

МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
имени М.В. ЛОМОНОСОВА

ФАКУЛЬТЕТ БИОИНЖЕНЕРИИ И БИОИНФОРМАТИКИ

**Отчёт по качеству РСА расшифровки структуры
комплекса с PDB-ID: 4kr2**

Отчетная работа студентки 4-го курса
Калашниковой Анастасии Ильиничны

Москва
2019 г.

Аннотация

В данной работе представлен отчет по качеству расшифровки структуры глицил-тРНК-синтетазы с PDB-идентификатором: 4kr2, полученной методом рентгеноструктурного анализа.

Введение

Глицил-тРНК-синтетаза человека (GARS) – фермент, который принадлежит ко второму классу аминоксил-тРНК-синтетаз. К каноническим свойствам относят способность фермента катализировать образование глицил-тРНК в реакции этерификации глицина и непосредственно тРНК [1].

Данный фермент имеет целый ряд интересных свойств, включая неканонические. Одним из ключевых неканонических свойств GARS является связывание с IRES элементом полиовируса. Было показано, что GARS довольно прочно связывает пятый домен IRES элемента данного вируса, тем самым стимулируя его трансляцию. Механизм данного явления на данный момент не известен [2].

Глицил-тРНК-синтетаза является одной из немногих синтетаз, которые обладают различными олигомерными структурами в разных организмах (типы « $\alpha_2\beta_2$ » и « α_2 »). К организмам, обладающим глицил-тРНК-синтетазой типа « $\alpha_2\beta_2$ », относят, например: *Escherichia coli*, *Bacillus subtilis* и *Lactobacillus brevis*; типа « α_2 »: *Saccharomyces cerevisiae*, *Drosophila melanogaster* и *Homo sapiens*. *Arabidopsis thaliana* является, в этом плане, уникальным организмом: у данной водоросли присутствуют оба типа фермента. Также можно отметить, что ферменты типа « $\alpha_2\beta_2$ » имеют сходство с фенил-тРНК-синтетазой (также фермент типа « $\alpha_2\beta_2$ »).

Структура глицил-тРНК-синтетазы человека представлена на рисунке 1 – данный фермент является гомодимером типа « α_2 ». Мономер образуют 685 остатков, он состоит из 3 доменов: присоединенного к N-концу домена WHEP-TRS, каталитического домена и C-концевого антикодон-связывающего домена. Каталитический домен содержит три характерных консервативных мотива (1, 2 и 3) тРНК-синтетаз класса II и, кроме того, три вставки (I, II и III) между мотивами [3].

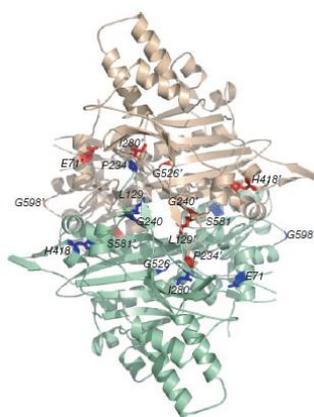


Рис. 1 – Структура GARS с обозначенными мономерами (зеленым и бежевым цветом соответственно) и некоторыми функционально важными остатками [3]

Результаты и обсуждение

Общая информация о модели

В состав комплекса с PDB-идентификатором 4kr2 входят глицил-тРНК-синтетаза и глицил-тРНК (рис. 2) [4]. С белком связан ампициллин и молекулы воды. Всего файл содержит 5223 атома, атомы водорода не представлены.

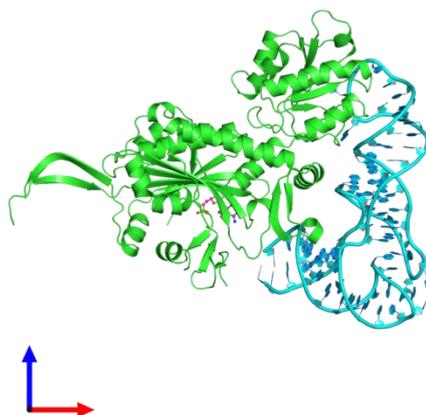


Рис. 2 - Структура комплекса с PDB-идентификатором 4kr2: глицил-тРНК-синтетаза обозначена зеленым; глицил-тРНК обозначена голубым [4]

Структура комплекса была расшифрована в 2013 году рядом авторов: Xiangjing Qin, Zhitai Hao, Qingnan Tian, Zhemin Zhang, Chun Zhou, Wei Xie [5]. Для решения фазовой проблемы использовали метод молекулярного замещения с помощью программы Phenix.

Число измеренных рефлексов, указанное в файле структурных факторов: 29706.
Разрешение структуры: 3.292 Å; полнота набора структурных факторов: 88.9 %; диапазон разрешений структурных факторов, указанный на странице EDS: 32.25 - 3.29 Å.

Параметры кристаллографической ячейки: a: 135.964 Å; b: 88.499 Å; c: 80.497 Å; α : 90.00°; β : 90.00°; γ : 90.00°; кристаллографическая группа: P 21 21 2.

Значения индикаторов качества модели

Значения индикаторов качества модели были получены с помощью нескольких сайтов/сервисов: PDB, MolProbity, EDS. Я использовала для оценки такие параметры, как: R-фактор, R_{free} , карта Рамачандрана, перекрытие Ван-дер-Ваальсовых радиусов атомов (ClashScore), RSR.

Сравнение моей модели с другими моделями PDB представлено на рисунке 3. Как можно заметить, моя модель, в целом, лучше моделей со схожим разрешением, но хуже большинства моделей PCA. Поэтому, можно предположить, что качество модели – среднее в общем, однако, для данного разрешения оно довольно неплохое на фоне остальных.

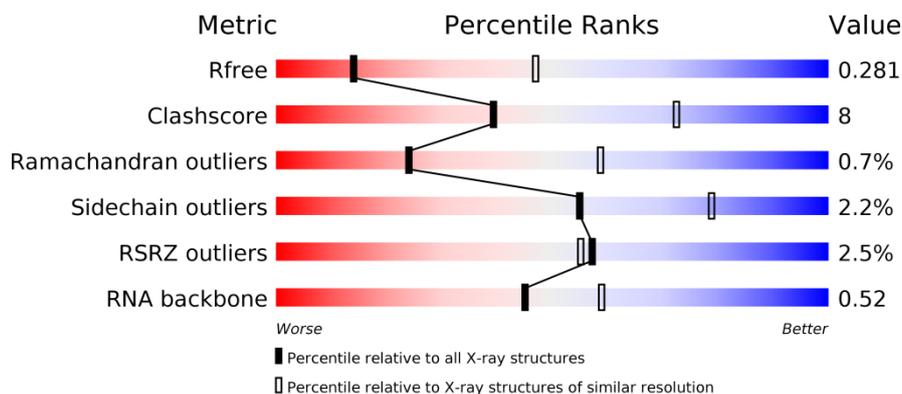


Рис. 3 – Показатели качества модели 4kr2 относительно всех моделей PCA (темные прямоугольники), относительно моделей PCA со схожим разрешением (светлые прямоугольники)

***R*-фактор и R_{free}**

R-фактор отражает соответствие модели экспериментальным данным. В данном случае R-фактор = 0.236 (данные PDB). Данное значение меньше 0.25, поэтому R-фактор моей модели можно считать «хорошим».

R_{free} – величина, в целом, аналогичная R-фактору, но рассчитанная по меньшему количеству рефлексов. R_{free} моей модели равен 0.277 (данные PDB), поэтому значение можно назвать «хорошим» (меньше 0.31).

Также можно посмотреть разность этих 2 параметров на предмет переоптимизации модели. Значение R_{free} -R равно 0.041 (<0.1), поэтому можно сделать вывод, что переоптимизации модели отсутствует.

Карта Рамачандрана

Для построения карты Рамачандрана, выявления маргинальных остатков и недопустимых наложений атомов, а также оценки качества структуры был использован сервис MolProbity [6]. Во время работы сервис добавил водороды и предложил инверсию боковых цепей His378 и Gln640, которая была одобрена, так как при данных инверсиях достигаются наиболее оптимальные значения параметров качества модели. Один из результатов работы сервиса – сводная таблица с суммарными характеристиками представлена на рисунке 4.

All-Atom Contacts	Clashscore, all atoms:	7.85	97 th percentile* (N=37, 3Å - 9999Å)
	Clashscore is the number of serious steric overlaps (> 0.4 Å) per 1000 atoms.		
Protein Geometry	Poor rotamers	12	3.28% Goal: <0.3%
	Favored rotamers	329	89.89% Goal: >98%
	Ramachandran outliers	3	0.67% Goal: <0.05%
	Ramachandran favored	427	94.89% Goal: >98%
	MolProbity score*	2.17	100 th percentile* (N=895, 3.29Å ± 0.25Å)
	Cβ deviations >0.25Å	1	0.23% Goal: 0
Peptide Omegas	Bad bonds:	0 / 3647	0.00% Goal: 0%
	Bad angles:	2 / 4953	0.04% Goal: <0.1%
Nucleic Acid Geometry	Cis Prolines:	1 / 23	4.35% Expected: ≤1 per chain, or ≤5%
	Twisted Peptides:	1 / 454	0.22% Goal: 0
Low-resolution Criteria	Bad bonds:	5 / 1645	0.30% Goal: 0%
	Bad angles:	0 / 2561	0.00% Goal: <0.1%
Low-resolution Criteria	CaBLAM outliers	4	0.90% Goal: <1.0%
	CA Geometry outliers	3	0.68% Goal: <0.5%

Рис. 4 – Таблица с суммарными характеристиками, построенная сервисом MolProbity

Данные таблицы говорят о том, что у меня много остатков с маргинальными по отклонению от ротамеров боковыми цепями (12), полных маргиналов по карте Рамачандрана (3) и недостаточно остатков в предпочитаемой области (у меня 94.89%, что, в целом, не такой плохой результат, но до «приемлемого» с точки зрения сервиса (> 98%) не дотягивает). ClashScore (число недопустимых наложений атомов на 1000) для моей структуры равен 7.85; перцентиль – 97 – стремится к идеальному (такой перцентиль обозначает, что 97% структур похожего разрешения имеют ClashScore хуже, чем моя модель).

Карта Рамачандрана - карта запрещенных и разрешенных конформаций остатка, изображенная в координатах (ϕ , ψ). Карта Рамачандрана по всей моей модели представлена на рисунке 5, по отдельным остаткам – на рисунке 6. 427 остатков из 450 (94.9%) находятся в предпочтительной области, 447 из 450 (99.3%) – в разрешенной. В моей модели также есть 3 полных маргинала по карте Рамачандрана: Pro234, Ser411, Pro575. Также был найден пролин в цис-конформации (Pro266).

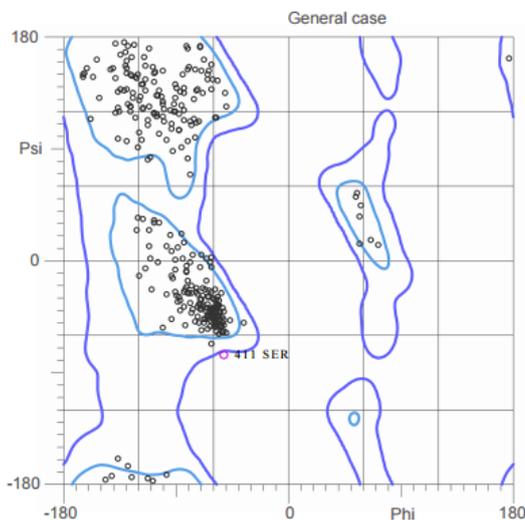
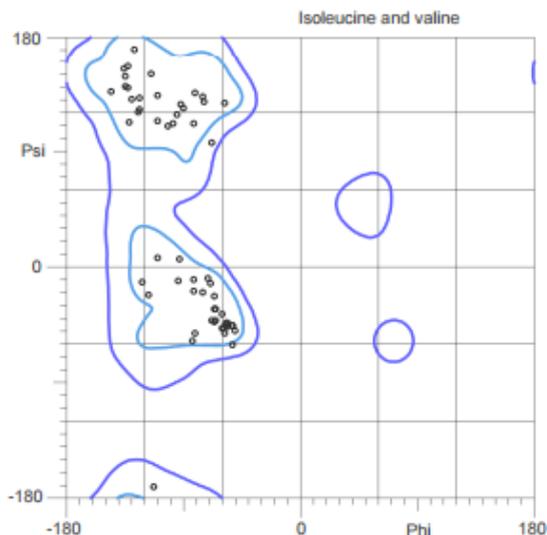


Рис. 5 – Карта Рамачандрана для всей модели. Синим контуром обозначены разрешенные области, голубым – предпочтительные области. Сами остатки изображены в виде точек. Фиолетовая точка (Ser411) – маргинальный остаток по карте Рамачандрана, лежит вне допустимой области.



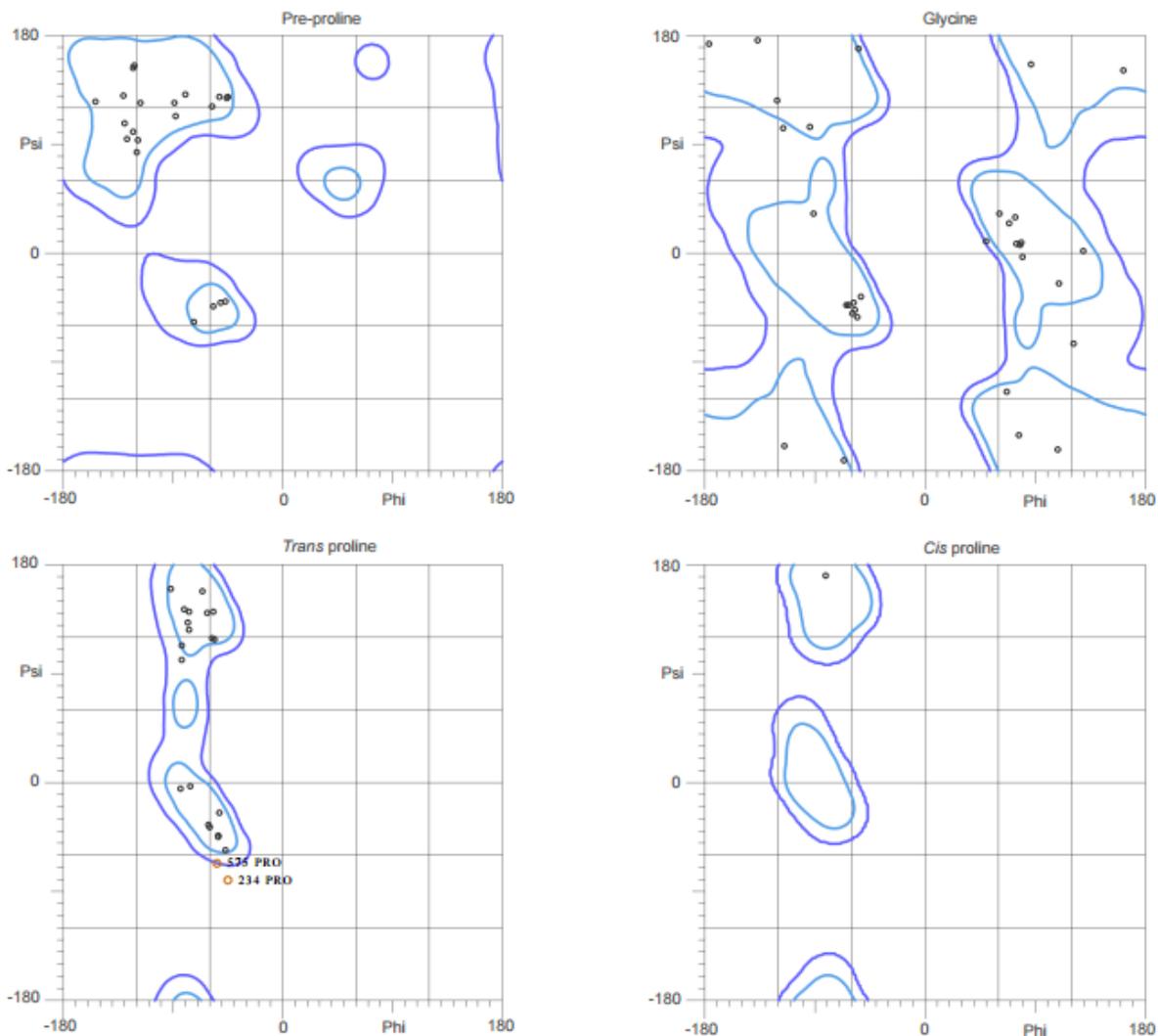


Рис. 6 – Карта Рамачандрана для отдельных остатков. Синим контуром обозначены разрешенные области, голубым – предпочтительные области. Сами остатки изображены в виде точек. Оранжевые точки (Pro234, Pro575) – маргиналы, лежат вне допустимой области.

RSR

RSR — пространственный R-фактор для всех остатков и молекул растворителя. Если для остатка RSRZ (Z-score) больше или равен 2.0, то остаток является маргиналом. Таким образом, моя модель содержит 10 маргиналов по фактору RSR (цепь A).

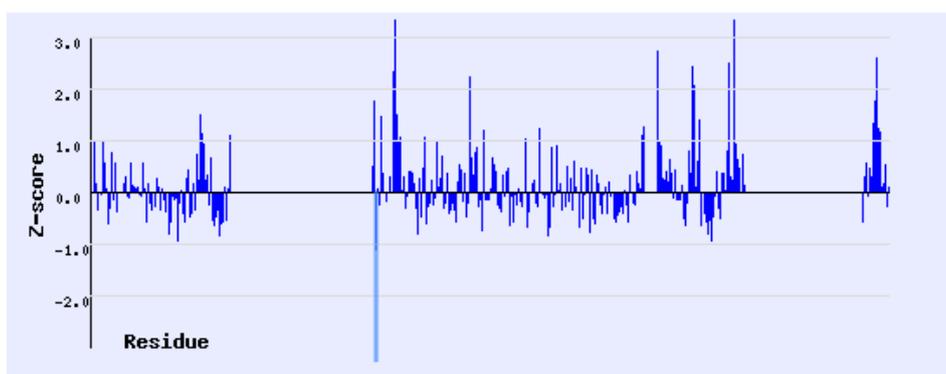


Рис. 7 – RSRZ для полипептидной цепи A (GARS) [7]

Маргинальные остатки

Общая таблица по 7 маргинальным остаткам с указанием показателя, по которому они отобраны. Для анализа были выбраны 5 маргинальных остатков: Pro234, Asn237, Pro266, Cys388, Ser411.

Таблица 1. Маргинальные остатки

<i>Маргинальный остаток</i>	<i>Параметр отбора</i>
Pro234	Запрещенная область карты Рамачандрана
Asn237	RSRZ = 4.2 (> 2)
Pro266	Остаток в cis конформации
Leu280	RSRZ = 2.2 (> 2)
Cys388	RSRZ = 2.4 (> 2)
Ser411	Запрещенная область карты Рамачандрана; перекрытие Ван-дер-Ваальсовых радиусов атомов на 0.42 Å
Pro575	Запрещенная область карты Рамачандрана

Pro234

Данный остаток является маргиналом по карте Рамачандрана. На рисунке 8 представлено изображение электронной плотности вокруг остатка при $\sigma = 1.0$. Электронная плотность, даже при уровне подрезки равном 1, почти не покрывает остаток, значит, можно сделать вывод, что в данном случае имеет место быть ошибка расшифровки, а не особенность структуры.

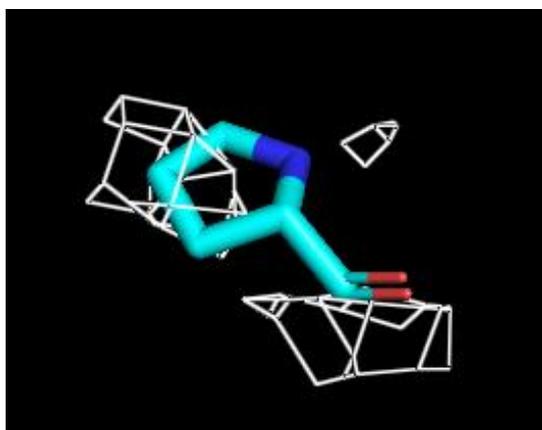


Рис. 8 – Изображение электронной плотности вокруг Pro234 при $\sigma = 1.0$

Asn237

Данный остаток является маргиналом по RSRZ ($4.2 > 2$). На рисунке 9 представлено изображение электронной плотности вокруг остатка при $\sigma = 1.0$. Как можно заметить, электронная плотность покрывает лишь «остовную», но не «радикальную» часть аминокислоты (при уровнях подрезки ЭП больше или равных 0.5), и не отражает структуру аспарагина, поэтому, скорее всего, маргинальность этого остатка – это ошибка расшифровки.

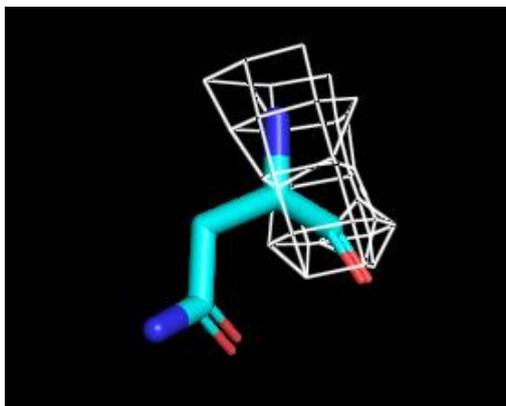


Рис. 9 – Изображение электронной плотности вокруг Asn237 при $\sigma = 1.0$

Pro266

Данный остаток является маргиналом по неблагоприятному окружению (остаток в цис-конформации). На рисунке 10 представлено изображение электронной плотности вокруг остатка при $\sigma = 1.5$. Как можно заметить, электронная плотность покрывает остаток довольно хорошо, поэтому, думаю, данный остаток является скорее особенностью, нежели ошибкой расшифровки.

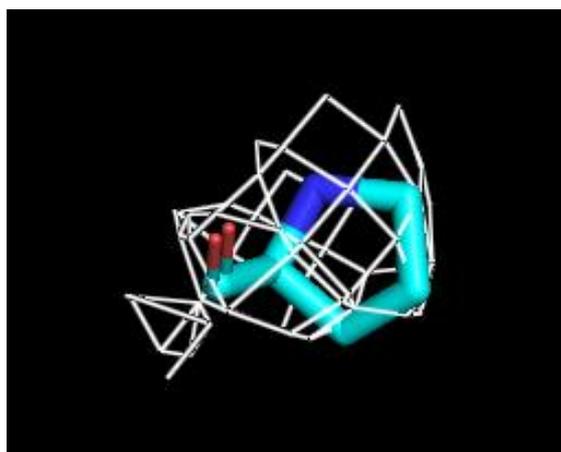


Рис. 10 – Изображение электронной плотности вокруг Pro266 при $\sigma = 1.5$

Cys388

Данный остаток является маргиналом по RSRZ ($2.4 > 2$). На рисунке 11 представлено изображение электронной плотности вокруг остатка при $\sigma = 1.0$. Как можно заметить, визуализированная электронная плотность не соответствует заявленному остатку – радикал аминокислоты не покрыт ЭП даже при $\sigma = 1.0$. При $\sigma = 1.5$ ЭП покрывает амидную группу и немного соприкасается с кислородом. Поэтому я склоняюсь к выводу, что в данном случае мы также имеем дело с ошибкой расшифровки.



Рис. 11 – Изображение электронной плотности вокруг Cys388 при $\sigma = 1.0$

Ser411

Данный остаток является маргиналом по нескольким критериям: запрещенная область карты Рамачандрана; перекрытие Ван-дер-Ваальсовых радиусов атомов на 0.42 \AA . На рисунке 12 представлено изображение электронной плотности вокруг остатка при $\sigma = 1.0$. Электронная плотность покрывает хорошо весь остаток, кроме радикала, что отражено при $\sigma = 1.0$ и только «усугубляется» при $\sigma = 1.5$. Ситуация довольно спорная, но я бы сделала вывод, что в данном случае это – ошибка расшифровки.

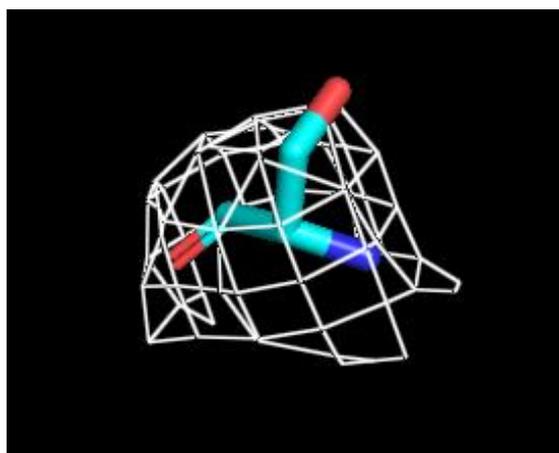


Рис. 12 – Изображение электронной плотности вокруг Ser411 при $\sigma = 1.0$

Сравнение моделей PDB и PDB_redo

В данном подпункте я сравнивала модель из PDB с оптимизированной моделью, которая содержится в базе PDB_redo [8]. Некоторые параметры, по которым я сравнивала модели, представлены в таблице 2. ТРНК, входящие в состав комплексов (и файлов PDB, PDB_redo) не сравнивала.

Таблица 2. Сравнение исходной и оптимизированных моделей

Параметр	База данных	
	PDB [данные, согласно PDB_redo]	PDB_redo
R	0,2586	0,3214
R _{free}	0,2955	0,382
R _{free} - R	0,0369	0,0606
Число остатков, вписанных в электронную плотность лучше	55	48

Сравнение моделей, произведенное на сайте PDB_redo, показывает, что по многим параметрам модель из PDB соизмерима с моделью из PDB_redo или даже лучше. Например, оптимизация привела к тому, что 48 остатков модели из PDB_redo вписываются в электронную плотность лучше, а 55 – хуже.

Также я решила подать структуру из PDB_redo на вход сервису MolProbity. Результаты представлены на рисунке 13. По одному параметру (Bad bonds) модели совпадают, по остальным – модель из PDB оказывается лучше модели PDB_redo.

All-Atom Contacts	Clashscore, all atoms:	11.34	97 th percentile* (N=37, 3Å - 9999Å)
	Clashscore is the number of serious steric overlaps (> 0.4 Å) per 1000 atoms.		
Protein Geometry	Poor rotamers	49	12.34% Goal: <0.3%
	Favored rotamers	287	72.29% Goal: >98%
	Ramachandran outliers	8	1.78% Goal: <0.05%
	Ramachandran favored	408	90.67% Goal: >98%
	MolProbity score [^]	2.93	89 th percentile* (N=895, 3.29Å ± 0.25Å)
	CB deviations >0.25Å	4	0.94% Goal: 0
	Bad bonds:	0 / 3776	0.00% Goal: 0%
Bad angles:	6 / 5107	0.12% Goal: <0.1%	
Peptide Omegas	Cis Prolines:	2 / 23	8.70% Expected: ≤1 per chain, or ≤5%
	Twisted Peptides:	7 / 454	1.54% Goal: 0
Low-resolution Criteria	CaBLAM outliers	21	4.74% Goal: <1.0%
	CA Geometry outliers	6	1.35% Goal: <0.5%

Рис. 13 – Данные, полученные сервисом MolProbity, для оптимизированной структуры

На основе выданных сервисов PDB_redo и MolProbity можно сделать вывод, что структура из базы PDB значительно лучше оптимизированной структуры из базы PDB_redo.

Выводы

Структура с PDB-ID: 4kr2, с которой я работала, имеет довольно невысокое (среднее или даже ниже) качество расшифровки структуры. Несмотря на то, что модель лучше других со схожим разрешением, у нее довольно много маргиналов. Я предполагаю, что одна из причин, по которым качество расшифровки структуры – среднее – это разрешение структуры, а не особенности комплекса. Возможно, будь разрешение выше, качество было бы намного лучше. Также было показано, что оптимизация с помощью PDB_redo не требуется.

Список литературы

- [1] W. Freist, D. T. Logan, and D. H. Gauss, "Glycyl-tRNA synthetase.," *Biol. Chem. Hoppe. Seyler.*, vol. 377, no. 6, pp. 343–356, Jun. 1996.
- [2] D. E. Andreev, J. Hirnet, I. M. Terenin, S. E. Dmitriev, M. Niepmann, and I. N. Shatsky, "Glycyl-tRNA synthetase specifically binds to the poliovirus IRES to activate translation initiation," vol. 40, no. 12, pp. 5602–5614, 2012.
- [3] L. A. Nangle, W. Zhang, W. Xie, X. Yang, and P. Schimmel, "Charcot – Marie – Tooth disease-associated mutant tRNA synthetases linked to altered dimer interface and neurite distribution defect," vol. 104, no. 27, 2007.
- [4] "RCSB PDB - 4KR2_ Glycyl-tRNA synthetase in complex with tRNA-Gly." .
- [5] X. Qin, Z. Hao, Q. Tian, Z. Zhang, C. Zhou, and W. Xie, "Cocrystal structures of glycyl-tRNA synthetase in complex with tRNA suggest multiple conformational states in glycylation," *J. Biol. Chem.*, vol. 289, no. 29, pp. 20359–20369, Jul. 2014.
- [6] "MolProbity." .
- [7] "Z-Score for 4kr2." .
- [8] "<https://pdb-redo.eu>." .