации удалось установить положение лишь с 31 по 338 остатки), двух ионов Ca^{2+} , какодилат-иона и протопорфирина IX в комплексе с Fe (гема) [4, 5].

В качестве экспрессионной системы использовалась *Escherichia coli* [4].

Метод решения фазовой проблемы: молекулярное замещение структурой HRP C в комплексе с феруловой кислотой в качестве субстрата и без молекул воды (PDB ID = 6ATJ) [4].

Число измеренных рефлексов = 21040 [4].

Разрешение = 2.00 Å [4].

Полнота набора структурных факторов = 92.6 % [6].

Диапазон разрешений: 2.00 Å – 19.48 Å [6].

Параметры кристаллографической ячейки:

- длины направляющих векторов кристалла: $a = 40.4 \text{ \AA}$, $b = 66.7 \text{ \AA}$, $c = 117.1 \text{ \AA}$;
- углы между векторами: $\alpha = 90.0$, $\beta = 90.0$, $\gamma = 90.0$;
- кристаллографическая группа примитивная (P 21 21 21) с тремя осями винтовой симметрии 2_1 , то есть объект поворачивается на 180° и сдвигается на половину периода вдоль каждой оси;
- число молекул в ячейке = 4.

Асимметрическая ячейка совпадает с биологической и в ней отсутствуют некристаллографические симметрии.

Значения индикаторов качества модели в целом.

К таковым индикаторам относятся R-фактор, характеризующий соответствие построенной модели экспериментальным данным, и R_{free} , контроль которого не позволяет переоптимизировать модель. Для данной структуры их значения оказались 0.168 и 0.203 (по аннотации авторов) или 0.162 и 0.200 (по вычислениям программы DCC) соответственно. При этом доля структурных факторов, отобранных на вычисление R_{free} 4.71 % (990 рефлексов) по данным PDB файла. Однако wwPDB-VP сообщает, что эта доля

составляет 4.83 % (1017 рефлексов). R-фактор постоянно минимизируется в процессе оптимизации модели, и когда он становится меньше 0.25, модель уже можно считать достаточно приближенной к реальности. Однако часть рефлексов специально остается незадействованной в процессе оптимизации для осуществления контроля R_free. Качественная модель характеризуется одновременно низкими и близкими значениями R-фактора и R_free, причем последний индикатор всегда будет немного больше. В нашем случае R_free – R-фактор = 0.035 или 0.038, что составляет 17.2 – 19 % от R_free. Подобные значения удовлетворяют критериям качественной модели [7].

Теперь рассмотрим качество модели 1KZM по сравнению с другими моделями PDB по следующим показателям (рис. 4):

- R_free – мера соответствия модели небольшому набору экспериментальных данных, не использовавшихся в ее уточнении (вычисляется программой DCC);
- 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) [7];

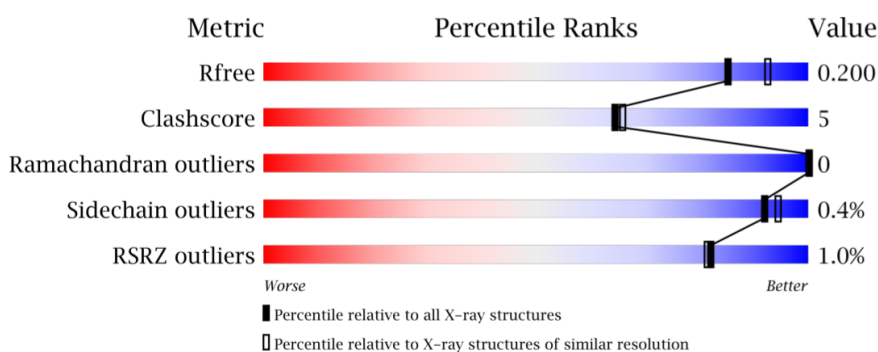


Рисунок 4. Сравнение качества модели 1KZM с другими моделями PDB [7]

Видно, что качество модели 1KZM в основном выше среднего среди других PDB структур. Лишь показатель Clashscore оставляет желать лучшего.

Исследуемая структура состоит из 2764 атомов (2419 атомов белковой цепи). Сервер MolProbity добавил 2364 атома водорода к полипептидной цепи для анализа показателя Clashscore. Всего было найдено 24 пары перекрытий Ван-дер-Ваальсовых радиусов, в 21 из которых участвуют аминокислотные остатки (суммарно 30 штук таких), а в 3 – только атомы кислорода молекул воды. В результате детального поиска маргинальных остатков, мне удалось обнаружить 34 таковых из 306 (длина белка составляет 308 а.о., однако 2 из них - Asn307 и Ser308 - реально отсутствуют в изучаемой структуре) и еще 2 остатка с подозрением на необходимость инверсии боковой цепи [7]. Данные приведены в таблице 1.

Критерий	Число маргинальных остатков	Процент маргинальных остатков
Длина связи	0	0
Углы связи	0	0
Хиральность	0	0
Планарность	0	0
Clashscore	30	9.80 %
Торсионные углы остова белка (карта Рамачандрана)	0	0
Торсионные углы боковых цепей остатков белка (ротамеры)	1	0.33 %
Соответствие модели "экспериментальным" данным (RSRZ)	3	0.98 %

Также было обнаружено 5 отклонений от нормы в длине связи и 3 отклонения – в углах связи в молекуле протопорфирина IX [7], но их я в дальнейшем не рассматривала.

На рисунке 5 приведено распределение найденных маргинальных остатков по цепи. Желтый цвет означает, что геометрия остатка не соответствует одному критерию, зеленый – геометрия остатка в пределах нормы, серый – остатки, не вошедшие в модель, красная точка – маргиналы по RSRZ.



Рисунок 5. Распределение маргинальных остатков [7]

Видно, что качество модели существенно снижено из-за большого количества абnormally близко перекрывающихся атомов. Однако тот факт, что большая часть остатков хорошо вписывается в экспериментальные данные электронной плотности и отсутствуют маргиналы по картам Рамачандрана (рис.б), позволяет предполагать, что структура расшифрована верно. Но, возможно, условия кристаллизации были таковы, что привело к «уплотнению» структуры и сближению ряда атомов. Могу предположить, что ионы какодилата натрия в растворе существенно увеличили его ионную силу, что гипотетически могло привести к наблюдаемой картине.

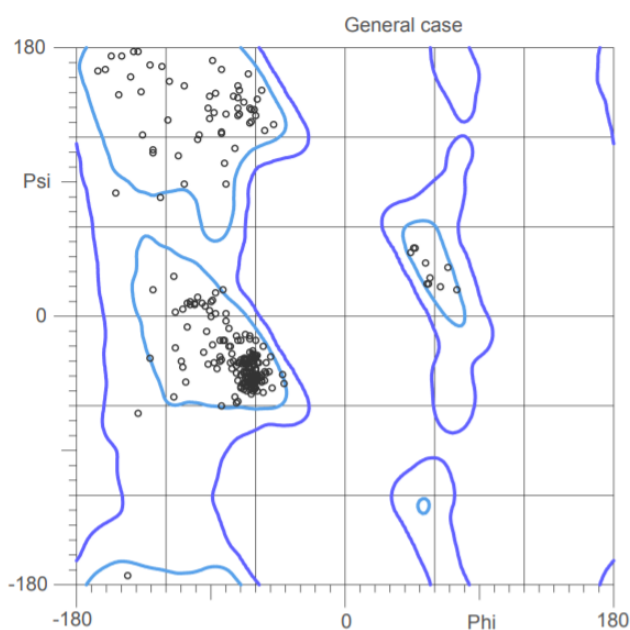


Рисунок б. Карта Рамачандрана для структуры 1KZM. Голубые линии ограничивают участки предпочтительных конформаций, фиолетовые – допустимых. Положение точек соответствует аминокислотным остаткам белка [8]

По карте Рамачандрана все остатки Pro оказались в trans-конформации [8].

Примеры маргинальных остатков.

Теперь перейдем к более частному рассмотрению найденных маргинальных остатков. Из их общего числа я выбрала по Clashscore те, у которых наибольшая длина перекрывания Ван-дер-Ваальсовых радиусов, а также все остальные маргинальные остатки, найденные по другим критериям. Результат выборки приведен в таблице 2. Интересно заметить, что Cys301 обладает двумя изоформами в структуре А и В с коэффициентами осцирапсу 0.52 и 0.41 (странно, что в сумме не дают 1.00), к тому же с данным остатком связано целых 5 пар перекрытий Ван-дер-Ваальсовых радиусов (в 3 из

которых участвует В-форма). Однако в PDB файле структуры авторы пояснили, что нормальная S-S связь между Cys97 и Cys301 была разрушена в результате облучения. Что также наводит на мысль о влиянии радиационного облучения в ходе эксперимента на качество структуры, в частности – на Clashscore [4].

Таблица 2. Перечень маргинальных остатков [7]	
Остаток	Показатель маргинальности
Cys97	Перекрытие Ван-дер-Ваальсового радиуса атома HG с HG атомом остатка Cys301 на 0.88 Å Перекрытие Ван-дер-Ваальсового радиуса атома SG с SG атомом остатка Cys301 на 0.55 Å.
Arg93	Перекрытие Ван-дер-Ваальсового радиуса атома HG2 с HD21 и CD2 атомами остатка Leu299 на 0.57 Å и 0.41 Å соответственно.
Phe172	Не ротамерная боковая цепь.
Asn16	Подозрение на необходимость инверсии боковой цепи Перекрытие Ван-дер-Ваальсового радиуса атома ND2 с O атомом 4001 молекулы воды на 0.40 Å.
Gln128	Подозрение на необходимость инверсии боковой цепи.
Gln271	Подозрение на необходимость инверсии боковой цепи.
Ser306	Плохое соответствие экспериментальной электронной плотности RSRZ = 3.3 > 2.
Ala74	Плохое соответствие экспериментальной электронной плотности RSRZ = 2.5 > 2.
Asn212	Плохое соответствие экспериментальной электронной плотности RSRZ = 2.1 > 2. Перекрытие Ван-дер-Ваальсового радиуса атома HB2 с O атомом остатка Leu211 на 0.47 Å.

Из приведенных остатков я выбрала 5 для детального рассмотрения: Cys97 в паре с Cys301, Phe172, Asn16, Ala74.

Анализ 5 маргинальных остатков.

1) Пара Cys97 и Cys301.

Рассмотрим сперва ситуацию с А-изоформой Cys301. По результатам работы сервера MolProbity после добавления атомов водорода в систему и оптимизации их расположения следующие пары атомов оказались расположенными слишком близко (при норме < 0.40 Å)[8]:

- Cys97 HG – Cys301 HG на 0.88 Å
- Arg298 HD2 – Cys301 HA на 0.43 Å

Визуализация же этих трех остатков показала, что атомы HG у цистеиновых остатков были добавлены ошибочно, так как атомы серы образуют дисульфидный мостик и, следовательно, не имеют водородные атомы при

себе (рис. 7). Вторая пара атомов водорода специально изображена сферами, чтобы показать перекрытие, однако оба остатка – Arg298 и Cys301 – принадлежат относительно подвижному участку вторичной структуры – петле. Учитывая этот факт и то, что их перекрытие немногим выше нормы, можно допустить, что расположение остатков приемлемо.

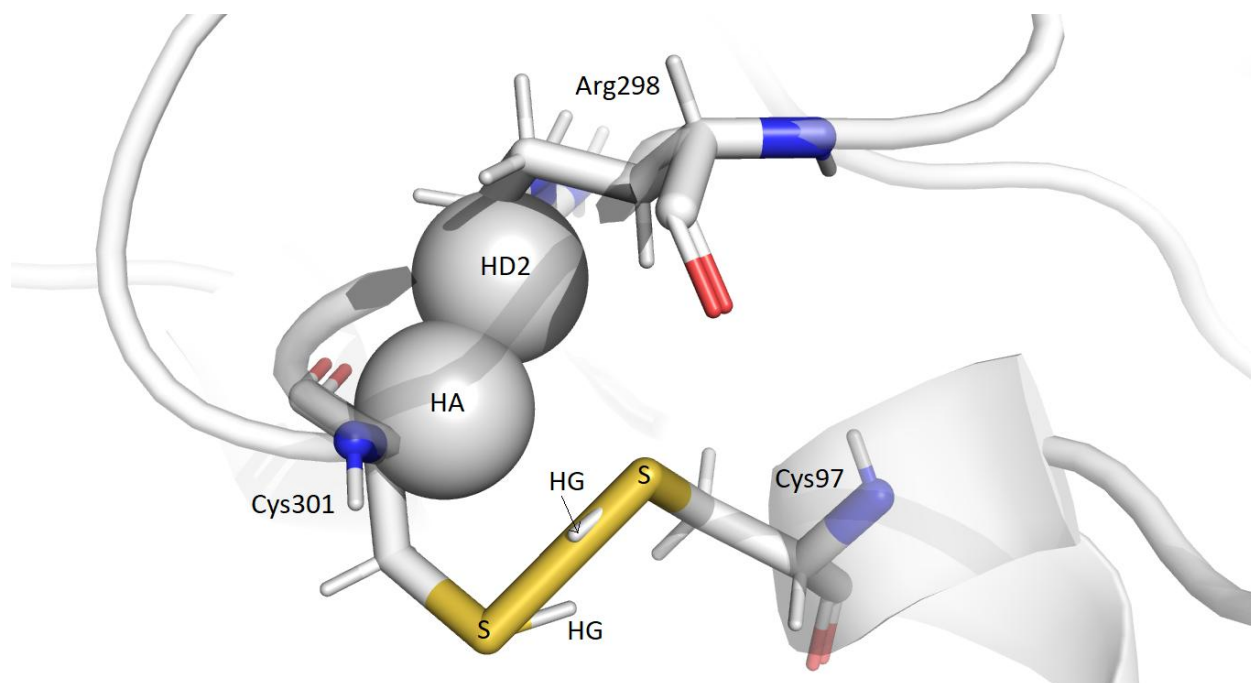


Рисунок 7. Геометрия аминокислотных остатков Cys97, Arg298 и изоформы A Cys301

Теперь обратимся к таинственной В-изоформе Cys301, которая появилась в результате разрушения S-S связи в ходе эксперимента. Кстати, то, что сумма коэффициентов осципации обеих изоформ не равна 1, говорит о незначительном присутствии иных конформаций 301 остатка в кристалле. В этом случае следующие пары атомов оказались расположенными слишком близко (при норме $< 0.40 \text{ \AA}$)[8]:

- Cys97 HG – Cys301 HB3 на 0.68 \AA
- Cys97 SG – Cys301 HB3 на 0.59 \AA
- Arg298 HD2 – Cys301 HA на 0.66 \AA

Как можно видеть на рисунке 8, положение 301 остатка сместилось, и его атом серы вывернулся в сторону поверхности белка (не показана). Это привело к еще большему наложению Ван-дер-Ваальсовых радиусов помеченных атомов. Тут стоит предположить, что, так как разрыв связи произошёл в уже закристаллизованном белке, то, вероятно, атомы серы после разрыва не успели бы найти в окружении лишние протоны для

образования связей. Поэтому считаю неуместных их тут изображать. Тогда получаем, что первое стерическое напряжение на самом деле отсутствует.

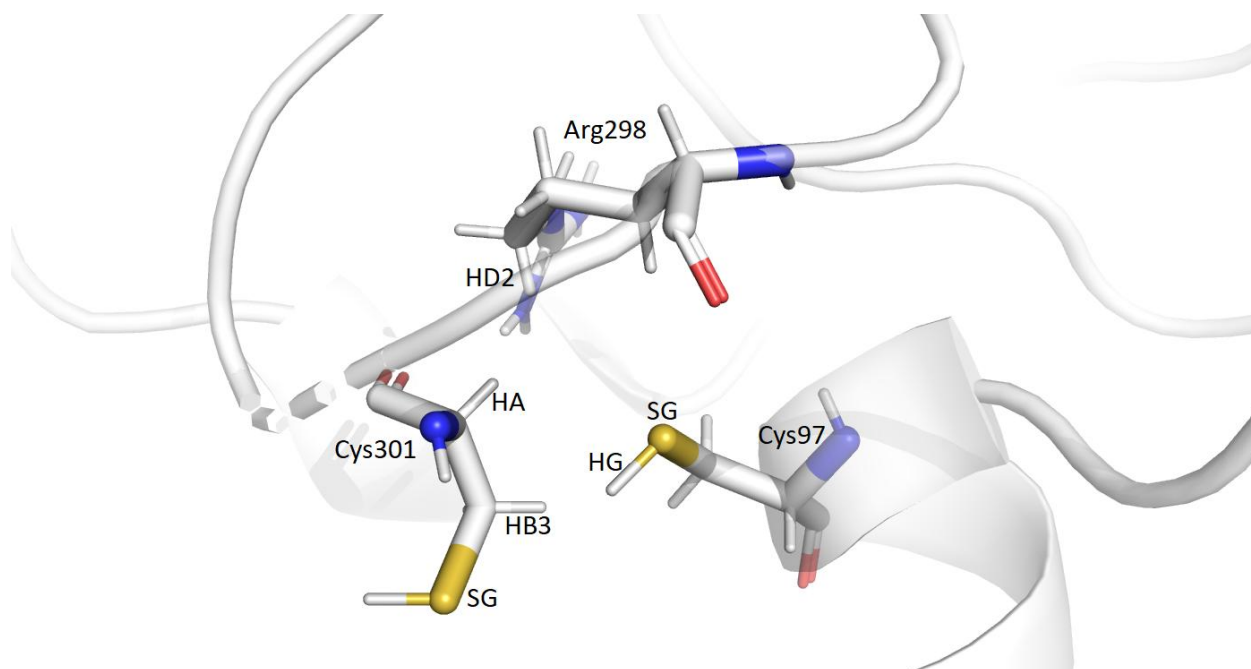


Рисунок 8. Геометрия аминокислотных остатков Cys97, Arg298 и изоформы B Cys301

Далее мне было интересно посмотреть, как в случае обеих изоформ остатки вписываются в карту электронной плотности (рис. 9).

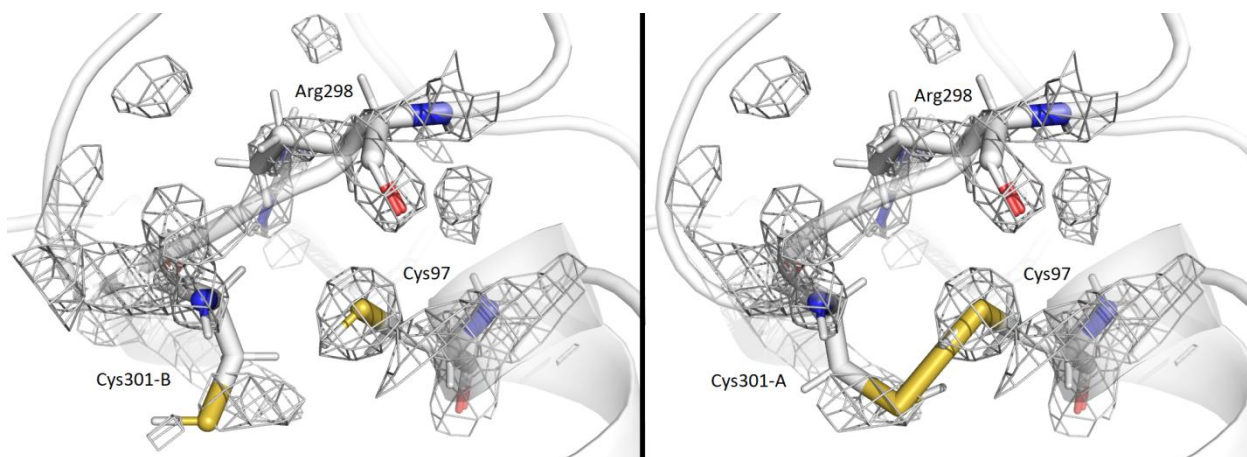


Рисунок 9. Электронная плотность вокруг остатков Cys97, Arg298 и Cys301 на уровне подрезки = 2σ

Оказалось, что, действительно, электронная плотность вокруг атома серы Cys301 V-образно «размазана» в пространстве с большим смещением в область A-изоформы. Из всего описанного вывод такой, что авторы очень точно построили модель, изоформа B не имеет биологического смысла и является экспериментальной ошибкой, программа сервера MolProbity

ошибочно добавила Н-атомы там, где они не уместны, из-за чего упало качество структуры по параметру Clashscore.

2) Phe172.

Теперь хотелось бы рассмотреть единственный неротамерный остаток в структуре. Оказалось, что двухгранный в его боковой цепи (отмечен на рисунке 10А) сильно превышает по модулю допустимые пределы. Я сделала выборку из 10 других остатков фенилаланина в данной структуре и измерила у них этот же двухгранный угол. Допустимые значения оказались в пределах от 90° до 120°.

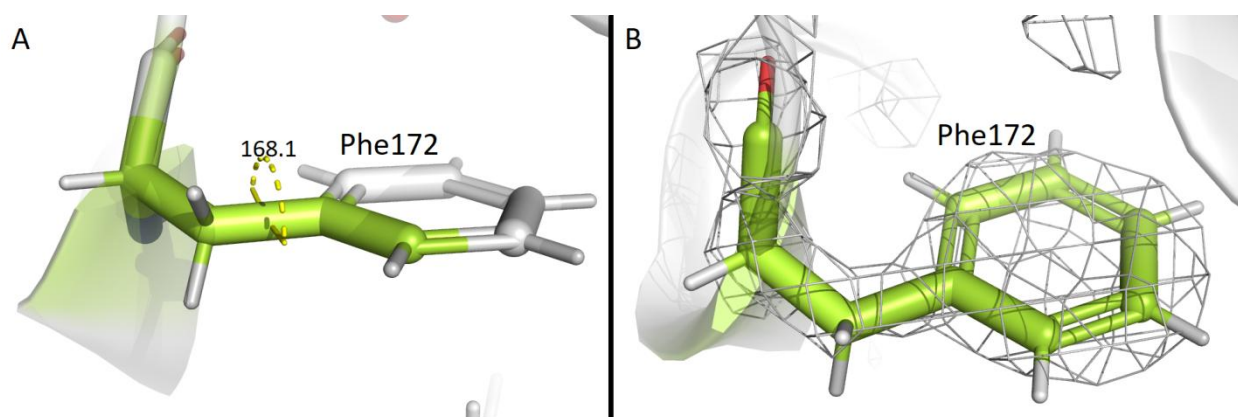


Рисунок 10. Остаток Phe172. А – с измеренным неротамерным двухгранным углом. В – вписанный в электронную плотность на уровне подрезки = 2 σ

Если же посмотреть на то, как остаток вписан в электронную плотность (рис. 10В), то нет никаких сомнений, что такая неротамерная конформация Phe172 является особенностью данного белка, а не ошибкой. Попытаемся привести разумное тому обоснование и посмотрим на ближайшее окружение белка (рисунок 11).

Оказывается, что Phe172 занимает крайне неудобное положение в пространстве. Его боковая цепь «зажата» между карманом для гемма и гидрофильной поверхностью. Видно, что его окружение образовано преимущественно гидрофобными остатками, формируя таким образом гидрофобное ядро. Также можно заметить, что Phe229 и Ala34 стерически мешают вращению плоскости ароматического кольца Phe172 в ротамерную область. Иначе это привело бы к увеличению Clashscore. Таким образом, еще один остаток из «подозрительных» оказался верно расшифрованным авторами. Интересно было бы дополнительно проверить, какая

конформация у этого остатка в других структурах HRP, но данная задача выходит за рамки настоящей работы.

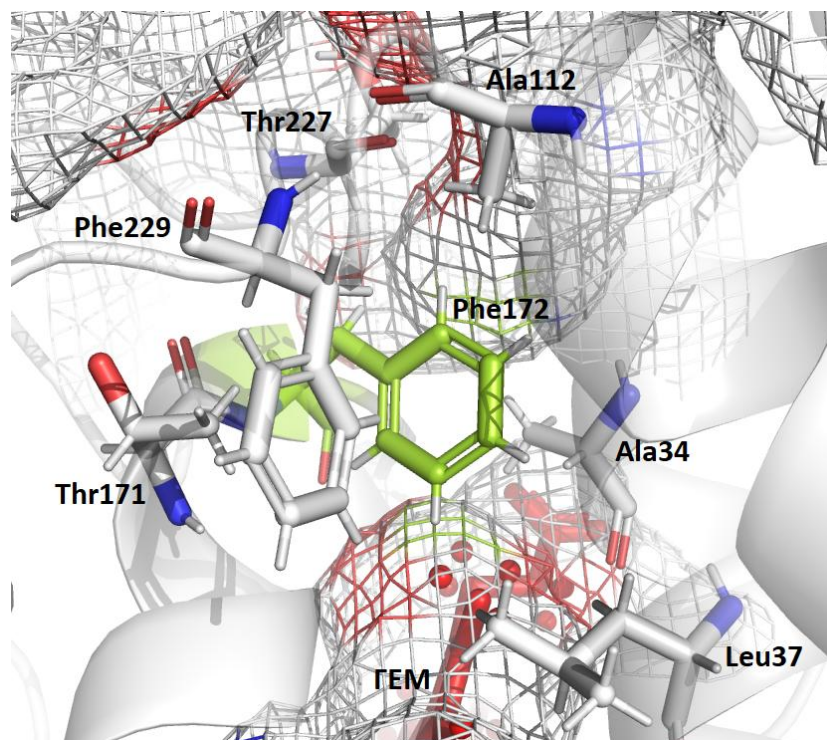


Рисунок 11. Ближайшее окружение Phe172. Сеткой изображена поверхность белка

3) Asn16.

Обратимся к аминокислотному остатку с подозрением на необходимость инверсии боковой цепи. С помощью сервера MolProbity я получила PDB файл, в котором N- и O-атомы боковой цепи Asn16 поменяны местами. На рисунке 12 представлено, как модели с (рис. 12B) и без (рис. 12A) инверсии вписываются в электронную плотность.

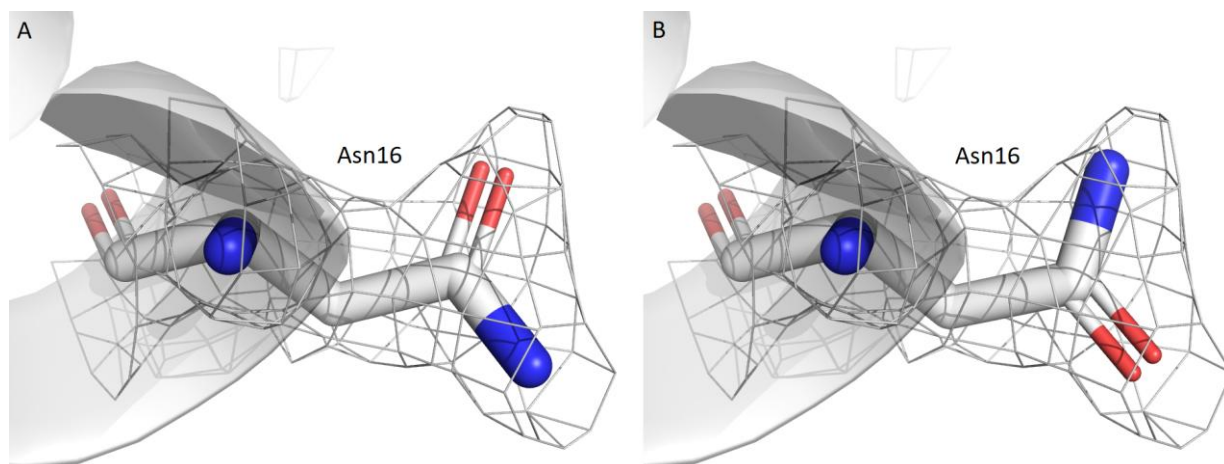


Рисунок 12. Электронная плотность вокруг Ans16 до (A) и после (B) инверсии на уровне подрезки = 1 σ

После визуализации кажется не понятным, почему авторы выбрали именно такое расположение атомов, ведь N-атом меньше O-атома, и потому электронное облако вокруг него будет немного меньше. На рисунке 12 верхняя область электронной плотности по диаметру меньше, чем нижняя. Следовательно, вариант с инверсией больше соответствует картине распределения электронной плотности. Тем более, что в опубликованной модели атом азота боковой группы оказывается слишком близко расположен по отношению к атому кислорода 4001 (вода). После инверсии эта проблема исчезает.

4) Ala74.

Разберем теперь подробно остаток с плохим соответствием экспериментальной электронной плотности ($RSRZ = 2.5 > 2$). Для этого посмотрим, как она покрывает Ala74 на разных уровнях подрезки (рис. 13).

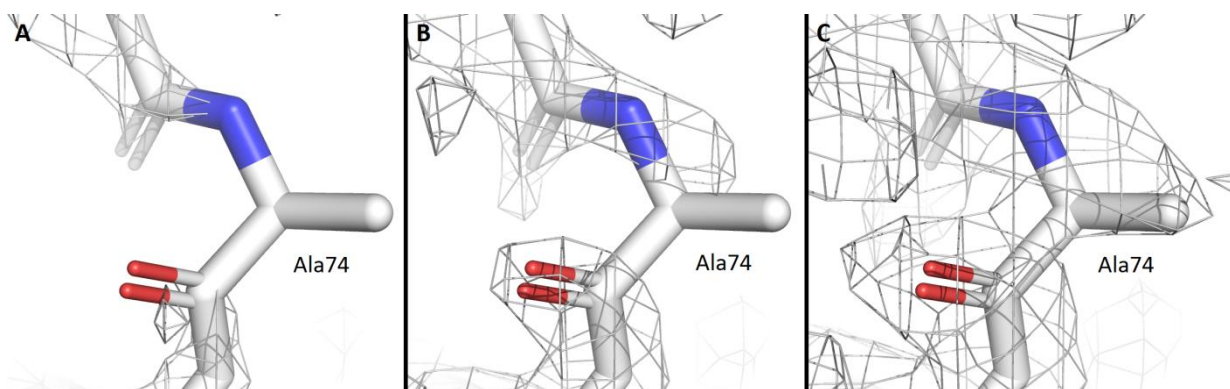


Рисунок 13. Электронная плотность вокруг Ala74 на уровне подрезки 1.5σ (A), 1.0σ (B) и 0.5σ (C)

Действительно, очень странно, что плотность вокруг боковой цепи остатка начинает появляться только на уровнях подрезки $< 1.0 \sigma$. Тем более, что соседние аминокислотные остатки хорошо описываются электронной плотностью. Единственное обоснование, которое приходит на ум, это подвижность Ala74. Дело в том, что он расположен в петле и не имеет каких-либо стерических «притеснений» со стороны окружающих его остатков. Тем не менее, наиболее частое положение Ala74 авторы установили.

Сравнение с моделью из PDB_redo.

База данных PDB_redo содержит оптимизированные версии существующих записей PDB с картами электронной плотности. В «улучшенной» версии структуры 1KZM было изменено положение 4 ротамеров и удалено 11

молекул воды (в том числе 4001, которая стерически мешала Asn16). В результате 12 остатков стали лучше вписываться в электронную плотность, а 4 – хуже. Также изменились общие показатели, R-фактор стал 0.150, а R_free стал 0.187. Z-score по картам Рамачандрана немного ухудшился после такой оптимизации, а по качеству ротамеров, наоборот, улучшился, причем на существенную величину (рис. 14) [9].

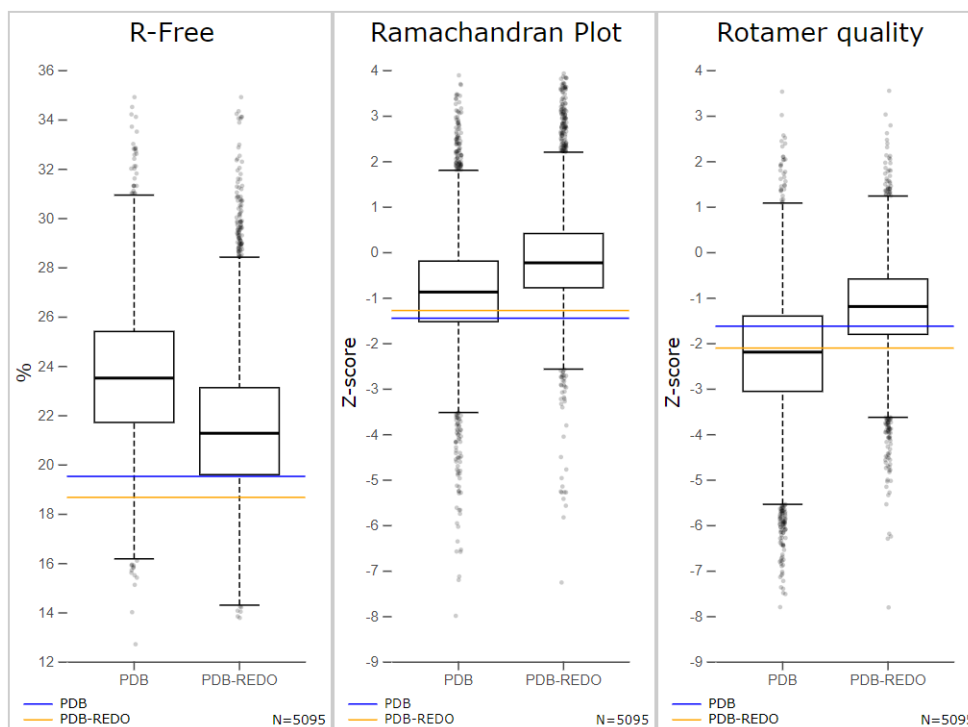


Рисунок 14. Сравнение общих параметров модели 1KZM до и после оптимизации [9]

Далее я скачала PDB файл с «улучшенной структурой» и дала его на вход программе сервера MolProbity. После добавления атомов водорода и анализа структуры по стандартным критериям качества получилась следующая картина (рис. 15)

All-Atom Contacts	Clashscore, all atoms:	9.44	84 th percentile* (N=715, 2.00Å ± 0.25Å)
	Clashscore is the number of serious steric overlaps (> 0.4 Å) per 1000 atoms.		
Protein Geometry	Poor rotamers	8	2.92% Goal: <0.3%
	Favored rotamers	258	94.16% Goal: >98%
	Ramachandran outliers	0	0.00% Goal: <0.05%
	Ramachandran favored	296	97.37% Goal: >98%
	MolProbity score [^]	2.07	69 th percentile* (N=12522, 2.00Å ± 0.25Å)
	Cβ deviations >0.25Å	6	2.02% Goal: 0
	Bad bonds:	8 / 2532	0.32% Goal: 0%
Bad angles:	25 / 3477	0.72% Goal: <0.1%	
Peptide Omegas	Cis Prolines:	0 / 17	0.00% Expected: ≤1 per chain, or ≤5%
	Twisted Peptides:	1 / 313	0.32% Goal: 0

Рисунок 15. Оценка качества «улучшенной» структуры 1KZM из PDB_redo сервером MolProbity

Очень странно, но MolProbity утверждает, что состояние по картам Рамачандрана не ухудшилось, однако стало больше плохих ротамеров и появились отклонения от нормальных длин связей и углов. Думаю, не стоит всецело доверять оптимизированной модели из PDB_redo. Тем более, что оптимизация полностью автоматизирована, а, как мы уже убедились, выбранная структура нуждается в механической проверке «подозрительных» аминокислотных остатков, особенно в случае двух изоформ.

Заключение

В результате оценки качества расшифровки структуры белка пероксидазы хрена, полученной методом рентгеноструктурного анализа (РСА), с PDB ID = 1KZM я пришла к заключению, что в целом аминокислотные остатки очень хорошо вписаны в карту электронной плотности. Однако мне удалось выявить один остаток, который точно нуждается в инверсии боковой цепи. Конформация единственного неротамерного остатка является особенностью его окружения. А изоформу В Cys301 по-хорошему надо вообще убрать из модели, так как она возникла из-за ошибки эксперимента и не несет никакого биологического смысла. Более того, она существенно снижает качество структуры. Еще стоит решить проблему того, что сервер MolProbity почему-то добавляет атомы водорода к атомам серы, образующим S-S связь, из-за чего качество модели тоже рассчитывается неверно. Что же касается большого числа остатков с плохим Clashscore, то стоит провести сравнение с другими аналогичными структурами в PDB. Возможно, для белка HRP подобные показатели – норма, или же повлияли условия кристаллизации или рентгеновское облучение, повлекшее разрыв S-S связи. Тем не менее, я довольна качеством расшифрованной структуры.

Список литературы

1. Veitch, N. C. (2004) Horseradish peroxidase: a modern view of a classic enzyme, *Phytochemistry*, 65, 249-259.
2. Meno, K., Jennings, S., Smith, A. T., Henriksen, A. and Gajhede, M. (2002) Structural analysis of the two horseradish peroxidase catalytic residue variants H42E and R38S/H42E: implications for the catalytic cycle, *Acta Crystallographica Section D Biological Crystallography*, 58, 1803-1812.
3. <https://www.uniprot.org/uniprot/P00433>
4. <https://www.ebi.ac.uk/pdbe/entry/pdb/1kzm>

5. <https://www.rcsb.org/structure/1kzm>
6. <http://eds.bmc.uu.se/cgi-bin/eds/uusfs?pdbCode=1kzm>
7. https://files.rcsb.org/pub/pdb/validation_reports/kz/1kzm/1kzm_full_validation.pdf
8. <http://molprobitry.biochem.duke.edu/>
9. <https://pdb-redo.eu/>