Отчет по качеству РСА расшифровки структуры лакказы *Coriolopsis caperata* (PDB ID: 4jhu)

(T2-depleted laccase from Coriolopsis caperata soaked with CuCl)

Оглавление

Аннотация	2
Введение	2
Результаты и обсуждение	3
Общая информация о модели	3
Значения индикаторов качества модели в целом	4
R-фактор	4
Диаграмма качества остатков	4
Анализ геометрии структуры	5
Карта Рамачандрана	7
Неротамерные остатки	7
Таблица со списком маргинальных остатков	8
Анализ маргинальных остатков и гетеромолекул	9
ASP 205	9
ILE 220	10
МЕТ 261 и НОН 736	11
PRO 285 и ASP 284	12
PRO 294	13
Сравнение модели из PDB с моделью из PDB redo	13
Заключение	15
Список использованной литературы и веб-сервисов	16
Список используемых программ	16

Аннотация

В данной работе был проведен анализ структуры лакказы (4jhu). Эта структура создавалась для того, чтобы понять, как устроен T2(Cu)-центр данного белка.

В отчете рассмотрены характеристики качества этой структуры, приведенные в отчете о расшифровке из PDB [2], проанализирована геометрия структуры с помощью сервиса MolProbity [3], рассмотрены более внимательно 5 остатков, которые показывали плохие значения при анализе качества структуры, и проведено сравнение структуры 4jhu из базы данных PDB с ее оптимизированной версией из базы данных PDB_redo.

Введение

Лакказы относятся к классу оксидаз, они окисляют различные органические (полифенолы, естественный субстрат - клеточная стенка растений) и неорганические (ферроцианид, удобно изучать), используя в качестве акцептора электронов молекулу кислорода [1].

Это лакказа гриба *Coriolopsis caperata*, данный фермент имеет 4 координированных иона меди I, которые образуют сайты связывания субстрата трех типов: T1, T2 и T3 (изображения можно видеть на сайте: <u>http://kodomo.fbb.msu.ru/~potanina.darya/term7/pr1/pr1_term7.html</u>). По каким-то причинам получить кристалл лакказы со всеми 4мя ионами меди сразу не получалось, поэтому был получен кристалл с 3мя ионами меди, который затем вымачивался в растворе CuCl, и из новоприсоединенных ионов меди был выбран тот, который вероятнее всего координирован не случайным образом, а в T2 сайте (см рис 1) [1].



Рисунок 1. Общая структура лакказы *С. caperata*: (a) T2-depleted структура, (b) T2depleted+Cu⁺ структура. В соответствии с доменной организацией белковая цепь окрашена разными цветами. Ионы меди показаны в виде красных сфер [1].

Результаты и обсуждение

Общая информация о модели

Таблица 1. Общая информация о модели.

РDВ код	4jhu
Название структуры	T2-depleted laccase from Coriolopsis caperata soaked with CuCl
Состав комплекса	Молекула лакказы (моносубъединичный фермент) в присутствии ионов Cu ²⁺ , тетраэтиленгликоля и N-ацетил-D-глюкозамина.
Год	Deposited: 2013
	Released: 2014
Авторы	Polyakov, K.M.,
	Fedorova, T.V.,
	Glazunova, O.A.,
	Kurzeev, S.A.,
	Maloshenok, L.G.,
	Koroleva, O.A.
Статья [1]	DOI: 10.2210/pdb4JHU/pdb
Метод решения фазовой проблемы	Молекулярное замещение
	The T2D laccase structure resolution was solved by molecular replacement using MOLREP (Vagin & Teplyakov, 2010) from the CCP4 package (Winn et al., 2011) [1].
Число измеренных рефлексов [2]	49260
Разрешение [2]	1.89Å
Полнота набора структурных факторов [2]	98.5%
Диапазон разрешений структурных факторов [2]	1.89Å – 19.36Å
Параметры кристаллографической ячейки [2]	a: 62.9 Å b: 84.42 Å c: 116.12 Å

	α: 90° β: 90° γ: 90°
Кристаллографическая группа [2]	P 21 21 21
Наличие некристаллографических симметрий в асимметрической ячейке [2]	нет

Значения индикаторов качества модели в целом

Ниже представлена диаграмма, отражающая значения некоторых параметров качества структуры [2]. Как можно видеть, все параметры модели находятся в областях оптимальных или допустимых значений. В следующих разделах будут более подробнее рассмотрены некоторые из них.



сторону, синий – в лучшую.

R-фактор

R-value = 0.17 (DCC, EDS) [2]

 $R_{\text{free}} = 0.21 \text{ (DCC) } [2]$

Хорошее значение R_{free} — меньше 20%. В данном случае 21% - я считаю это вполне качественный результат. Свободный R-фактор должен не более чем на 10% превышать R-фактор, иначе это свидетельствует о переобученнии. В данном случае разница составляет 4%, значит восстановление достоверное.

Диаграмма качества остатков

В PBD full report [2] представлен общий график по качеству всех остатков в структуре (см рисунок 3). Как можно видеть, в структуре присутствуют outlier'ы по двум параметрам: D142, M261, E308. Остатков, плохо вписанных в электронную плотность значительно больше, но интерес представляют остатки D284, Q362, тк они одновременно являются outlier'ами с точки зрения геометрии.



Рисунок 3. Нижняя строка показывает последовательность белка, где все позиции аннотированы с точки зрения качества воссозданной геометрии и электронной плотности. Верхняя строка показывает суммарное соотношение остатков разного качества. Цвет остатков свидетельствует о том, по скольки параметрам они являются outlier'ами: зеленый = 0, желтый = 1, оранжевый = 2, красный = 3 и более. Красные точки поставлены над теми остатками, которые плохо вписываются в электронную плотность (RSRZ > 2) [2].

Анализ геометрии структуры

Ниже представлена таблица 2, которая содержит некоторые результаты анализа контактов между атомами и геометрии всей молекулы белка с помощью сервиса MolProbity [3]. Перед началом работы с помощью встроенных инструментов сервера [3] в структуру были добавлены атомы водорода и исправлена инверсия боковой цепи ASN 51.

Среди данных, представленных в таблице следующие.

ClashScore — число недопустимых наложений электронных облаков атомов (> 0.4 Å) на 1000; перцентиль данной структуры по отношению к структурам примерно такого же разрешения [4]. В данном случае получился 99й перцентиль, то есть с точки зрения наложения электронных облаков атомов структура расшифрована очень хорошо.

Poor rotamers — число остатков с маргинальными по отклонению от ротамеров боковыми цепями [4]. В данном случае 7 из 496 (1.65%). Сервис [3] считает это плохим значением. Интересно, что в отчете PDB [2] указано 8 неротамерных остатков.

Favored rotamers – число и процент "хороших" ротамеров. Судя по значению, в данной структуре 398 ратамеров с оптимальными углами, 91 – с субоптимальными и 7 с неоптимальными. При этом % "хороших" ротамеров, выданный программой – 94.09%, не очень понятно откуда взялась эта цифра, но ладно.

Ramachandran outliers — полные маргиналы по карте Рамачандрана, лежат вне допустимой области [4]. Здесь получилось 0.2% (1 из 496). Сервис [3] считает это средним значением. В отчете PDB [2] также указано, что в данной структуре есть один остаток с неоптимальными углами по карте Рамачандрана (это ASP205).

Ramachandran favored — число и процент остатков в предпочитаемой области [4]. То есть 477 из 496 остатков имеют оптимальные углы по карте Рамачандрана, 18 – субоптимальные и 1 неоптимальные. По мнению сервиса [3] это среднее значение, но, судя по структурам,

которые анализировали мои коллеги, обычно outlier'ов по карте Рамачандрана сильно больше, поэтому я считаю этот результат хорошим.

MolProbity score — интегральная оценка структуры по данным этого сервиса [3][4]. Структура попадает в 97й перцентиль, что свидетельствует о хорошем качестве расшифровки.

С β deviations >0.25Å — число С β -атомов с неприемлемым отклонением от ожидаемого положения [4]. В данной структуре оказался 1 атом с неоптимальным положением С β -атома.

Bad backbone bonds — число ковалентных связей, существенно отклоняющихся от теории [4]. 4 из 3969 считается плохим значением.

Bad backbone angles — число валентных углов, существенно отклоняющихся от теории [4]. 15 из 5449 тоже считается плохим значением.

Все контакты между атомами	Clashscore, все атомы:	2.38		99 th перцентиль [*] (N=771, 1.89Å ± 0.25Å)
	Poor rotamers	7	1.65%	Оптимально: <0.3%
	Favored rotamers	398	94.09%	Оптимально: >98%
Геометрия белка	Ramachandran outliers	1 0.20%		Оптимально: <0.05%
	Ramachandran favored	477	96.56%	Оптимально: >98%
	MolProbity score [^]	1.41		97 th перцентиль [*] (N=11926, 1.89Å ± 0.25Å)
	Cβ deviations >0.25Å	1	0.22%	Оптимально: 0
	Bad bonds:	4 / 3969	0.10%	Оптимально: 0%
	Bad angles:	15 / 5449	0.28%	Оптимально: <0.1%

Таблица 2. Анализ структуры с помощью сервиса MolProbity [3].

Из всего вышеперечисленного можно сделать вывод, что параметры структуры неидеальны с точки зрения сервиса MolProbity [3], однако среди других структур такого же разрешения данная структура расшифрована весьма качественно, хотя и имеет некоторые проблемы с геометрией.

Рассмотрим теперь некоторые другие параметры.

Карта Рамачандрана

Как можно видеть на рисунке 4, подавляющее большинство остатков в данной структуре лежит в области предпочтительных значений, некоторое количество в области допустимых, и один на самой границе допустимых.



Рисунок 4. Карта Рамачандрана для всех остатков 4jhu, полученная с помощью сервиса MolProbity [3]. Темно-синим выделены области допустимых значений углов ф и ψ, светлосиним – наиболее предпочтительные.

В таблице 2 представлен один outlier по карте Рамачандрана, взятый из PBD full report [2]. Видимо именно этот остаток обозначен точкой, находящейся в правом верхнем углу на границе допустимой области.

Таблица 2. Маргиналы по карте Рамачандрана, согласно PBD full report	ort [2]].
--	---------	----

Mol	Chain	Res	Туре
1	А	205	ASP

Неротамерные остатки

В PBD full report [2] указано 8 неротамерных остатков (см таблицу 3), то есть в процентном соотношении из 496 остатков 1.61% неротамеров.

Mol	Chain	Res	Туре
1	А	120	LEU
1	А	242	ARG
1	А	261	MET
1	А	284	ASP
1	А	297	GLN
1	А	349	PRO
1	А	447	РНЕ
1	А	460	РНЕ

Таблица 3. Неротамерные остатки, согласно PBD full report [2].

Таблица со списком маргинальных остатков

Согласно выдаче сервиса MolProbity [3] и с уточнениями в PBD full report [2] была составлена таблица 4 со списком маргинальных остатков и критериями, по которым была определена их маргинальность. Таблица по сравнению с выдачей сервера сильно сокращена — в ней представлены только outlier'ы и остатки в «субоптимальной области» по более, чем одному критерию. И по каждому критерию выбрано не более одного остатка, то есть, если в данной структуре найдено семь неротамерных остатков, то в таблице 4 будет представлен только один из них — с самым выбивающимся значением.

Таблица 4. Список маргинальных остатков с указанием критерия маргинальности, согласно MolProbity [3] и PBD full report [2]. Серым отмечены остатки, рассмотренные в следующем разделе.

Nº	Остаток	Критерий	ритерий						
		Пересечение электронной плотности > 0.4 Å	Ramachandran outliers	Неротамерные остатки	RSRZ	Сβ в неправильном положении	Угол связи	Длина связи	Ис то чн ик
A 142	ASP							Неоптимальная длина связи между Сұ атомом и Об1 (карбоксильной группы)	[3]
A 205	ASP		Маргинал 0.02% Нормальное значение Pre-Pro = -156.0 и -70.4						[3] [2]
A 220	ILE	0.45Å перекрывания HB с CD2 133HIS	Отклонение 1.09% Нормальное значение						[3]

			Ile = -135.0 и -3.1					
A 261	MET	0.54Å перекрывания НЕ2 с О 736НОН		Отклонение 0.4% типичные χ углы: 183.4,191.3,14.2				[3]
A 285	PRO		Отклонение 1.91% Нормалное значение для Trans-Pro = -70.6 и 57.9	Отклонение 0.5% Типичные <i>Сү_endo</i> <i>ху</i> тлы: 2.6,15.4,333.2	3.5			[3] [2]
A 294	PRO				4.4			[2]
A 308	GLU					Маргинал по углу между Сү-Сб-Оє1	Маргинал по длине связи между Сб-Оє1	[3]
A 349	PRO			Маргинал 0.1% типичные χ углы: 320.8,44.6,327.3				[3]
A 447	PHE			Маргинал 0% типичные χ углы: 234.1,345.4				[3]
A 736	НОН	0.54Å перекрывания О с НЕ2 261МЕТ						[3]

Анализ маргинальных остатков и гетеромолекул

ASP 205

Данный остаток указан как маргинал по карте Рамачандрана [2][3]. По пептидной цепи ASP 205 идет перед пролином, и для него должны быть характерны значения углов φ и ψ : -156.0 и -70.4. Визуализацию данного остатка в РуМОL можно видеть на рисунке 5. В электронную плотность этот остаток вписан хорошо (см рисунок 5). Странно, что боковая группа этой аспарагиновой кислоты не образует никаких контактов с атомами окружения остатка на расстоянии 8 Å (см рисунок 6), хотя рядом от него есть атом Cu, но он закрыт от аспартата 205 изолейцином 452. Затрудняюсь сказать, почему этот остаток имеет неоптимальные углы, я не вижу причин этого, возможно, это особенность белка.



Рисунок 5. ASP 205. Уровень подрезки 1.5, буфер 0.5.



Рисунок 6. Окружение ASP 205. Аспартат изображен в виде sticks, окружение — в виде lines. Желтыми пунктирными линиями показаны полярные контакты.

ILE 220

Данный остаток определен как маргинальный по двум параметрам: 0.45Å перекрывания электронной плотности атома HB с атомом CD2 133HIS; отклонение 1.09% от оптимальных углов по карте Рамачандрана (нормальные значения для Ile = -135.0 и -3.1). Причины маргинальности этого остатка визуализировать не получается, в электронную плотность он вписан хорошо (см рисунок 7), перекрывание электронной плотности с HIS 133 образовано атомом водорода, который был автоматически добавлен в структуру (см рисунок 8). Видимо, этот остаток не маргинален, а просто это особенность белка.



Рисунок 7. ILE 220. Уровень подрезки 1.5, буфер 0.3.



Рисунок 8. (не)Перекрывание электронной плотности ILE 220 и HIS 133. Уровень подрезки 1.5, буфер 0.3.

МЕТ 261 и НОН 736

Данный метионин был выбран для анализа, так как его электронная плотность HE2 атома этого остатка на 0.54Å перекрывается с атомом кислорода воды 736. И кроме того, данный остаток указан как "не совсем" ротамер, тк его углы на 0.4% отклоняются от типичных χ углов: 183.4, 191.3, 14.2.

При визуализация электронной плотности (см рисунок 9) оказалось, что данный остаток не очень хорошо вписан в нее. ЕС атом данного остатка не вписан вообще, но перекрывания электронной плотности с молекулой воды визуализировать не получилось (см рисунок 10). Скорее всего отклонение от ротамерных углов связано с низким качеством «изображения» в этой точке, то есть это ошибка расшифровки.



Рисунок 9. МЕТ 261, уровень подрезки 1.3, буфер 0.3.



Рисунок 10. (не)Перекрывание электронной плотности МЕТ 261 и воды 736, уровень подрезки 1.3, буфер 0.3.

PRO 285 и ASP 284

Это транс-пролин. Для него в отчете PDB [2] указан RSRZ = 3.5. Также замечено отклонение от ротамерных углов 0.5% (типичные углы: 2.6, 15.4, 333.2), и от оптимальных углов по карте Рамачандрана (для транс-пролина типичны углы: -70.6 и 57.9).

При визуализации электронной плотности этого остатка в PyMOL, соседние остатки изображены линиями, ЭП вокруг них не строится) оказалось, что он не очень хорошо в нее вписан. Причем соседние остатки тоже вписаны довольно плохо (см рисунок 11), остаток ASP 284 — неротамерный [3], он имеет отклонение 0.1% от оптимальных углов (268.8, 76.0). Здесь причина маргинальности определенно в ошибке расшифровки.



Рисунок 11. PRO 285 и ASP 284 (sticks), уровень подрезки 1.5, буфер 0.3. Соседние остатки изображены линиями, электронная плотность вокруг них не построена.

PRO 294

Данный остаток имеет RSRZ = 4.4. При визуализации электронной плотности становится понятно, что такой плохой Z-score пространственного R-фактора обусловлен полным отсутствием электронной плотности вокруг остова боковой цепи данного пролина (см рисунок 12). Маргинальность данного остатка является следствием его "выдуманности", видимо отсутствие данных об электронной плотности в данной области вынудило экспериментаторов "сочинить" его структуру.



Рисунок 12. PRO 294, уровень подрезки 1.5, буфер 0.3.

Сравнение модели из PDB с моделью из PDB_redo

База данных PDB_redo [5] содержит оптимизированные расшифровки из базы данных PDB.

Характеристики оптимизированной модели 4jhu [6] не очень существенно отличаются от исходной. Как можно видеть в таблице 5 и на рисунке 13, R-фактор и свободный R-фактор у оптимизированной модели ниже, чем у исходной, хотя разница между ними, равная примерно 4%, сохранилась. Кроме того, в данной структуре увеличился % остатков, хорошо вписывающихся в карты Рамачандрана и ротамерные углы, модель стала попадать в более высокие перцентили по этим параметрам среди структур с таким же разрешением.

Таблица 5. Сравнение характеристик качества расшифровки структуры 4jhu PDB и PBD_redo [6].

	PDB	PDB-REDO
R	0,1670	0,1551
R-free	0,2072	0,1895
Качество модели (п	ерценті	или)
Ramachandran plot appearance	30	37
Rotamer normality	55	67



Рисунок 13. Сравнение показателей качества моделей 4jhu из базы данных PDB и PDB_redo [6].

При анализе оптимизированной структуры сервисом MolProbity [3] мне было предложено заменить уже два инвертированных аспарагина: ASN 51 и ASN 226. Хотя по параметру clashscore новая структура попадает в 100й перцентиль, то есть атомы вписаны в электронную плотность идеально (раньше был 99й — тоже очень-очень хорошо), но параметры геометрии скорее ухудшились. Немного возросло количество хороших ротамеров и остатков с хорошими углами по карте Рамачандрана, но при этом в новой структуре оказалось уже 2 маргинала по карте Рамачандрана. Увеличилось количество bad bonds и незначительно упало количество bad angles, увеличилось количество число Сβ-атомов с неприемлемым отклонением от ожидаемого положения (см рисунки 14.1-14.2).

All-Atom	Clashscore, all atoms:	2.38		99 th percentile [*] (N=771, 1.89Å ± 0.25Å)			
Contacts	Clashscore is the number of serious steric overlaps (> 0.4 Å) per 1000 atoms.						
	Poor rotamers	7	1.65%	Goal: <0.3%			
	Favored rotamers	398	94.09%	Goal: >98%			
	Ramachandran outliers	1	0.20%	Goal: <0.05%			
Protein	Ramachandran favored	477	96.56%	Goal: >98%			
Geometry	MolProbity score [^]	1.41		97 th percentile [*] (N=11926, 1.89Å ± 0.25Å)			
	Cβ deviations >0.25Å	1	0.22%	Goal: 0			
	Bad bonds:	4 / 3969	0.10%	Goal: 0%			
	Bad angles:	15 / 5449	0.28%	Goal: <0.1%			
Peptide Omegas	Cis Prolines:	4 / 42	9.52%	Expected: ≤1 per chain, or ≤5%			
Additional validations	Chiral volume outliers	0/614					
Additional validations	Waters with clashes	9/341	2.64%	See UnDowser table for details			

Рисунок 14.1. Характеристики геометрии исходной модели 4jhu из базы данных PDB. Таблица получена с помощью сервиса MolProbity [3].

All-Atom	Clashscore, all atoms:	1.32		$100^{\text{th}} \text{ percentile}^* (\text{N}=771, 1.89\text{\AA} \pm 0.25\text{\AA})$		
Contacts	lashscore is the number of serious steric overlaps (> 0.4 Å) per 1000 atoms.					
	Poor rotamers	7	1.65%	Goal: <0.3%		
	Favored rotamers	401	94.80%	Goal: >98%		
	Ramachandran outliers	2	0.40%	Goal: <0.05%		
Protein	Ramachandran favored	482 97.57%		Goal: >98%		
Geometry	MolProbity score^	1.11		$100^{\text{th}} \text{ percentile}^* (N=11926, 1.89\text{\AA} \pm 0.25\text{\AA})$		
	Cβ deviations >0.25Å	4	0.87%	Goal: 0		
	Bad bonds:	8 / 3969	0.20%	Goal: 0%		
	Bad angles:	12 / 5449	0.22%	Goal: <0.1%		
Peptide Omegas	Cis Prolines:	4 / 42	9.52%	Expected: ≤ 1 per chain, or $\leq 5\%$		
Additional validations	Tetrahedral geometry outliers	1				
Additional validations	Waters with clashes	3/339	0.88%	See UnDowser table for details		

Рисунок 14.2. Параметры геометрии оптимизированной структуры 4jhu из базы данных PDB_redo [5]. Таблица получена с помощью сервиса MolProbity [3].

Я также попробовала наложить новую и старую структуры в PyMOL (см рисунок 15). Расхождений между ними визуально оказалось довольно мало, хотя они явно есть. В основном различия касаются положения и длин β-тяжей. Однако, очень хорошо, что в данных структурах практически полностью совпадают локации атомов Cu, ведь именно в выяснении сайта связывания T2-Cu и была цель создания данной структуры.



Рисунок 15. Наложение структур 4jhu из базы данных PDB (зеленый) и PDB_redo (голубой).

Общий вывод, который я могу сделать, что вторичная расшифровка структуры была необязательной. Оптимизированная структура немного лучше по ряду формальных параметров, однако проигрывает по качеству геометрии.

Заключение

Данная структура имеет хорошее разрешение — 1.89Å. С точки зрения геометрии остатков она не идеальна, однако большинство из них хорошо вписано в электронную плотность, то есть искажение геометрии может быть особенностью белка. В данной структуре мало маргинальных остатков.

В связи с ее хорошим качеством, я считаю, авторы статьи имели право делать по ней выводы о положении T2-Cu, то есть цель, с которой изучался кристалл данного белка, выполнена.

Хорошее качество структуры этого белка может быть связано с тем, что в природе гомологи этого белка встречаются, в частности, в необычном "кристаллическом состоянии", например в "ранце" "взрывающихся" термитов [7]. На мой взгляд, тот факт, что гомолог данного белка был обнаружен в виде естественного кристалла белка, говорит о уменьшенном риске получения большого количества артефактов при создании кристалла (раз он может кристаллизоваться в природе), и это повышает достоверность данной модели с точки зрения соответствия полученного в данном эксперименте белка и его строения в природе.

Список использованной литературы и веб-сервисов

[1] O. A. Glazunova, K. M. Polyakov, T. V. Fedorova, P. V. Dorovatovskii, and O. V. Koroleva, "Elucidation of the crystal structure of Coriolopsis caperata laccase: Restoration of the structure and activity of the native enzyme from the T2-depleted form by copper ions," *Acta Crystallogr. Sect. D Biol. Crystallogr.*, vol. 71, pp. 854–861, 2015.

[2] https://www.ebi.ac.uk/pdbe/entry-files/download/4jhu_full_validation.pdf

- [3] http://molprobity.biochem.duke.edu/
- [4] https://kodomo.fbb.msu.ru/wiki/Main/RSA/Validation
- [5] https://pdb-redo.eu/
- [6] https://pdb-redo.eu/db/4jhu

[7] Scharf, M. E., & Tartar, A. (2008). Termite digestomes as sources for novel lignocellulases. Biofuels, Bioproducts and Biorefining, 2(6), 540–552. doi:10.1002/bbb.107

Список используемых программ

PyMOL (TM) 2.3.3 (Open-Source PyMOL)