

Московский государственный университет
имени М.В. Ломоносова

Факультет биоинженерии
и биоинформатики

Глеб Гавриш, 402гр.

Отчёт по качеству PCA расшифровки структуры с
PDB ID 5TYB

Москва, 2018

0) Аннотация

В ходе работы был произведен анализ качества расшифровки PCA структуры белка pol μ (PDB ID: 5TYB).

1) Введение

DNA polymerase μ (pol μ) – это ДНК-зависимая полимераз, участвующая в нехомологичном исправлении двойного разрыва (Non-homologous end joining, NHEJ). Pol μ относится к X семейству полимераз, куда относятся другие полимеразы, отвечающие за репликацию и починку, такие как Pol β , Pol σ и Pol λ [1].

Было показано на мышах, что Pol μ играет важную роль в кроветворении, а также участвует в поддержании генетической целостности во всех тканях [2]. Далее я буду изучать структуру 5tyb (рис. 1), которая представляет собой Pol μ человека в комплексе с молекулой ДНК, выступающей в роли субстрата.

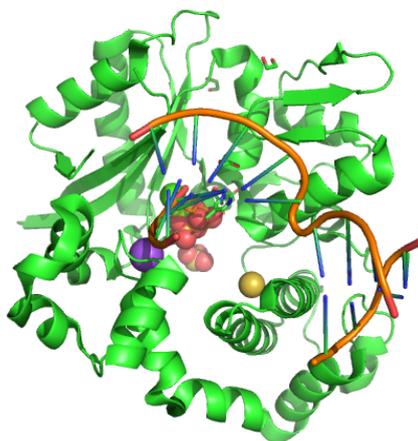


Рисунок 1. Pol μ человека из структуры 5tyb

2) Результаты

2.1) Общая информация о модели

Структура с PDB ID: 5TYB была загружена в PDB 30 августа 2017 коллективом лаборатории геномной целостности и структурной биологии, национальный институт здоровья, США (Jamsen JA, Beard WA, Pedersen LC, Shock DD, Moon AF, Krahn JM, Bebenek K, Kunkel TA, Wilson SH). Разрешение структуры – 1.85 Å. Комплекс состоит из цепи белка, двухцепочечной днк (с разрывом) и лигандов. Таким образом в структуре имеется 4 макромолекулы, из которых 3 - ДНК. Фазовая проблема была решена молекулярным замещением на основе структуры 4M04, для этого исследователи использовали программу PHENIX версии 1.11.1_2575. [3]

Всего в структуре 3631 атом, атомы водородов отсутствуют. Из 345 остатков в структуре, 17 являются остатками нуклеиновых кислот, остальные - аминокислот. 31 аминокислотный остаток имеет альтернативное положение.

Число измеренных рефлексов - 39477. Полнота набора структурных факторов составила 99.4%, диапазон разрешений - 31.39Å – 1.85Å. Параметры кристаллографической ячейки: длина направляющих векторов (в ангстремах): 60.05Å 68.49Å 110.45Å; все углы 90.00°.

2.2) Индикаторы качества модели

Для определения качества модели расшифровки РСА структуры обычно используют такие показатели, как R-фактор, R-free, и доли маргинальных остатков по различным параметрам, таким как: перекрытия Ван-дер-Ваальсовых радиусов, положение боковой цепи, а также RSR. На рис. 2 представлена информация о сравнении модели 5tyb с другими моделями из PDB.

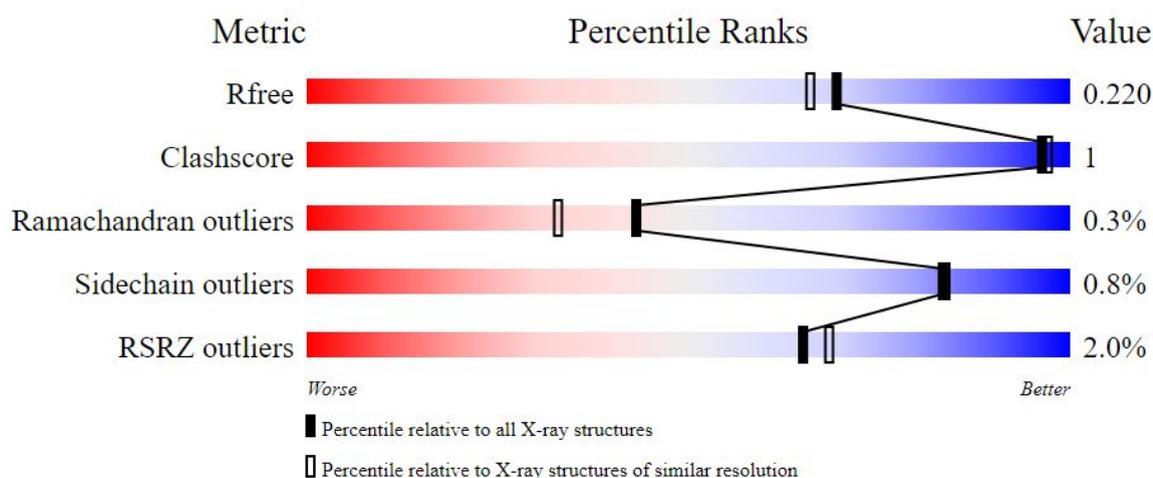


Рисунок 2. Основные показатели качества модели. Черным показаны значения относительно всех РСА моделей, белым - относительно моделей схожего разрешения. Сдвиг в красную область означает, что модель “хуже” аналогичных, а сдвиг в синюю - что модель “лучше”.

Подробнее рассмотрим каждый из пунктов:

R-фактор

R-фактор отражает насколько модель описывает экспериментальные данные. При значении R-фактора ниже 0.25 можно говорить о хорошей расшифровке структуры. Значение R-фактора выбранной мной модели составляет 0.169, что меньше 0.25. Таким образом, из значения R-фактора следует, что модель хорошо описывает экспериментальные данные.

R-free

При моделировании структуры опираются на значение R-фактора, стараясь его минимизировать. Таким образом часто возможна ситуация, что модель “переоптимизирован”, т.е. просто напросто подогнана под экспериментальные данные. Для избежания таких ситуаций используют показатель R-free, который считается аналогичным с R-фактором образом, однако для подсчета которого заранее случайным образом выбирают рефлексы (т.е. часть экспериментальных данных). Модель оптимизируют по оставшимся данным, а после этого считают R-free, т.е. R-фактор для рефлексов, не участвовавших в оптимизации. “Хорошим” читают значение R-free меньше 0.2, и если разница R-фактора и R-free меньше 0.1. В

полученной модели R-free очень близок к границе (значение R-free = 0.199), но все же подходит к ней с нижней стороны. Разница R-фактора и R-free составляет 0.03, что вполне приемлемо. Таким образом значение R-free скорее говорит о “хорошести” модели, и не может свидетельствовать о переоптимизации.

Таблица 1 Значение R-фактора и R-free структуры 5TYB

Индикатор качества	Значение
R-фактор	0.169
R-free	0.199
Доля структурных факторов, отобранных для R-free	5.0% (1975 рефлексов)

Карта Рамачандрана

Карта Рамачандрана (Ramachandran outliers) - распределение аминокислот по значениям двух торсионных углов (ϕ и ψ). Для подробного анализа этого параметра был использован сервис MolProbity [4], предназначенный для валидации молекул. Для структуры было добавлено 3007 атомов водорода, 80 из которых были скорректированы для улучшения водородных связей. Остатки His329 и His481 были переориентированы. Результаты выдачи представлены на рис. 3

All-Atom Contacts	Clashscore, all atoms:	1.12	100 th percentile* (N=865, 1.85Å ± 0.25Å)	
	Clashscore is the number of serious steric overlaps (> 0.4 Å) per 1000 atoms.			
Protein Geometry	Poor rotamers	3	1.03%	Goal: <0.3%
	Favored rotamers	280	96.55%	Goal: >98%
	Ramachandran outliers	1	0.31%	Goal: <0.05%
	Ramachandran favored	315	97.52%	Goal: >98%
	MolProbity score [^]	0.91	100 th percentile* (N=12654, 1.85Å ± 0.25Å)	
	C β deviations >0.25Å	0	0.00%	Goal: 0
	Bad bonds:	0 / 2881	0.00%	Goal: 0%
	Bad angles:	0 / 3943	0.00%	Goal: <0.1%
Peptide Omegas	Cis Prolines:	0 / 17	0.00%	Expected: ≤1 per chain, or ≤5%
	Cis nonProlines:	1 / 327	0.31%	Goal: <0.05%
Nucleic Acid Geometry	Bad bonds:	0 / 479	0.00%	Goal: 0%
	Bad angles:	0 / 732	0.00%	Goal: <0.1%

Рисунок 3. Индикаторы качества структуры 5tyb, определенные сервисом MolProbity [4]

Из выдачи следует, что есть 1 остаток маргинальный по картам Рамачандрана. Карта, построенная сервисом MolProbity для структуры 5tyb представлена на рис. 4. На рис. 5 представлены карты для отдельных остатков.

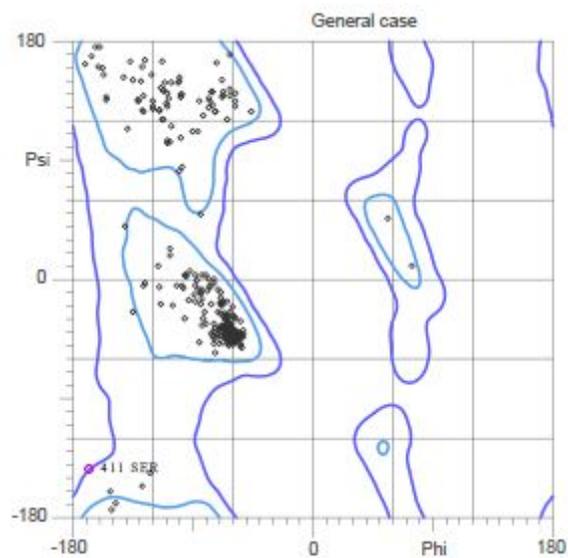
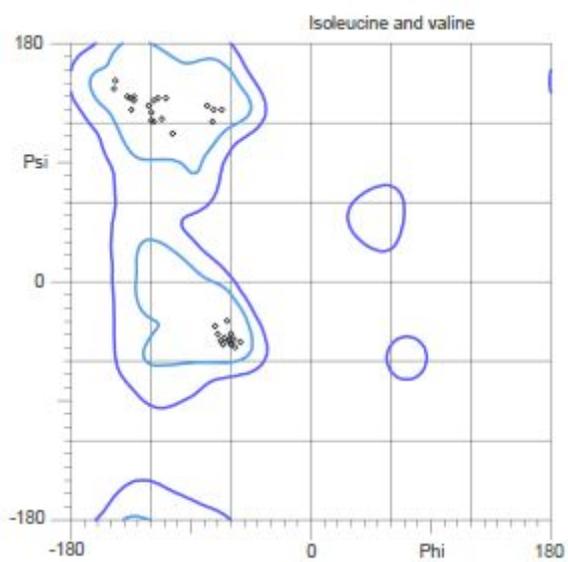


Рисунок 4. Карта Рамачандрана для всей структуры 3o48. Голубые линии ограничивают участки предпочтительных областей, синие – допустимых областей. Точками представлены аминокислотные остатки. Красным отмечен маргинальный остаток Ser411.



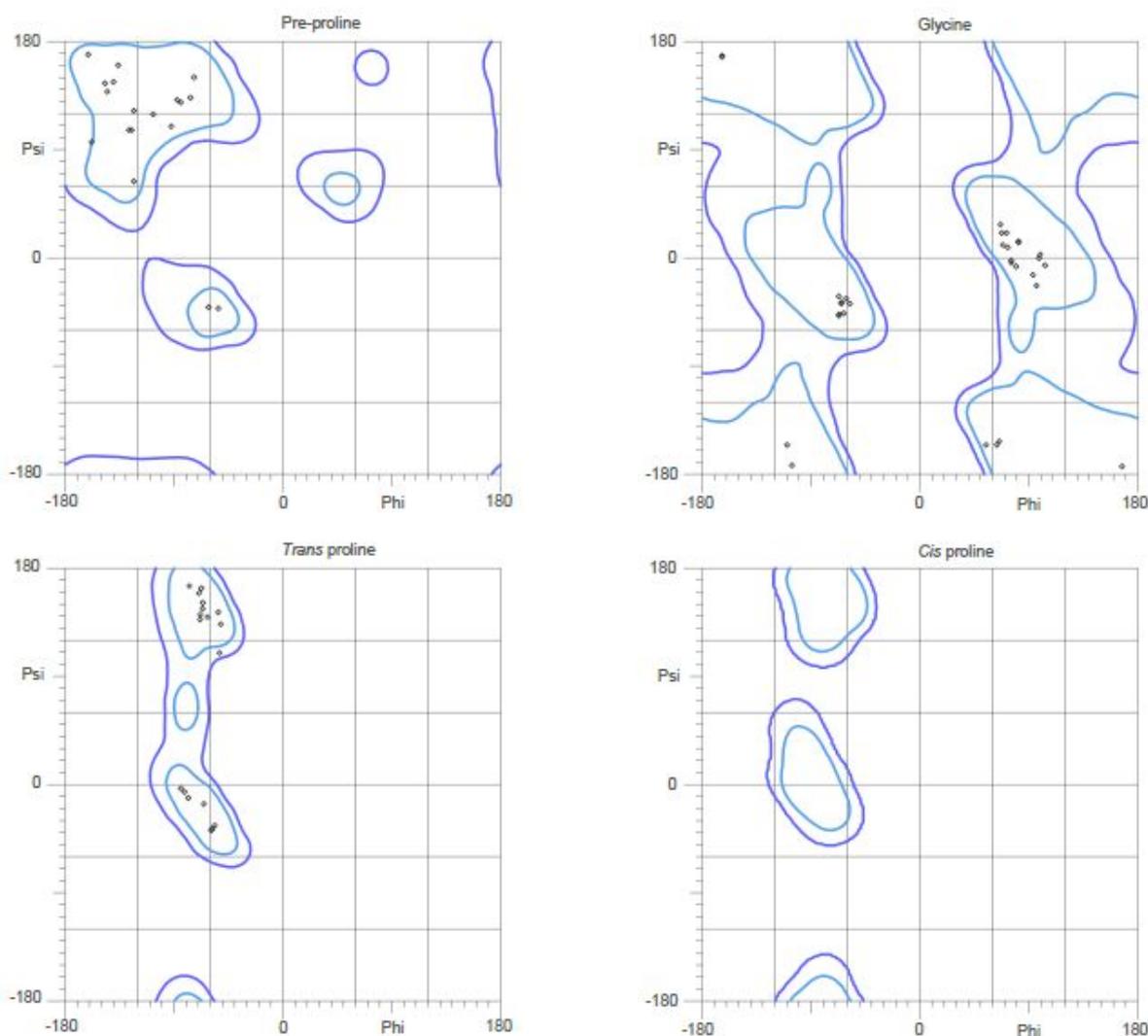


Рисунок 5. Карта Рамачандрана для отдельных остатков структуры 3o48. Голубые линии ограничивают участки предпочтительных областей, синие – допустимых областей. Точками представлены аминокислотные остатки.

97.6%% остатков лежат в предпочтительной зоне, 99.7%% - в разрешенной. Обнаружен 1 маргинальный остаток Ser411 со значениями углов (ϕ , ψ) = (-168.7, -144.1). Из анализа отдельных аминокислот примечательно то, что пролина в cis конформации найдено не было.

Ротамеры

Из выдачи MolProbity (рис. 3) следует, что есть 3 “плохих ротамера”. В структуре 5tyb доля плохих ротамеров составляет 1.03%, что заметно больше рекомендуемой нормы в 0.3%

Clashscore

Clashscore, или параметр перекрытия Ван-дер-Ваальсовых радиусов, равен количеству перекрытий Ван-дер-Ваальсовых радиусов больше, чем на 0.4 ангстрема, на 1000 атомов. На рис. 3 видно, он равен 1.12, что лучше чем почти у 100% PDB структур с разрешением $1.85\text{\AA} \pm 0.25\text{\AA}$. Таким образом по этому показателю структуру можно считать отличной.

RSR

Параметр RSR определяет на сколько каждый остаток вписывается в экспериментальные данные. Принято производить оценку в виде Z-score (RSRZ), которая показывает, насколько RSR для остатка в данной модели отличается от среднего значения RSR по структурам из PDB аналогичного разрешения. Показателем маргинальности считается RSRZ больше 2. Для модели 5 μ b значение RSRZ равно 2.0%, что чуть лучше, чем для средней модели PDB.

Для прочих параметров, таких как длина валентной связи или величина угла, отклонений обнаружено не было. Стоит отметить лишь величину Cis не пролиновых остатков, которая в относительном значении равна 0.31% при рекомендации в 0.05% (рис. 3). Реально же обнаружен лишь 1 такой остаток, а рекомендация 0.05% явно допускает отклонение от нуля, и большой процент обусловлен лишь небольшим суммарным числом остатков (327). Таким образом я бы заключил, что структура в целом неплохая.

2.3) Анализ маргинальных остатков

Значения RSRZ для остатков были получены сервисом EDS [5].

Таблица 2. Маргинальные остатки в структуре 5 μ b

Остаток	Обоснование маргинальности
Ser411	Карты Рамачандрана (нехарактерные углы (ϕ , ψ)=(-168.7, -144.1))
Trp137	RSRZ > 2 (RSRZ = 2.050)
His363	RSRZ ~ 2 (RSRZ = 1.963)
Phe385	RSRZ > 2 (RSRZ = 2.386)
Phe392	RSRZ > 2 (RSRZ = 2.133)

Проанализируем маргинальные остатки подробнее:

Ser411

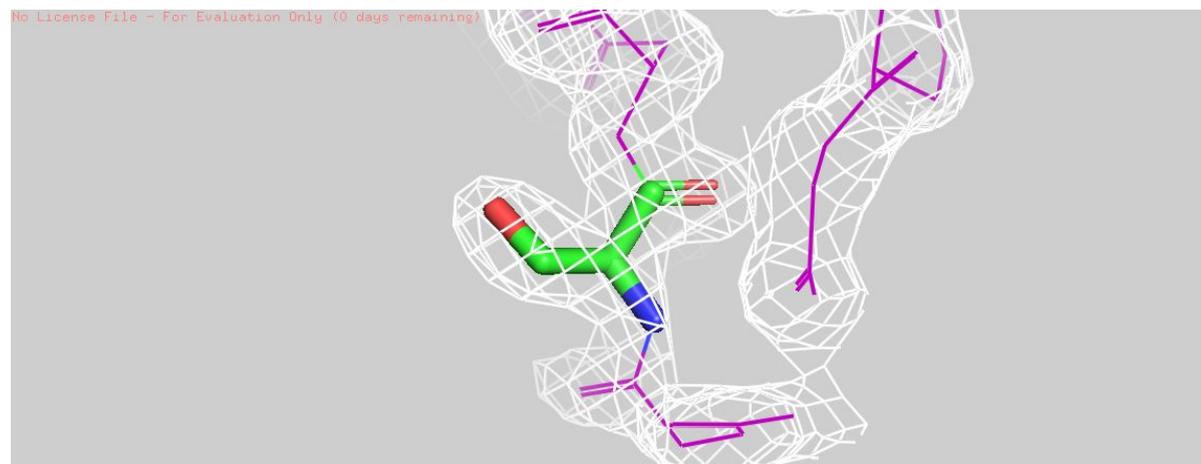


Рисунок 6. Остаток Ser411 с визуализированной электронной плотностью на уровне 1 σ .

Ser411 является маргинальным по картам Рамачандрана, т.е. он имеет нетипичные углы ϕ и ψ ($(\phi, \psi) = (-168.7, -144.1)$). Однако, судя по электронной плотности (рис. 6) это является явной особенностью белка, а не самой модели.

Trp137

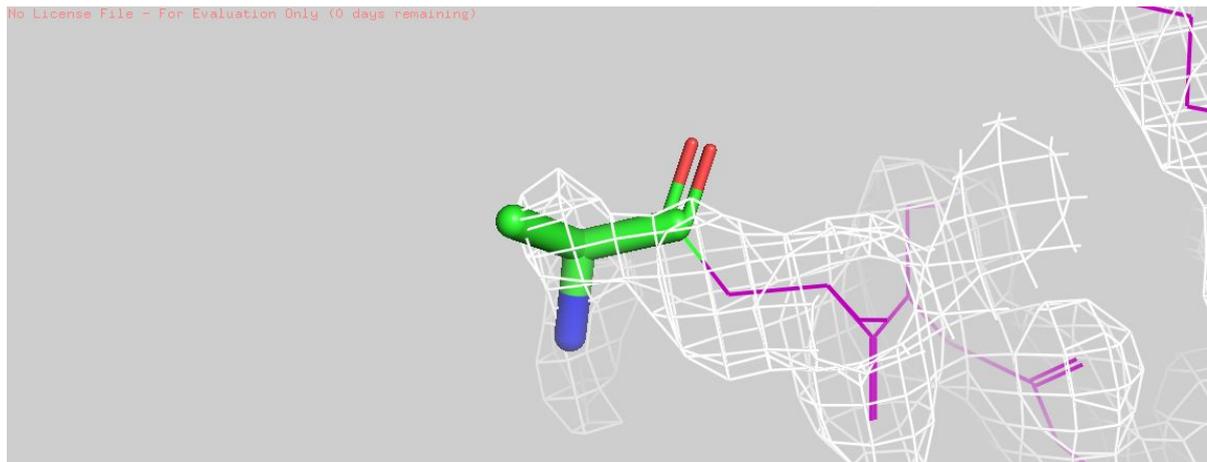


Рисунок 7. Остаток Trp137 с визуализированной электронной плотностью на уровне 1σ .

Trp137 является маргинальным по RSRZ ($RSRZ = 2.050$). Однако с данным остатком все далеко не очевидно, т.к. судя по электронной плотности и структуре (рис. 7) боковая цепь у него просто отсутствует. В файле PDB также указано, что для этого остатка отсутствуют почти все (!) атомы боковой группы. Возможно, именно этот остаток вообще не может нормально кристаллизоваться, и отсутствие электронной плотности вызвано большим количеством ротамеров для данного остатка.

His363

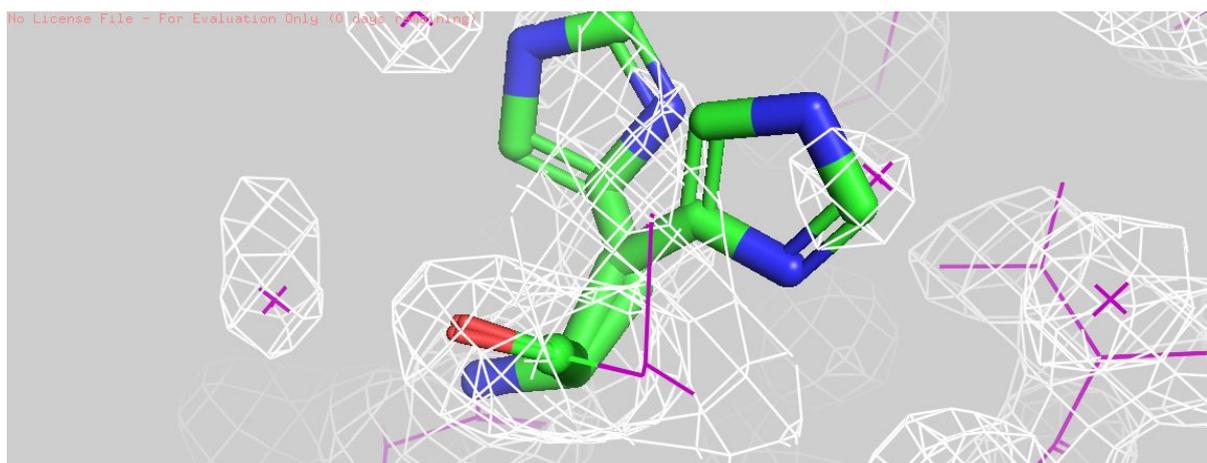


Рисунок 8. Остаток His363 с визуализированной электронной плотностью на уровне 1σ .

His363 является маргинальным по RSRZ ($RSRZ = 1.963$). Но видимо это может быть связано с наличием разных ротамеров (рис. 8). Я бы сказал что данный остаток описан нормально.

Phe385

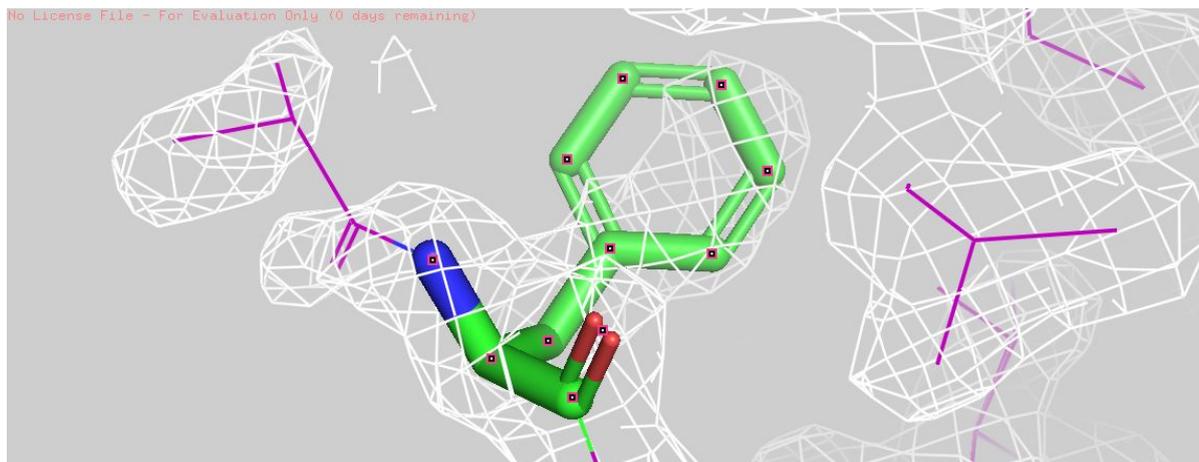


Рисунок 9. Остаток Phe385 с визуализированной электронной плотностью на уровне 1σ .

Phe385 является маргинальным по RSRZ ($RSRZ = 2.386$). Электронная плотность для него вполне нормальна, если не считать низкую плотность для двух атомов боковой цепи, которая не видна на уровне 1σ (рис. 9). Но по факту остаток в модели неплох.

Phe392



Рисунок 10. Остаток Phe392 с визуализированной электронной плотностью на уровне 1σ .

Phe392 является маргинальным по RSRZ ($RSRZ = 2.133$). Электронная плотность отлично описывает боковую группу, однако меня смущает наличие двух ротамеров в финальной версии. Я бы все же не стал приводить два различных положения, а привел бы лишь одно, как среднее - уж больно на маленький угол отличается позиции боковых групп для обоих случаев.

Так то единственный остаток, который я бы поправил - Phe392. Однако, возможно это лишь дело вкуса, так как на финальную структуру влияет очень слабо, но зато демонстрирует некую подвижность остатка, что может быть полезно для молекулярной динамики. Главное, чтобы оба положения были оправданы.

2.4) Сравнение модели из PDB с моделью из PDB_redo

Воспользовавшись сервисом PDB_redo [6] я получил “улучшенную” версию структуры 5tyb. Сравнительная таблица приведена на рис. 11; на рис. 12 приведена информация об изменениях,

внесенных в модель. На рис. 13 представлен вывод сервиса MolProbity [4] для структуры PDB_redo.

Validation metrics from PDB-REDO		
	PDB	PDB-REDO
Crystallographic refinement		
<i>R</i>	0,1701	0,1675
<i>R-free</i>	0,2025	0,1972
<i>Bond length RMS Z-score</i>	0,494	0,643
<i>Bond angle RMS Z-score</i>	0,752	0,985
Model quality (raw scores percentiles)		
<i>Ramachandran plot appearance</i>	91	86
<i>Rotamer normality</i>	87	73
<i>Coarse packing</i>	43	45
<i>Fine packing</i>	36	53
<i>Bump severity</i>	67	53
<i>Hydrogen bond satisfaction</i>	15	14
WHAT_CHECK	Report	Report

Рисунок 11. Результат сравнения PDB и PDB_redo на сайте PDB_redo [6]

Significant model changes	
Description	Count
<i>Rotamers changed</i>	2
<i>Side chains flipped</i>	15
<i>Waters removed</i>	12
<i>Peptides flipped</i>	0
<i>Chiralities fixed</i>	0
<i>Residues fitting density better</i>	2
<i>Residues fitting density worse</i>	1

Рисунок 12. Изменения внесенный в PDB_redo (информация с сервиса PDB_redo [6])

All-Atom Contacts	Clashscore, all atoms:	2.94		99 th percentile* (N=865, 1.85Å ± 0.25Å)
	Clashscore is the number of serious steric overlaps (> 0.4 Å) per 1000 atoms.			
Protein Geometry	Poor rotamers	7	2.30%	Goal: <0.3%
	Favored rotamers	288	94.43%	Goal: >98%
	Ramachandran outliers	0	0.00%	Goal: <0.05%
	Ramachandran favored	314	97.21%	Goal: >98%
	MolProbity score [^]	1.48		95 th percentile* (N=12654, 1.85Å ± 0.25Å)
	Cβ deviations >0.25Å	1	0.30%	Goal: 0
	Bad bonds:	4 / 2939	0.14%	Goal: 0%
	Bad angles:	7 / 4017	0.17%	Goal: <0.1%
Peptide Omegas	Cis Prolines:	0 / 17	0.00%	Expected: ≤1 per chain, or ≤5%
	Cis nonProlines:	1 / 327	0.31%	Goal: <0.05%
Nucleic Acid Geometry	Bad backbone conformations [#] :	9	52.94%	Goal: ≤ 5%
	Bad bonds:	0 / 479	0.00%	Goal: 0%
	Bad angles:	1 / 732	0.14%	Goal: <0.1%

Рисунок 13. Показатели качества структуры PDB_redo, полученные сервисом MolProbity [4].

В целом видно, что структура PDB_redo часто не лучше, а иногда и хуже PDB структуры. Хотя R-фактор и R-free у redo структуры лучше (рис. 11), однако глядя на количество новых “плохих” остатков, я бы заключил, что структура redo хуже “обычной”.

3) Выводы

Структура 5tyb на мой взгляд очень удачная - она имеет очень мало маргинальных остатков, да и среди тех лишь один можно улучшить. Также я бы сказал что правки в структуре PDB_redo излишни, и хоть и приводят к уменьшению R-фактора и R-free, но по факту вносят больше “спорных” остатков в структуру.

4) Список литературы:

- 1) URL: https://en.wikipedia.org/wiki/DNA_polymerase
- 2) Lucas D1, Escudero B, Ligos JM, Segovia JC, Estrada JC, Terrados G, Blanco L, Samper E, Bernad A. Altered hematopoiesis in mice lacking DNA polymerase mu is due to inefficient double-strand break repair. PLoS Genet, 2009 Feb
- 3) Joonas A. Jamsen, William A. Beard, Lars C. Pedersen, David D. Shock, Andrea F. Moon, Juno M. Krahn, Katarzyna Bebenek, Thomas A. Kunkel & Samuel H. Wilson. Time-lapse crystallography snapshots of a double-strand break repair polymerase in action. Nature Communications, 2017
- 4) URL: <http://molprobit.biochem.duke.edu/>
- 5) URL: <http://eds.bmc.uu.se/eds/>
- 6) URL: http://www.cmbi.ru.nl/pdb_redo/

