

Московский Государственный Университет  
Имени М.В. Ломоносова

Факультет биоинженерии и биоинформатики

Отчет по качеству расшифровки ДНК-лигазы A *Streptococcus Pneumoniae*  
(PDB ID 4GLW)

Выполнила студентка 4 курса  
Горбачева Дарья Викторовна

*Москва 2017*

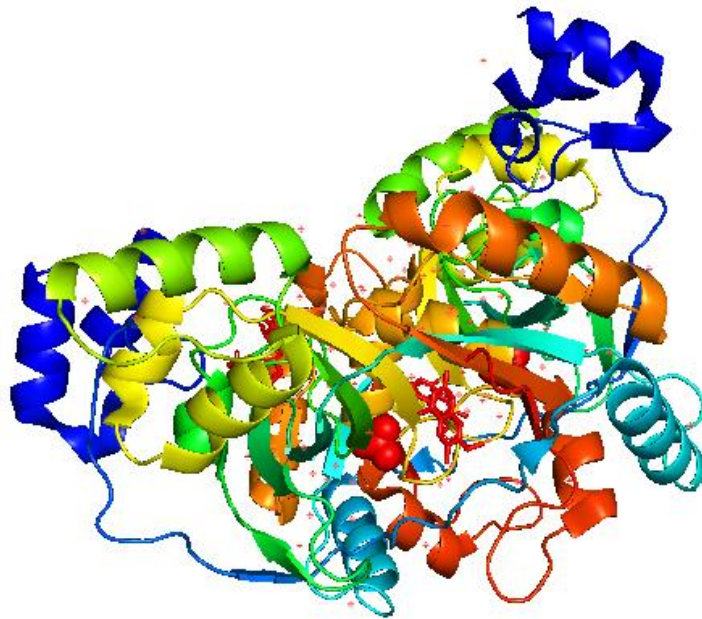
## Аннотация

В данном отчете проведена оценка качества структуры белка с идентификатором PDB 4GLW, расшифрованной методом рентгеноструктурного анализа, проанализированы некоторые отдельные аминокислотные остатки, произведено сравнение моделей из PDB и PDB\_redo.

## Введение

ДНК-лигаза А (или бактериальная НАД-зависимая ДНК-лигаза, LigA, КФ 6.5.1.2) – фермент, катализирующий образование фосфодиэфирной связи между 5'-фосфорильной и 3'-гидроксильной группами соседних дезоксирибонуклеотидов, что очень важно в процессах репарации, репликации, рекомбинации. Механизм этого фермента начинается с аденилирования лизина в активном центре посредством НАД с высвобождением НМН. Затем АМФ переносится с лизина на 5'-фосфорильную группу конца ДНК. Фосфодиэфирная связь образуется из-за нуклеофильной атаки активированного 5'-конца 3'-гидроксильной группой, высвобождается АМФ.

Предполагается, что LigA может служить хорошей мишенью для антибактериальной терапии благодаря своей консервативности в большом спектре как Грам-положительных, так и Грам-отрицательных бактерий и огромной важности для их жизнедеятельности. Кроме того, LigA сильно отличается от эукариотической ДНК-лигазы, что делает ее ингибиторы потенциально безвредными [1].



*Рис. 1. Пространственная структура LigA, окрашены радужно, переход цвета от N к C концу*

## Результаты и обсуждение

### Общая информация о модели

- ❖ Состав комплекса: цепи А и В, НМН, два сульфат-иона, 7-метокси-6-метилптеридин-2,4-диамин (ингибитор)
- ❖ Год: 2012
- ❖ Авторы работы: J.P. Surivet, R. Lange, C. Hubschwerlen, W.Keck, J.L. Specklin, D.Ritz, D. Bur, H. Locher, P.Seiler, D.S. Strasser, L. Prade, C. Kohl, C. Schmitt, G. Chapoux, E. Ilhan, N. Ekambaram, A. Athanasiou, A. Knezevic, D. Sabato, A. Chambovey, M. Gaertner, M. Enderlin, M. Voehme, V. Sippel, P. Wyss
- ❖ Метод решения фазовой проблемы: молекулярное замещение
- ❖ Число измеренных рефлексов: 48105
- ❖ Разрешение: 2.0 Å
- ❖ Полнота набора структурных факторов: 99.9%
- ❖ Диапазон разрешений структурных факторов: 39.41 – 2.0 Å
- ❖ Параметры кристаллографической ячейки:
  - $\alpha=90^\circ$ ,  $\beta=90^\circ$ ,  $\gamma=120^\circ$
  - $a=56.73 \text{ \AA}$ ,  $b=56.73 \text{ \AA}$ ,  $c=197.93 \text{ \AA}$
  - Кристаллографическая группа P31
  - Число молекул в ячейке кристалла 6
- ❖ Некристаллографические симметрии: нет (не упомянуты)

### Значения индикаторов качества модели

- ❖ R-фактор: 0.207
- ❖ R-free: 0.260

Хорошим считается R-factor  $< 0.2$ . В нашем случае значение немного больше, но лишь на сотые доли, так что я все же назову R-фактор хорошим. Для R-free хорошими также считаются значения  $< 0.2$ , плохими  $> 0.4$ , то есть данное значение можно считать удовлетворительным. Разность R-free и R-фактора составляет  $0.053 < 0,1$ , что говорит в пользу отсутствия переоптимизации модели.

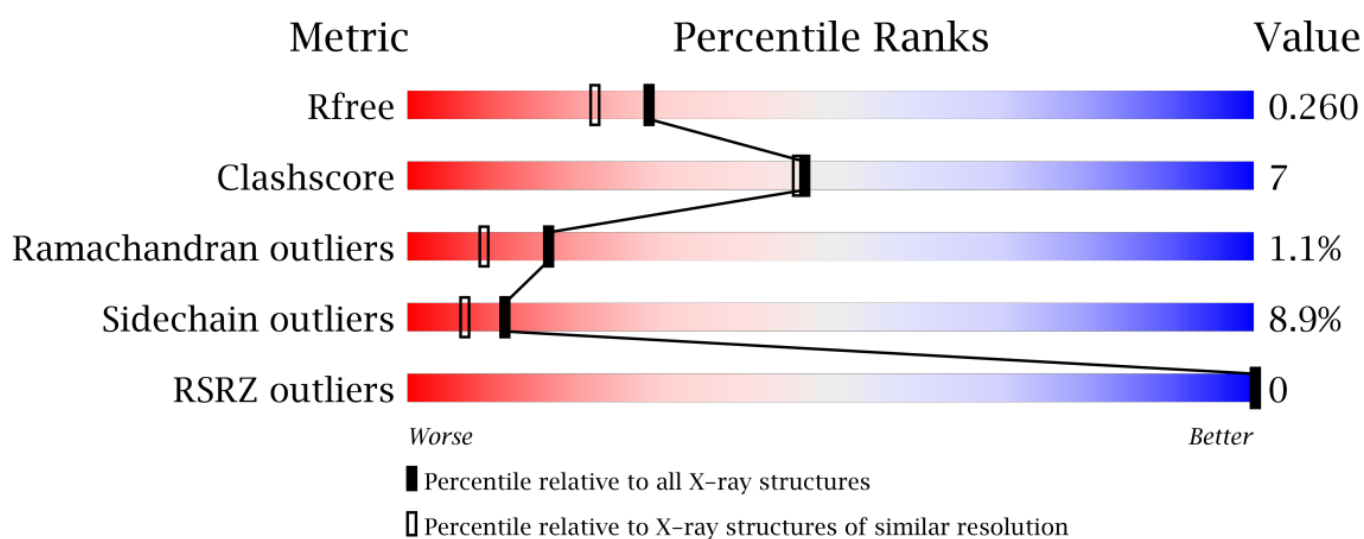


Рис. 2. Граф с основными характеристиками качества модели из Full wwPDB X-ray Structure Validation Report для 4GLW

Можно оценить параметры, приведенные в отчете по валидации со страницы PDB [2], основной граф из него на рисунке 2. Видно, что в среднем по PDB структура нехорошая, но среди структур такого же разрешения ее показатели являются хорошими (выше среднего), показатель слишком близкого расположения невзаимодействующих атомов (clashscore) на среднем уровне, но все равно очень приемлемый (< 20). Проблемы с боковыми цепями будут видны и далее, однако рискну назвать структуру хорошей для своего разрешения (довольно низкого).

### MolProbity [3]

Были построены карты Рамачандрана для модели белка (с предварительным добавлением атомов водорода). При добавлении водорода программа предложила внести изменения в некоторые остатки. Изменения были внесены при указании программой явной необходимости (на рисунке 3).

Flip?	Chain	Res#	Alt	Res ID	Orig	Flip	Flip-Orig	Code	Explanation
<input type="checkbox"/>	A	57		HIS	-1.3	0.26	1.56	FLIP	Some evidence recommending flip.
<input type="checkbox"/>	A	182		ASN	-0.55	-0.025	0.525	FLIP	Some evidence recommending flip.
<input type="checkbox"/>	A	242		ASN	0.62	1.9	1.28	FLIP	Some evidence recommending flip.
<input type="checkbox"/>	A	256		ASN	-1.1	-0.082	1.018	FLIP	Some evidence recommending flip.
<input checked="" type="checkbox"/>	B	6		ASN	-3.8	0.61	4.41	FLIP	Clear evidence for flip.
<input checked="" type="checkbox"/>	B	176		ASN	-6.1	-0.095	6.005	FLIP	Clear evidence for flip.
<input type="checkbox"/>	B	250		ASN	-1.8	0	1.8	FLIP	Some evidence recommending flip.

**Рис. 3. Внесение корректив в модель программой MolProbity**

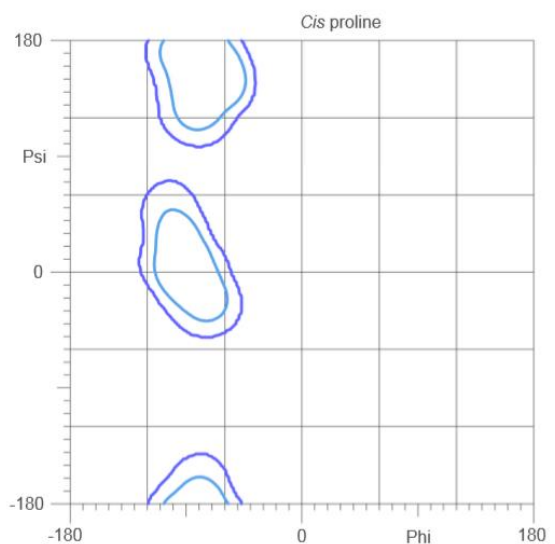
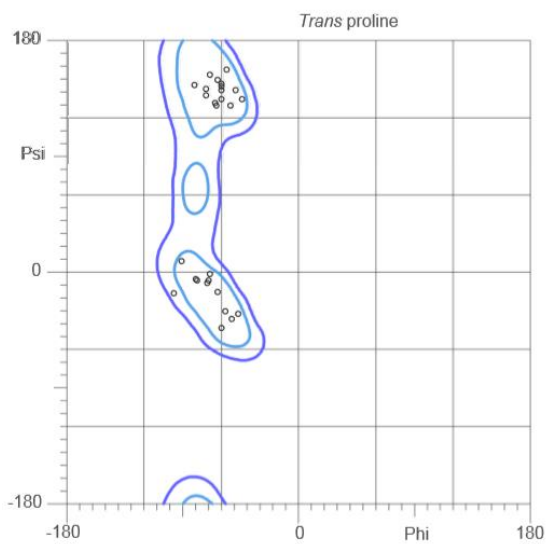
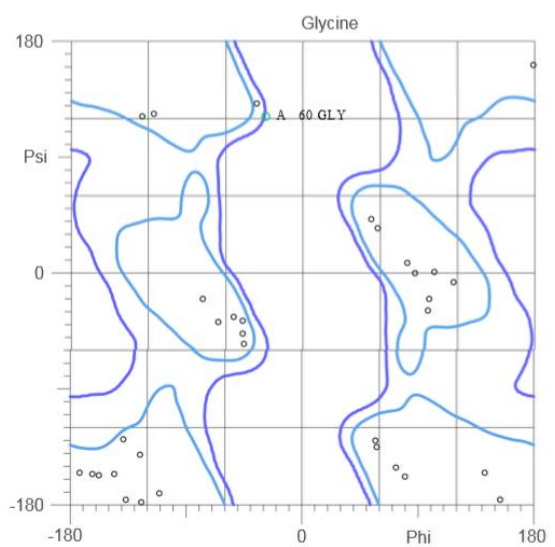
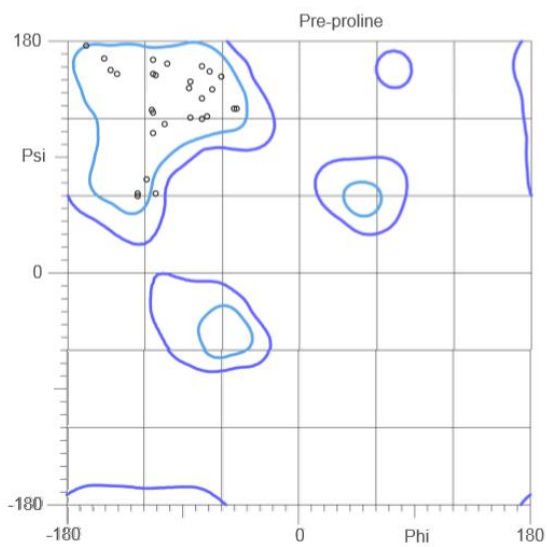
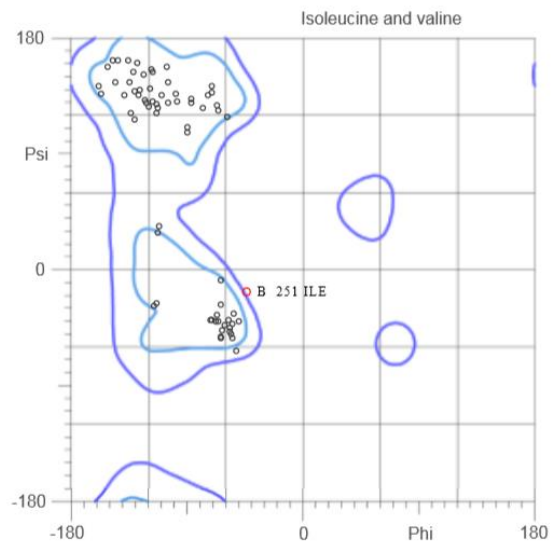
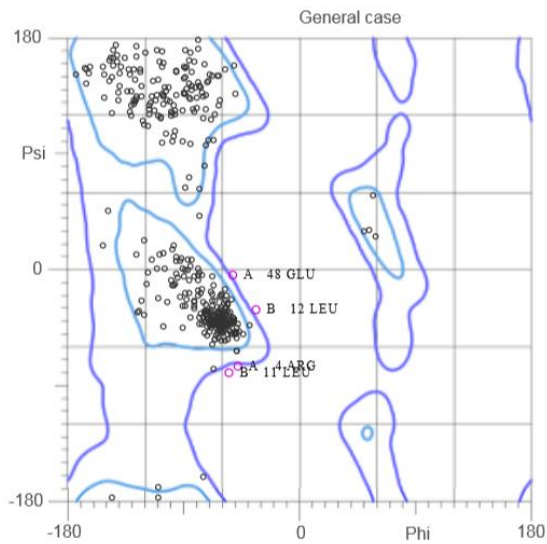
Согласно картам (на рисунке 5), 95.4% остатков находятся в благоприятных областях, 98.9% - в разрешенных. Есть перечисленные на рисунке 5 выбросы в числе шести, для них значения углов  $\phi$  и  $\psi$  лежат за пределами допустимых областей на картах Рамачандрана. Это плохой показатель, говорит не в пользу качества расшифровки.

Также MolProbity представил значения критериев качества модели (на рисунке 4).

All-Atom Contacts	Clashscore, all atoms:	6.51		95 <sup>th</sup> 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	39	9.33%	Goal: <0.3%
	Favored rotamers	339	81.10%	Goal: >98%
	Ramachandran outliers	6	1.09%	Goal: <0.05%
	Ramachandran favored	523	95.44%	Goal: >98%
	MolProbity score <sup>^</sup>	2.41		41 <sup>st</sup> percentile* (N=12522, 2.00Å ± 0.25Å)
	C $\beta$ deviations >0.25Å	6	1.15%	Goal: 0
	Bad bonds:	8 / 4300	0.19%	Goal: 0%
Bad angles:	13 / 5849	0.22%	Goal: <0.1%	
Peptide Omegas	Cis Prolines:	0 / 28	0.00%	Expected: ≤1 per chain, or ≤5%

**Рис.4. Значения критериев качества модели, представленные MolProbity. Красные значения «плохие», желтые удовлетворительные, зеленые хорошие**

Исходя из таблицы на рисунке 4, модель отстает по большому количеству параметров, в том числе выбросы на картах Рамачандрана, «плохие» ротамеры (остатки, маргинальные по углам  $\chi$ , определяющим положения атомов в боковых цепях), связи и углы, значения которых существенно отклоняются от теории. Clashscore, определяющий число Ван-дер-Ваальсовых перекрытий на 1000 атомов, оказался хорошим.



95.4% (524/549) of all residues were in favored (98%) regions.  
 98.9% (543/549) of all residues were in allowed (>99.8%) regions.

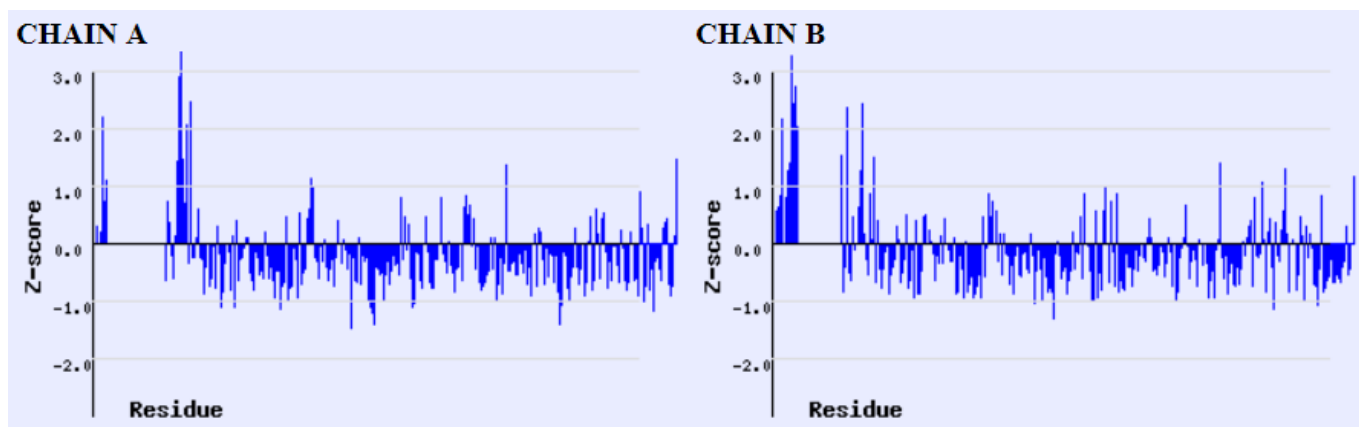
There were 6 outliers (phi, psi):

- A 4 ARG (-48.7, -75.2)
- A 48 GLU (-52.3, -4.7)
- A 60 GLY (-29.1, 122.6)
- B 11 LEU (-55.4, -80.5)
- B 12 LEU (-34.6, -31.4)
- B 251 ILE (-44.2, -17.9)

**Рис. 5. Карты Рамачандрана для всех и отдельных аминокислот**

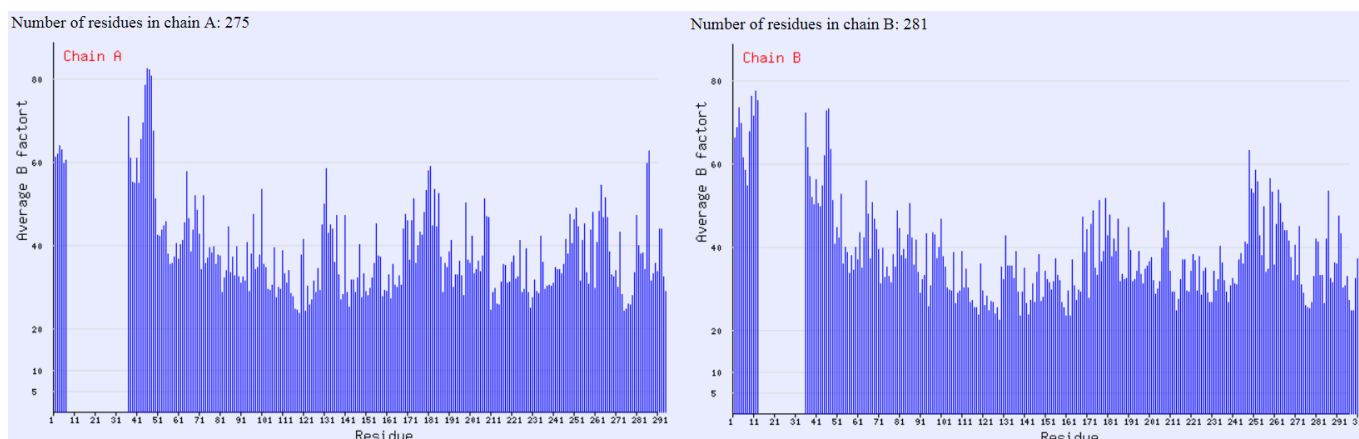
## EDS [4]

Z-score – параметр, позволяющий сравнить пространственный R-фактор (RSR) отдельного остатка модели со средним RSR остатков в моделях такого же разрешения. Для маргинального остатка Z-score > 2. Значения Z-score'ов для остатков 4GLW приведены на рисунке 6. Видно, что выбросы есть.



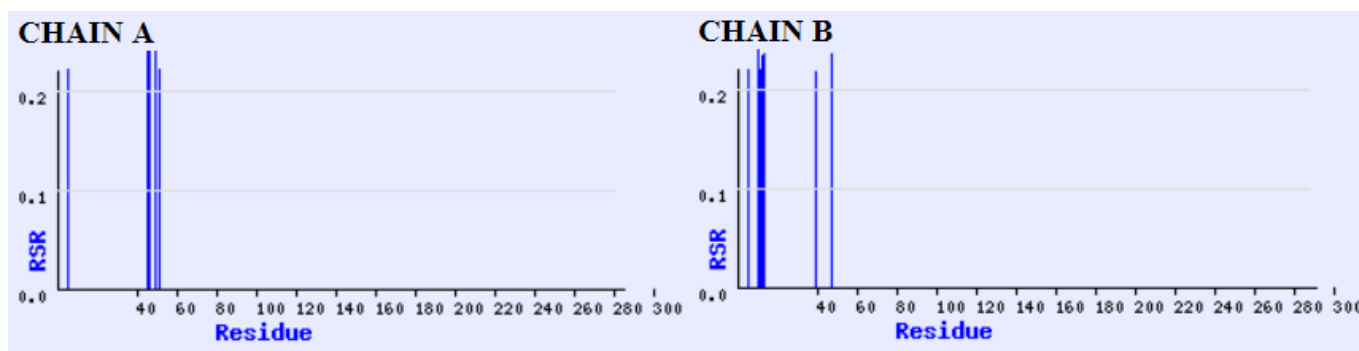
*Рис. 6. Значения Z-score для остатков двух цепей структуры 4GLW*

Температурный фактор (B-фактор) отражает подвижность атомов. B-фактор < 30 означает фиксированность позиции атома, B-фактор > 60 – высокую подвижность. На рисунке 7 представлены B-факторы для остатков структуры. Видно, что в среднем B-фактор ~35, что удовлетворительно, но не хорошо.



*Рис. 7. Значения B-фактора для остатков двух цепей структуры 4GLW*

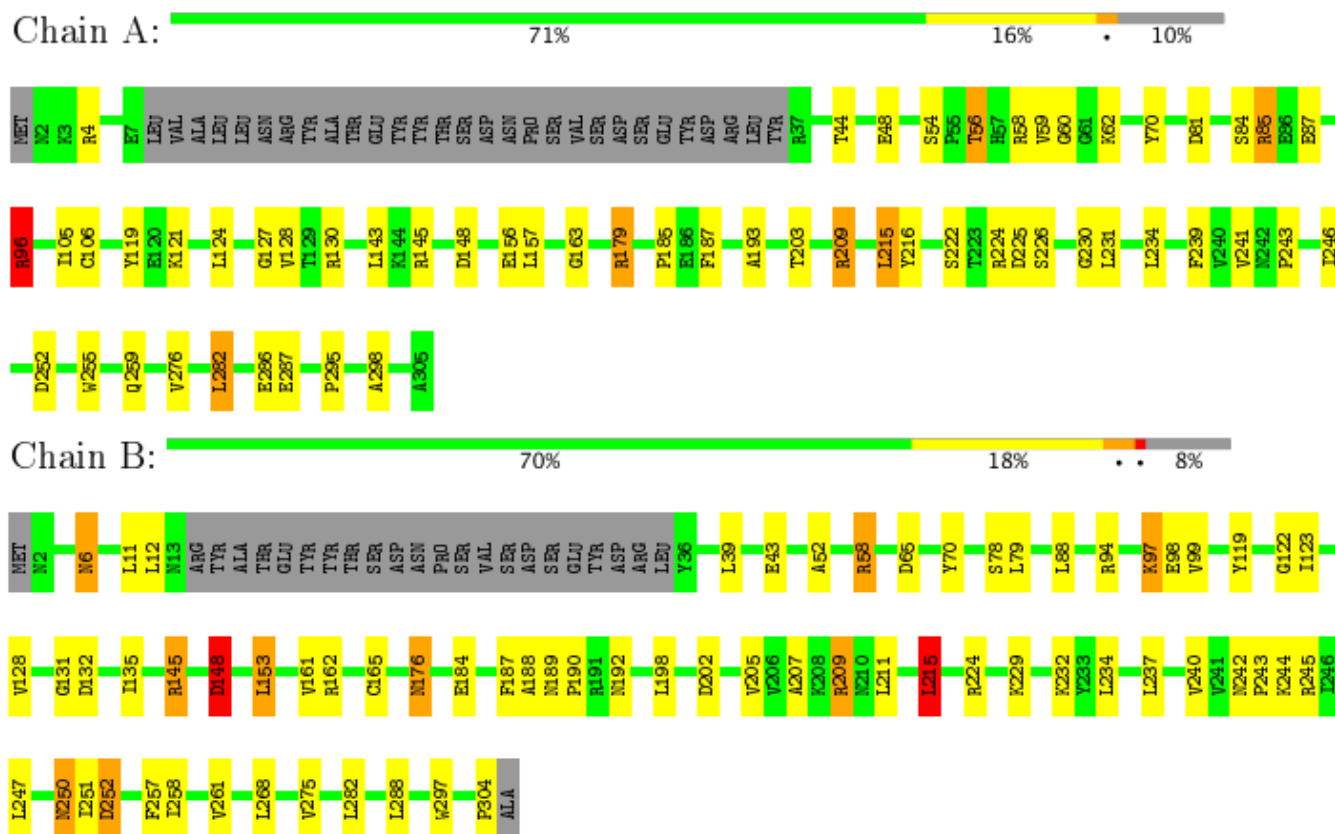
По рисункам 6 и 7 видно, что для обеих цепей белка N-конец оказался проблемным местом для определения структуры. Это подтверждается и графиком с сомнительными остатками с повышенным RSR (на рисунке 8).



*Рис. 8. Участки с повышенным RSR (significant regions) для двух цепей 4GLW*

## Примеры маргинальных остатков

Полная информация о маргинальных остатках содержится в полном отчете о валидации на странице структуры в PDB, в выдаче MolProbity (файл Multi-criterion chart). Данные из этих источников и из приведенных выше согласуются. Примеры маргинальных по разным причинам остатков в таблице 1. Удобное отображение маргинальности остатков представлено в отчете с PDB [2], на рисунке 9.



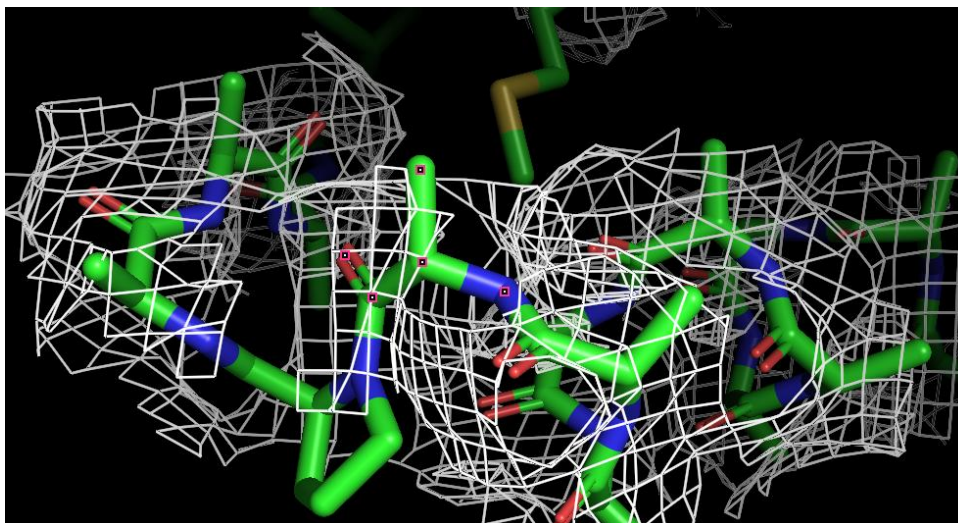
**Рис. 9. Качество остатков в белке 4GLW.** Зеленые – хорошие остатки, желтые – маргинальны по одному, оранжевые – по двум, красные – по трем и более признакам. Красная точка – плохая вписанность в электронную плотность (ЭП)

**Таблица 1. Примеры маргинальных остатков.** Выделены выбранные для следующего шага

Цепь	Остаток	Причина маргинальности
A	Glu 48	Выброс на карте Рамачандрана ( $\phi = -52.3$ , $\psi = -4.7$ )
B	Leu 11	Выброс на карте Рамачандрана ( $\phi = -55.4$ , $\psi = -80.5$ ), В-фактор 71.53
A	Trp 44	«Плохой» ротамер ( $\chi = 348.7$ )
A	Arg 209	«Плохой» ротамер ( $\chi = 335.6$ ; 253.9; 235; 213.9), аномальный угол CG-CD-NE: 4.5 $\sigma$
B	Asp 148	«Плохой» ротамер ( $\chi = 330$ ; 2.5), аномальный угол C-CA-CB: 6.8 $\sigma$
A	Tyr 46	Высокий Z-score = 3.37, высокий B-score = 82.56, высокий RSR = 0.252
B	Leu 215	«Плохой» ротамер ( $\chi = 284.2$ ; 110.9), аномальный угол CA-CB-CG: 6.1 $\sigma$ , высокий clash HD22 с B Val 161 HG12
A	Arg 96	«Плохой» ротамер ( $\chi = 275.4$ ; 134.8; 147.4; 244.4), аномальный угол NE-CZ-NH1 (но только по отчету с PDB, но не MolProbity), большой clash HD3 с A Arg 96 NH11
A	His 72	Аномальная длина связи CG-ND1: 4.1 $\sigma$
B	Ile 110	Аномальный угол C-CA-CB: 4.2 $\sigma$ , отклонение C $\beta$ (0.26 $\text{\AA}$ )

## Анализ некоторых маргинальных остатков

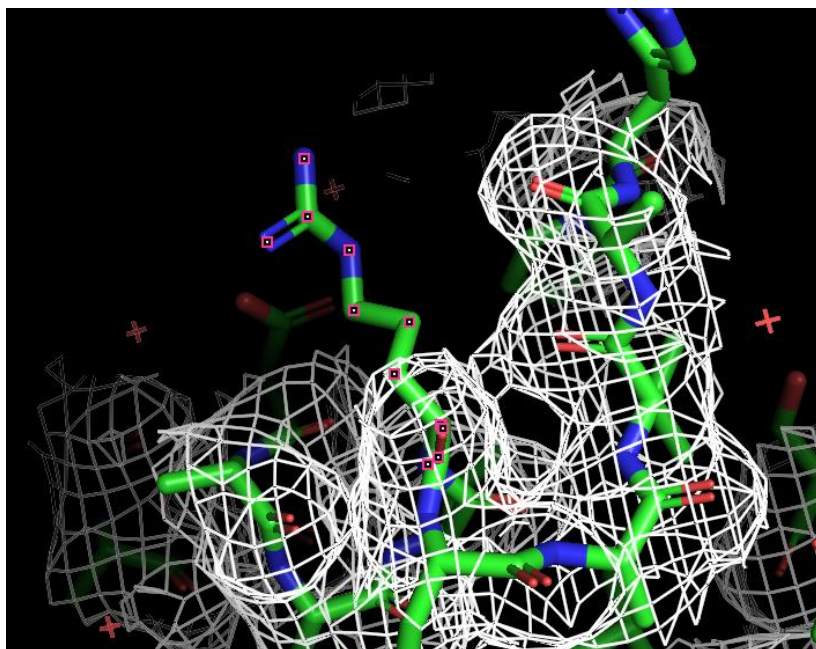
### ❖ A Tyr 46 (на рисунке 10)



*Рис. 10. A Tyr 46. Уровень подрезки ЭП 0.1  $\sigma$ . Остаток выделен красными точками*

Удивительным образом на месте тирозина в структуре оказался аланин! В соответствующей белку LigA *S. pneumoniae* записи в БД NCBI (NP\_358618.1 ссылка [5]) на 46 позиции стоит тирозин, значит, неточностей в последовательности нет. Видно, что остаток 46 (каким бы он ни был) плохо вписывается в ЭП даже при очень низком уровне подрезки, то есть определение его пространственного положения было сложной задачей для исследователей. Возможно, этот остаток не стали отображать, поскольку данных о его положении было слишком мало.

### ❖ A Arg 96 (на рисунке 11)

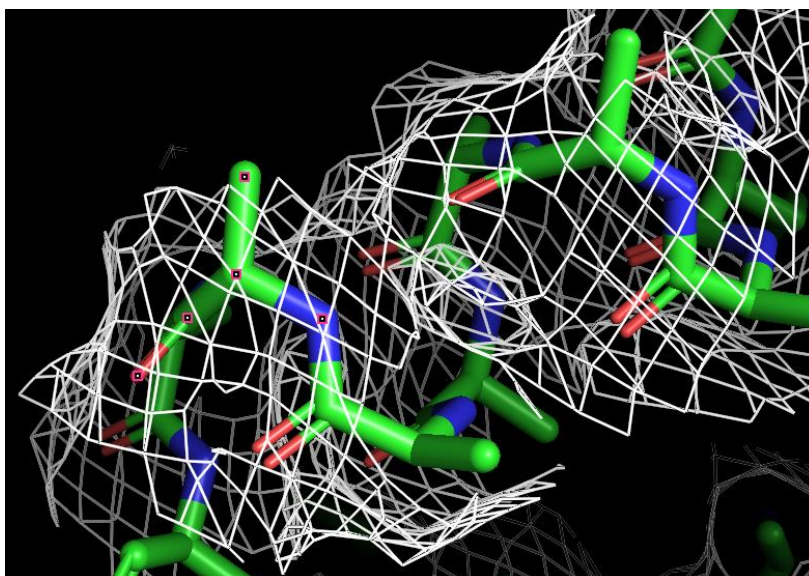


*Рис. 11. A Arg 96. Уровень подрезки ЭП 0.1  $\sigma$ . Остаток выделен красными точками*

Я не смогла добиться того, чтобы этот остаток был вписан в ЭП, даже уменьшив уровень подрезки до 0.1  $\sigma$ . Ясно, что положение остатка базируется только на представлениях авторов модели о его положении, но не на экспериментальных данных. Возможно, авторам модели следовало расположить этот аргинин в пространстве иначе, чтобы его конформация была менее напряженной.



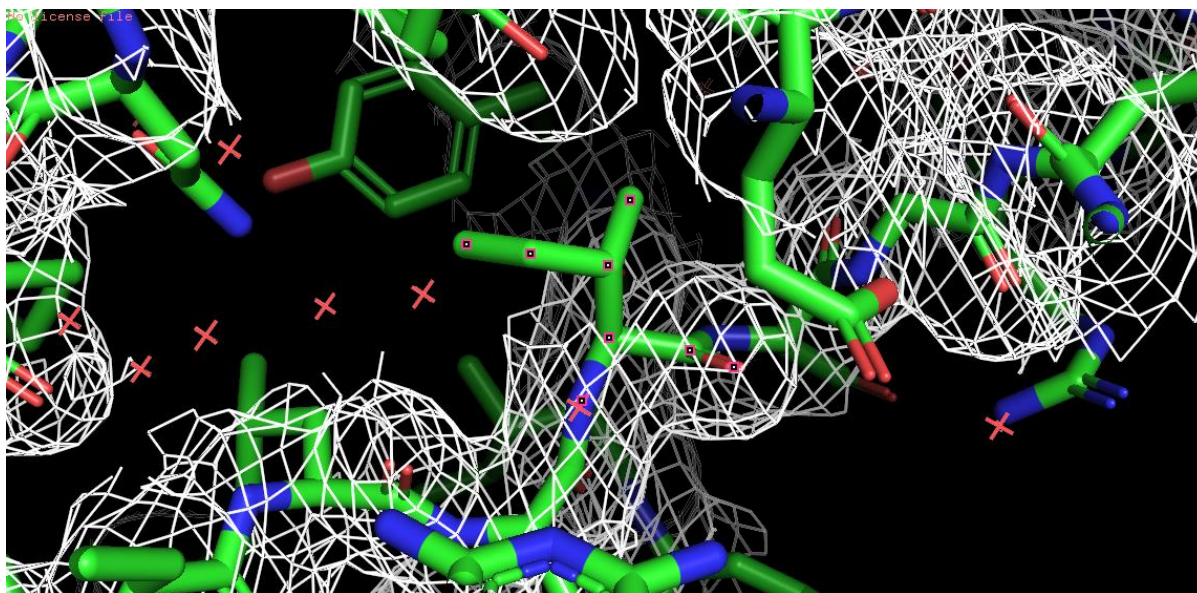
❖ В Leu 11 (на рисунке 12)



*Рис. 12. В Leu 11. Уровень подрезки ЭП 0.1  $\sigma$ . Остаток выделен красными точками*

Так же, как и в случае с А Туг 46, боковая цепь недоизобразена, остаток не удастся вписать в ЭП. Наверное, В Leu 11 не изображен по той же причине. Стоит отметить, что это остаток с начала полипептидной цепи, а они обычно сложно определяются. Еще можно предположить, что положения этого лейцина и тирозина неопределяемые теоретически из-за того, что они располагаются прямо на поверхности и, соответственно, никакие другие остатки не мешают им свободно колебаться в пространстве так, что невозможно зафиксировать одну конкретную позицию (в пользу этого говорят аномально высокие значения В-факторов обоих остатков).

❖ В Ile 110 (на рисунке 13)



*Рис. 13. В Ile 110. Уровень подрезки ЭП 0.1  $\sigma$ . Остаток выделен красными точками*

Если предыдущие остатки располагались в свободном положении на поверхности, то этот достаточно плотно окружен, то есть его конформация должна быть зафиксирована. Я думаю, что даже несмотря на очень плохую вписанность в ЭП необычные характеристики этого изолейцина могут быть особенностью структуры, а не артефактом, поскольку для остатка может быть непросто сохранить все наиболее выгодные характеристики углов и связей в столь плотном окружении.

❖ В Leu 215 (на рисунке 14)

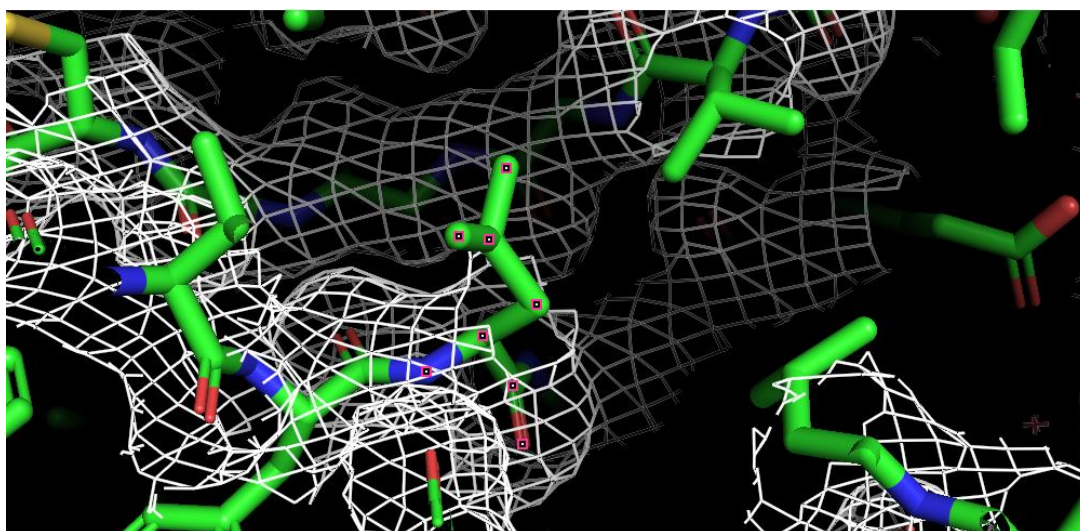


Рис. 14. В Leu 215. Уровень подрезки ЭП 0.1  $\sigma$ . Остаток выделен красными точками

В данном случае маргинальность остатка выглядит как недостаток построения модели, хотя ЭП дает слабое представление о положении остатка. Видно, что лейцин неоправданно сильно наклонен своим C $\beta$  в сторону Val 161, с которым находится в положении clash, а затем отклоняется от него, формируя неестественный аномальный угол. Этого вполне можно было избежать без особых потерь для вписанности структуры в ЭП, если бы положение C $\beta$  задали другим и, возможно, расположили бы весь остаток чуть дальше отвалина. И угол стал бы тупее, и не было бы clash.

### Сравнение модели из PDB с моделью из PDB\_redo [6]

PDB\_redo – база данных оптимизированных моделей из PDB. Для оптимизации перестраиваются конформации роротамеров, положения молекул воды, оптимизируются водородные связи и т.д. Из рисунка 15 видно, что показатели качества модели после оптимизации улучшились, ни один показатель не ухудшился.

Validation metrics from PDB-REDO		
	PDB	PDB-REDO
<i>Crystallographic refinement</i>		
<i>R</i>	0.2097	0.1775
<i>R-free</i>	0.2603	0.2085
<i>Bond length RMS Z-score</i>	0.877	0.719
<i>Bond angle RMS Z-score</i>	1.012	0.876
<i>Model quality (raw scores   percentiles)</i>		
<i>Ramachandran plot appearance</i>	17	65
<i>Rotamer normality</i>	2	15
<i>Coarse packing</i>	N/A	N/A
<i>Fine packing</i>	N/A	61
<i>Bump severity</i>	72	84
<i>Hydrogen bond satisfaction</i>	54	49

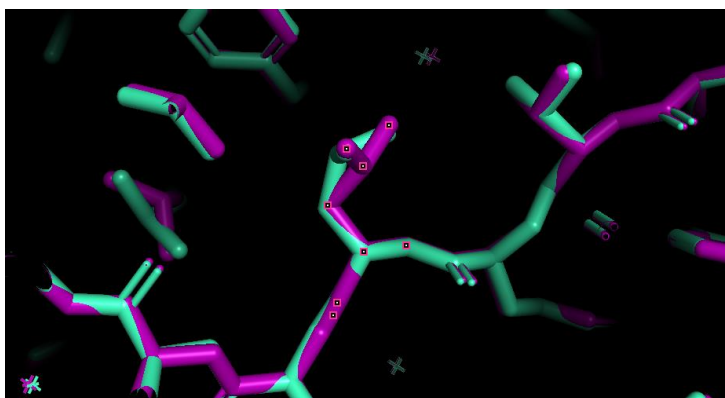
Рис. 15. Сравнение характеристик исходной модели 4GLW и оптимизированной в PDB\_redo

Для сравнения двух структур я начала с их простого наложения друг на друга, которое показало, что глобально оптимизация с помощью PDB\_redo ничего не изменила (см. рисунок 16).



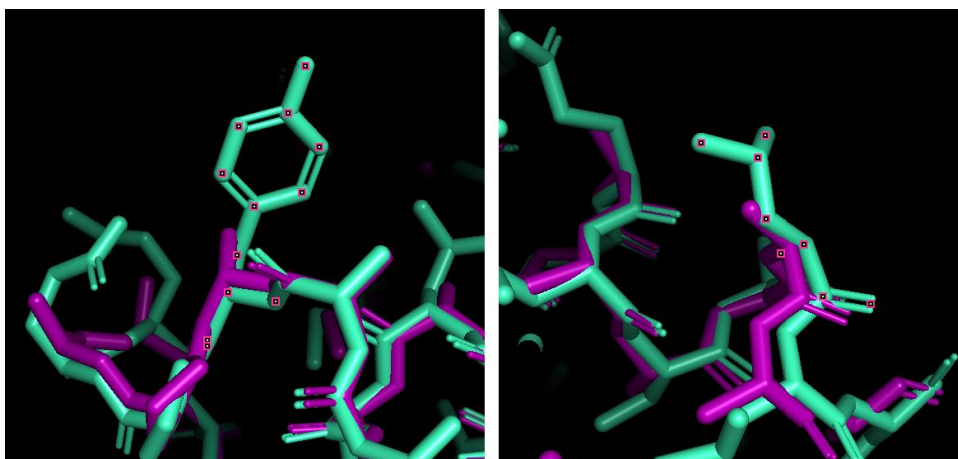
*Рис. 16. Пространственное выравнивание исходной (фиолетовый) и оптимизированной (голубой) структур*

Я решила посмотреть на ход оптимизации на примере рассмотренных ранее маргинальных остатков. Во-первых, для В Leu 215 сильных изменений сделано не было (изображено на рисунке 17). Спорный острый угол был немного сглажен, но не сильно, что свидетельствует в пользу «искривленности» этого остатка из-за особенностей белка, а не расшифровки структуры.



*Рис. 17. То же пространственное выравнивание в области В Leu 215*

А вот недоизображенные остатки А Tyr 46 и В Leu 11 были исправлены (рисунок 18). Это говорит о том, что отсутствие этих боковых цепей могло быть ошибкой модели.



*Рис. 18. То же пространственное выравнивание в области А Tyr 46 (слева) и В Leu 11 (справа)*

## Заключение

Нельзя сказать, что структура 4GLW расшифрована хорошо. Показатели ее качества значительно хуже среднего по PDB, хоть и хороши в классе структур с низким разрешением. В структуре много маргинальных остатков, особенно на поверхности. Однако использование PDB\_redo позволяет улучшить как характеристики модели в целом, так и определение отдельных остатков.

## Источники

- [1] Surivet JP, Lange R, Hubschwerlen C. et al. Structure-guided design, synthesis and biological evaluation of novel DNA ligase inhibitors with in vitro and in vivo anti-staphylococcal activity. *Bioorg Med Chem Lett* 2012; 22: 6705–11
- [2] [http://files.rcsb.org/pub/pdb/validation\\_reports/gl/4glw/4glw\\_full\\_validation.pdf](http://files.rcsb.org/pub/pdb/validation_reports/gl/4glw/4glw_full_validation.pdf)
- [3] <http://molprobity.biochem.duke.edu/>
- [4] <http://eds.bmc.uu.se/eds/>
- [5] [https://www.ncbi.nlm.nih.gov/protein/NP\\_358618.1](https://www.ncbi.nlm.nih.gov/protein/NP_358618.1)
- [6] [http://www.cmbi.ru.nl/pdb\\_redo/](http://www.cmbi.ru.nl/pdb_redo/)