

МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
имени М.В.ЛОМОНОСОВА

ФАКУЛЬТЕТ БИОИНЖЕНЕРИИ И БИОИНФОРМАТИКИ

**Отчет по расшифровке структуры
бифункционального белка GImU из *S. pneumoniae***

*Выполнил студент 4-го курса
Волик Павел*

Москва

2017 г.

Аннотация

Отчет представляет собой анализ качества расшифровки структуры 4aaw бифункционального белка GlmU из организма *Streptococcus pneumoniae*.

Введение

N-ацетилглюкозамин-1-фосфат уридилтрансфераза/глюкозамин-1-фосфат ацетилтрансфераза (GlmU) – бифункциональный фермент (гомотример), который катализирует реакцию образования УДФ-N-ацетилглюкозамина (UDP-GlcNAc), который входит в состав пептидогликанов и липополисахаридов грамположительных и грамотрицательных бактерий. На первой стадии реакции С-концевой домен GlmU катализирует перенос ацетильной группы с кофермента А на глюкозамин-1-фосфат (GlcN-1-P) с образованием N-ацетилглюкозамин-1-фосфата (GlcNAc-1-P). На второй стадии N-концевой домен GlmU катализирует перенос УМФ с УТФ на GlcNAc-1-P с образованием UDP-GlcNAc и пиррофосфата [1]. Схема данной реакции приведена на Рис. 1.

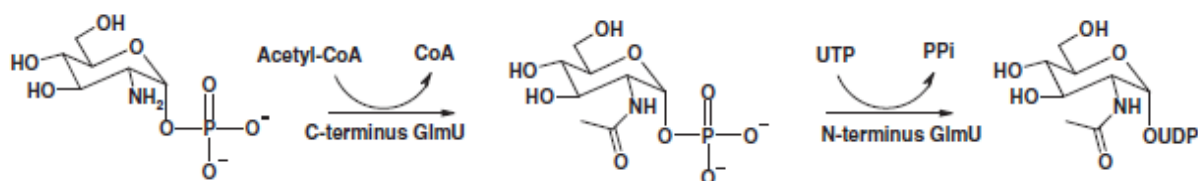


Рис. 1. Схема реакции, катализируемой GlmU [1].

Целью данной статьи ([1]) являлось нахождение ингибитора ацетилтрансферазного домена фермента GlmU посредством сканирования серии химических соединений и выявления среди них тех, которые демонстрировали антибактериальную активность. Была расшифрована структура данного фермента, которая подтвердила, что найденный ингибитор (R84) связывается в сайте, занимаемом субстратом ацетил-КоА (конкурентное ингибирование).

Результаты и обсуждения

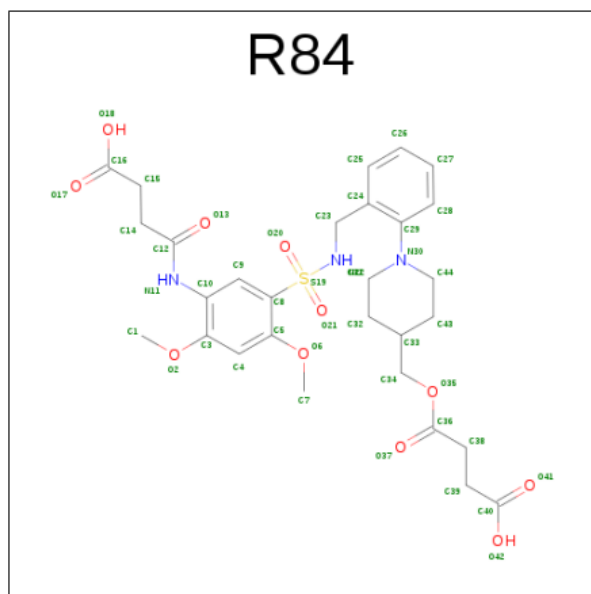
Общая информация о модели

Структура бифункционального белка GlmU в комплексе с ингибитором была расшифрована в 2012 году, расшифровку описывает статья «Inhibitors of acetyltransferase domain of N-acetylglucosamine-1-phosphate uridyltransferase/glucosamine-1-phosphate-acetyltransferase (GlmU). Part 1: Hit to lead evaluation of a novel arylsulfonamide series» (Green OM, McKenzie AR, Shapiro AB, Otterbein L, Ni H, Patten A, Stokes S, Albert R, Kawatkar S, Breed J) [1].

Метод решения фазовой проблемы – метод молекулярного замещения.

В состав кристалла помимо белка входят сульфат-анион и ингибитор данного белка – R84. Его структурная формула приведена на Рис. 2.

Molecule 2 is 4-{{[1-(2-{{[{{5-[(3-CARBOXYPROPANOYL)AMINO]-2,4-DIMETHOXYPHENYL}SULFONYL)AMINO]METHYL}PHENYL)PIPERIDIN-4-YL]METHOXY}-4-OXOBUTANOIC ACID (three-letter code: R84) (formula: C₂₉H₃₇N₃O₁₁S).



Полнота набора структурных факторов в этом диапазоне: 99,83%.

Параметры кристаллографической ячейки:

CRYST1 116.779 116.779 116.683 90.00 90.00 120.00 P 3 2 1 6

Первые три числа – длины направляющих векторов, следующие три – углы между ними.

Некристаллографические симметрии присутствуют. Белок представляет собой гомотример [2]. Асимметрическая единица включает в себя 6 субъединиц (или 2 гомотримера).

Значения индикаторов качества модели в целом

На Рис. 3 представлены характеристики структуры 4aaw по отношению ко всем PDB-структурам (темные символы), полученным методом PCA, и по отношению к структурам близкого разрешения (светлые символы). В целом, структура 4aaw расшифрована лучше, чем структуры близкого разрешения, и примерно так же (или лучше по отдельным критериям), чем все структуры PDB.

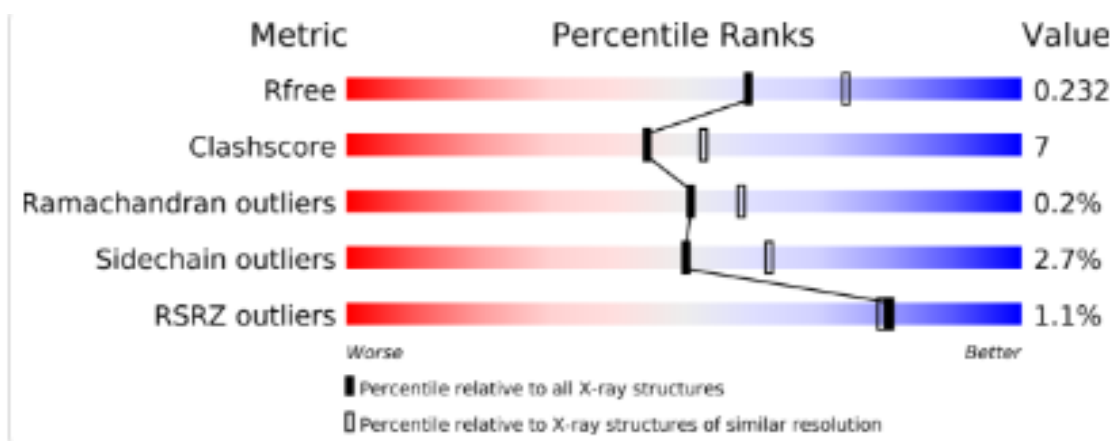


Рис. 3. Характеристика расшифровки структуры 4aaw.

R-фактор показывает, насколько модель соответствует экспериментальным данным. R-Free более точно описывает качество модели и позволяет учесть переоптимизацию, так как его считают по контрольным рефлексам, которые не используют для создания модели.

R-фактор: 0,20164

R-Free: 0,23662

R-Free < 25%, разность R-free – R < 10%. Это позволяет говорить о том, что у модели хорошее качество и переоптимизации не наблюдается.

Пространственный R-фактор (RSR) показывает насколько хорошо каждый аминокислотный остаток соответствует экспериментальным данным, то есть, насколько хорошо остаток в модели вписывается в электронную плотность (Рис. 4).

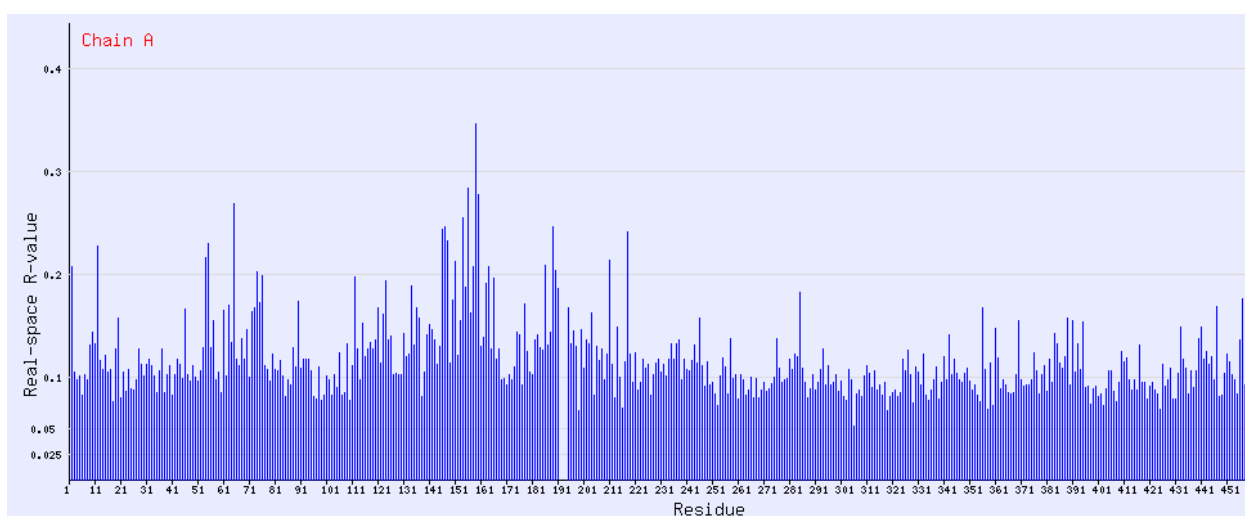


Рис. 4. RSR остатков (сервис EDS [4])

Видна четкая маргинальность (очень большое значение RSR) у остатка Thr159.

Оценка RSRZ (Рис. 5) демонстрирует значение RSR конкретного остатка в сравнении с таковой оценкой для аналогичных остатков.

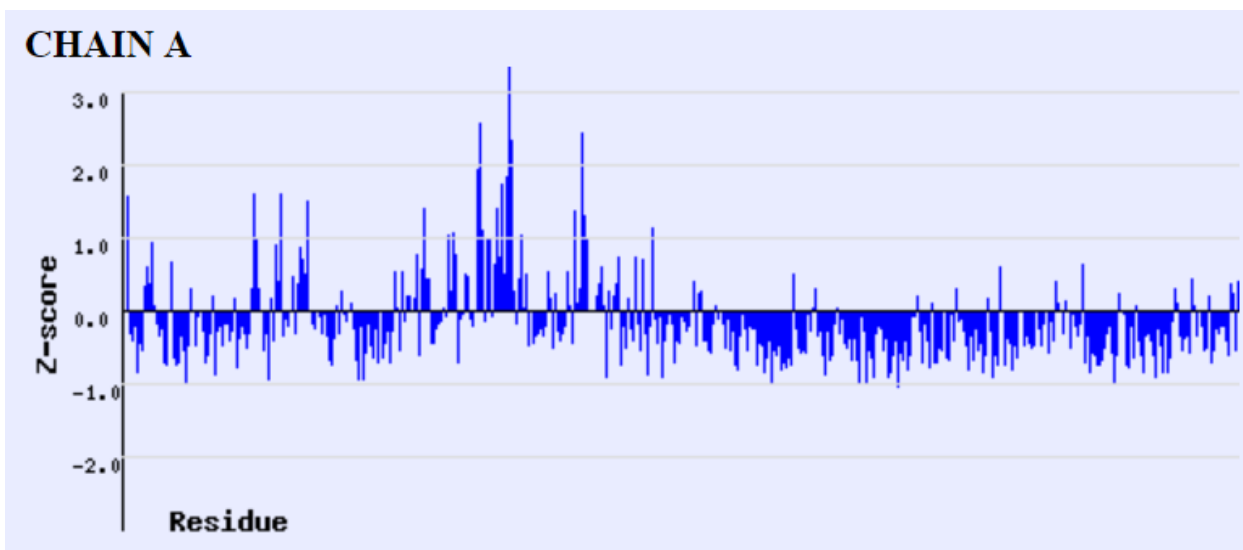


Рис. 5. RSRZ остатков (сервис EDS [4]).

Видна маргинальность остатков Thr159, Ala147, Thr189.

Выдача сервиса MolProbity [3] (Рис. 6).

All-Atom Contacts	Clashscore, all atoms:	7.05	97 th percentile* (N=456, 2.20Å ± 0.25Å)	
	Clashscore is the number of serious steric overlaps (> 0.4 Å) per 1000 atoms.			
Protein Geometry	Poor rotamers	14	3.73%	Goal: <0.3%
	Favored rotamers	339	90.40%	Goal: >98%
	Ramachandran outliers	1	0.22%	Goal: <0.05%
	Ramachandran favored	437	96.90%	Goal: >98%
	MolProbity score [^]	2.01		85 th percentile* (N=10167, 2.20Å ± 0.25Å)
	CB deviations >0.25Å	5	1.21%	Goal: 0
	Bad bonds:	11 / 3497	0.31%	Goal: 0%
Bad angles:	8 / 4746	0.17%	Goal: <0.1%	
Peptide Omegas	Cis Prolines:	2 / 13	15.38%	Expected: ≤1 per chain, or ≤5%
	Cis nonProlines:	1 / 440	0.23%	Goal: <0.05%

Рис. 6. Выдача сервиса MolProbity.

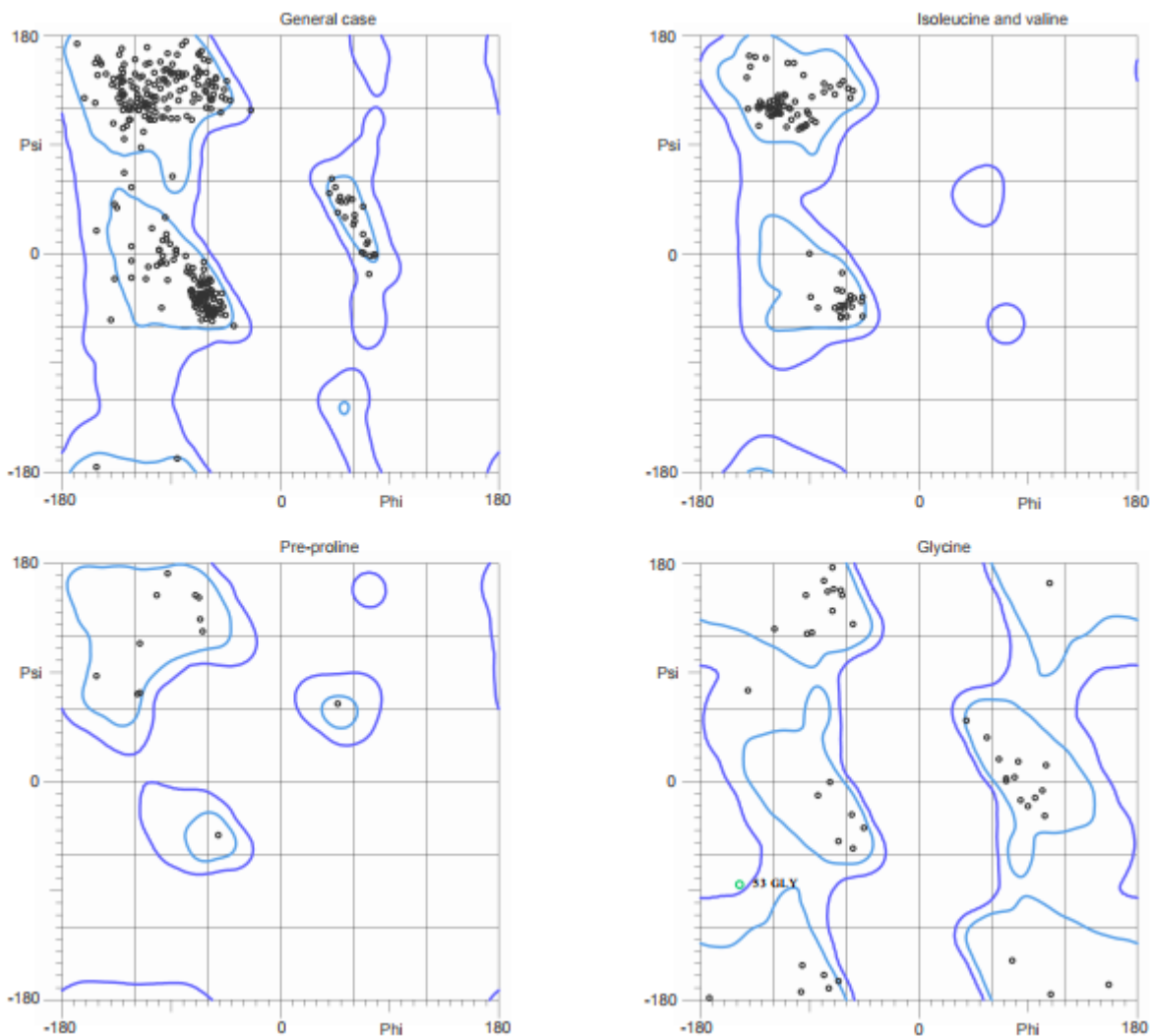
Результат выдачи показывает нелучшие показатели оценок качества данной структуры.

Тем не менее, Clashscore (число недопустимых стерических перекрытий атомов на 1000 (т.е. 5 значит 0.5% атомов); перцентиль данной структуры по отношению к структурам примерно такого же разрешения) довольно неплохой – 97-ой (сотый перцентиль отражает наилучшее качество для

данного индикатора среди моделей с близким разрешением, нулевой – наихудшее).

Стоит отметить присутствие одного маргинального остатка, который не является пролином и находится в цис-конформации – Ala147, и несколько остатков (14) с неблагоприятной конформацией боковых цепей, в частности: Asn345, Lys389, Val412, Thr450.

Карта Рамачандрана, построенная с помощью сервиса MolProbity (Рис. 7), выявила всего один маргинальный остаток – Gly53. Процент маргинальных остатков (расположенных в запрещенных областях): 0,2%.



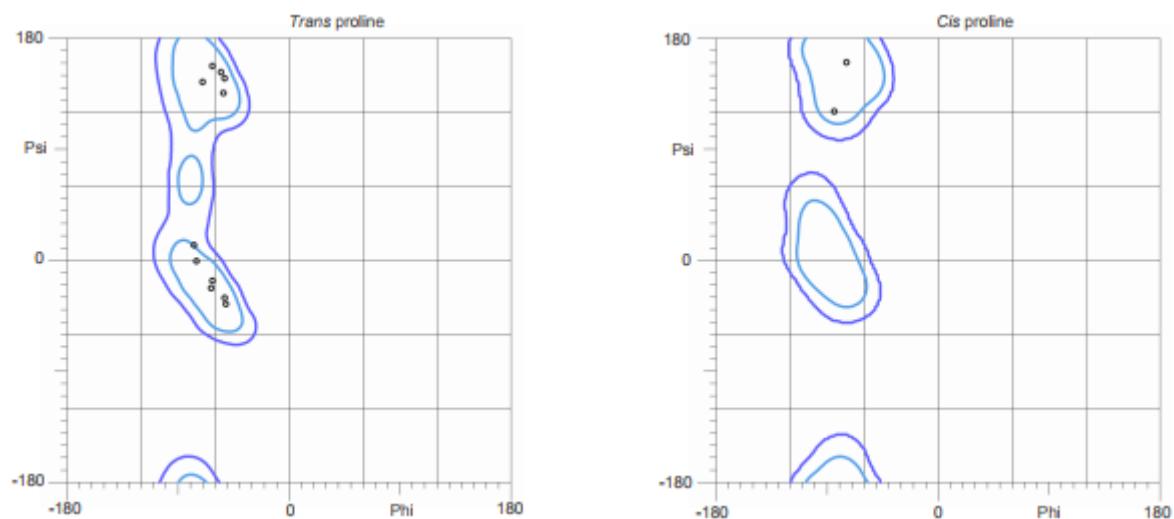


Рис. 7. Карты Рамачандрана для обычных остатков; для Пе и Val; для остатков, расположенных перед Pro; для Gly, для Pro в транс- и цис-конформациях. Разрешенные области обведены синим, благоприятные – голубым.

Построение карты Рамачандрана с помощью сервиса EDS (Рис. 6) выявило 2,2% маргинальных остатков (лежащих в недопустимых областях).

Выявленные маргиналы представлены в Табл. 1.

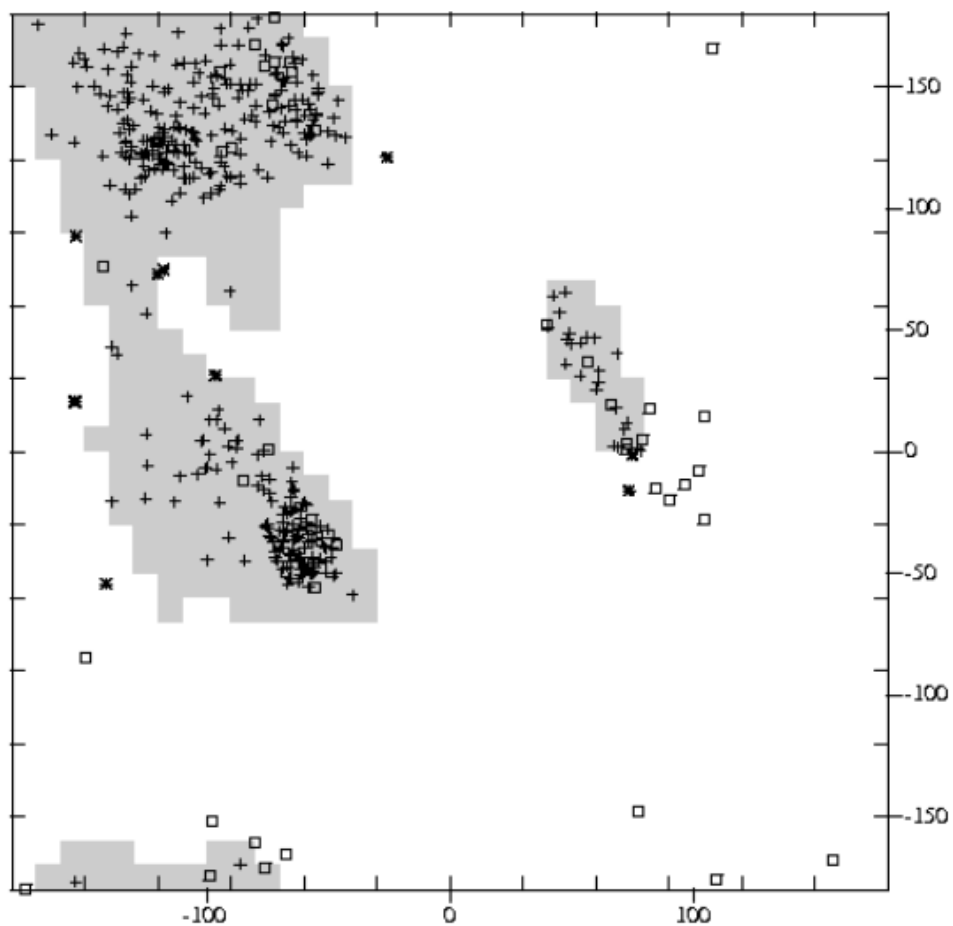


Рис. 8. Карта Рамачандрана, построенная с помощью сервиса EDS.

Таблица 1. Маргинальные остатки, выявленные по карте Рамачандрана (сервис EDS)

Residue	Type	Phi Angle	Psi Angle
A64	ALA	-25.9	120.9
A76	LEU	-154.0	20.4
A103	THR	-119.9	73.1
A134	ASN	-117.4	74.7
A190	ASN	-96.3	31.4
A256	ASN	-153.6	88.4
A357	ASN	75.0	-1.4
A389	LYS	-141.2	-54.5
A447	GLU	73.9	-16.1

Маргинальные остатки

В таблице 2 представлены некоторые маргинальные остатки и критерии, по которым они были выявлены.

Таблица 2. Некоторые маргинальные остатки из структуры 4aaw

№	остаток	Показатель маргинальности
1	Thr159	RSRZ
2	Ala147	RSRZ, цис-конформация
3	Thr189	RSRZ
4	Gly53	Карта Рамачандрана
5	Ala64	Карта Рамачандрана
6	Leu76	Карта Рамачандрана
7	Asn345	Боковая цепь
8	Lys389	Боковая цепь
9	Val412	Боковая цепь
10	Thr450	Боковая цепь

Анализ некоторых маргинальных остатков

Thr189

Показатель маргинальности – очень высокое значение RSRZ, равное 3.2 [2]

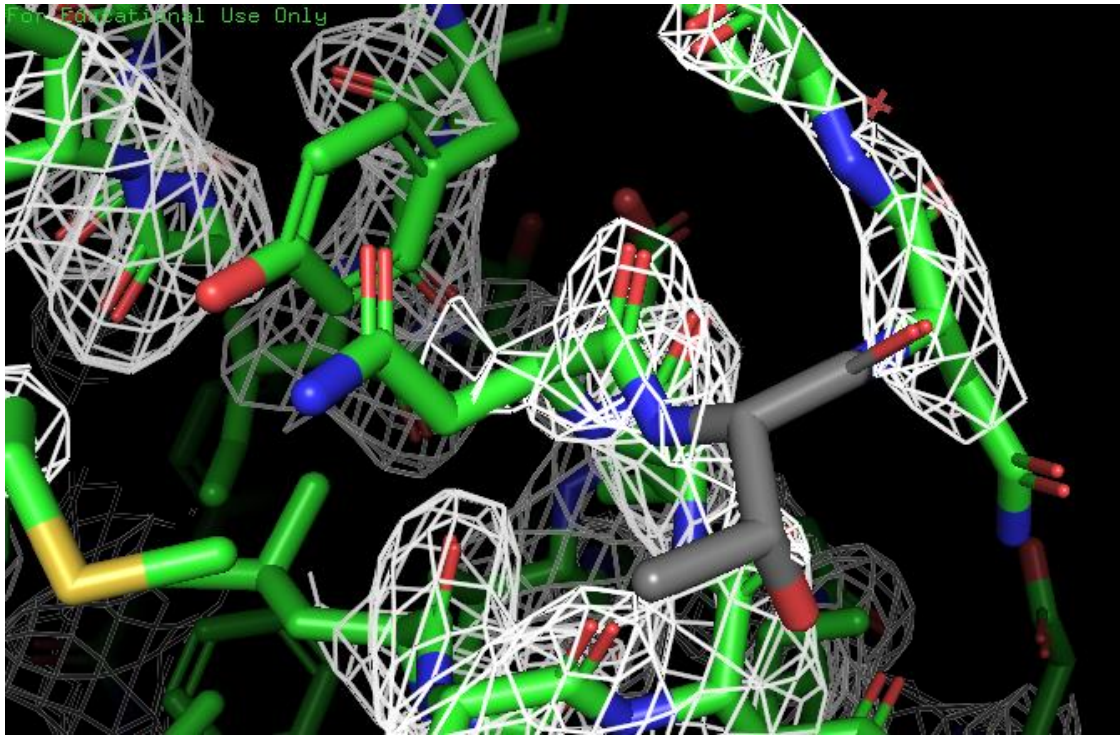


Рис. 9. Остаток Thr189 и его окружение. Электронная плотность показана на уровне подрезки $1,5\sigma$.

Видим, что остаток очень плохо вписывается в электронную плотность. Причина маргинальности скорее всего ошибка расшифровки. Это предположение подкрепляет то, что данный остаток располагается в альфа-спирали на поверхности глобулы белка, и, вероятно, имеет большую подвижность (Рис. 10).

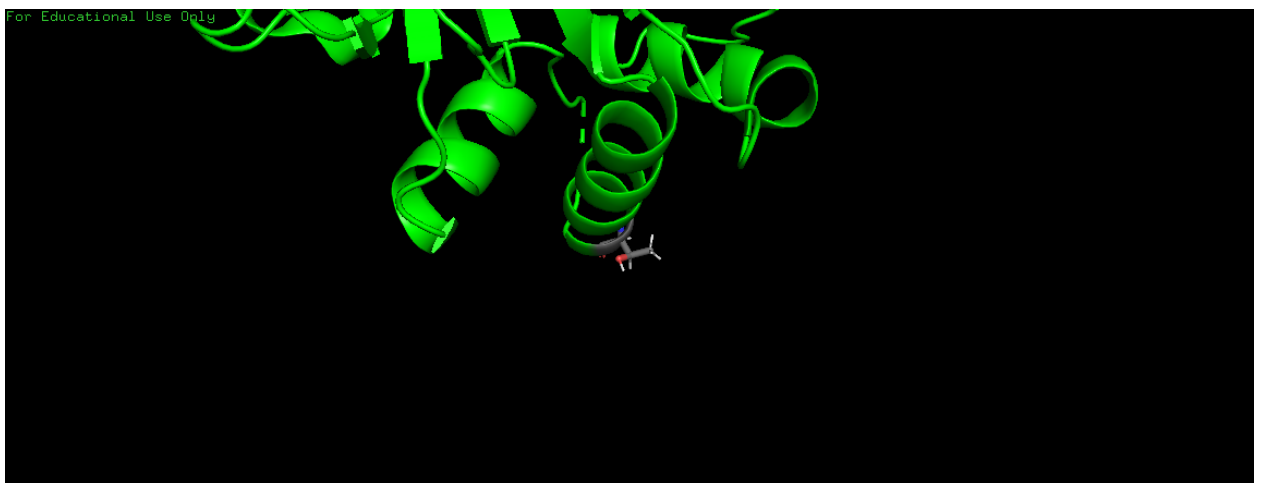


Рис. 10. Положение Thr189 в структуре белка.

Thr159

Показатель маргинальности – высокое значение RSRZ: 3,3 [2].

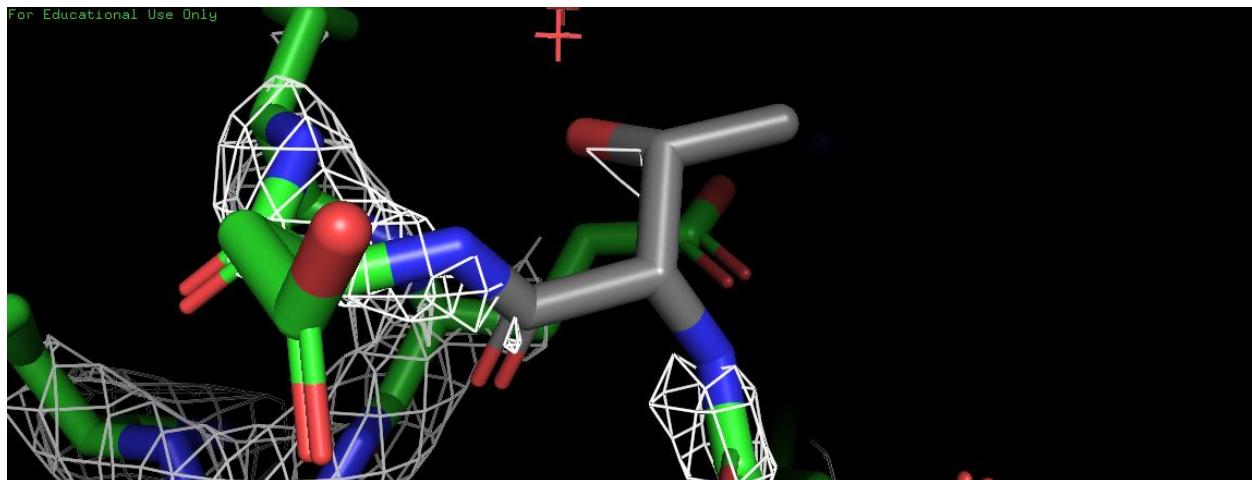


Рис. 11. Остаток Thr159 и его окружение. Электронная плотность показана на уровне подрезки $1,5\sigma$.

Аналогично предыдущему случаю, остаток очень плохо вписывается в электронную плотность. Вероятно, это тоже ошибка расшифровки. Остаток расположен также на поверхности глобулы (Рис. 12)

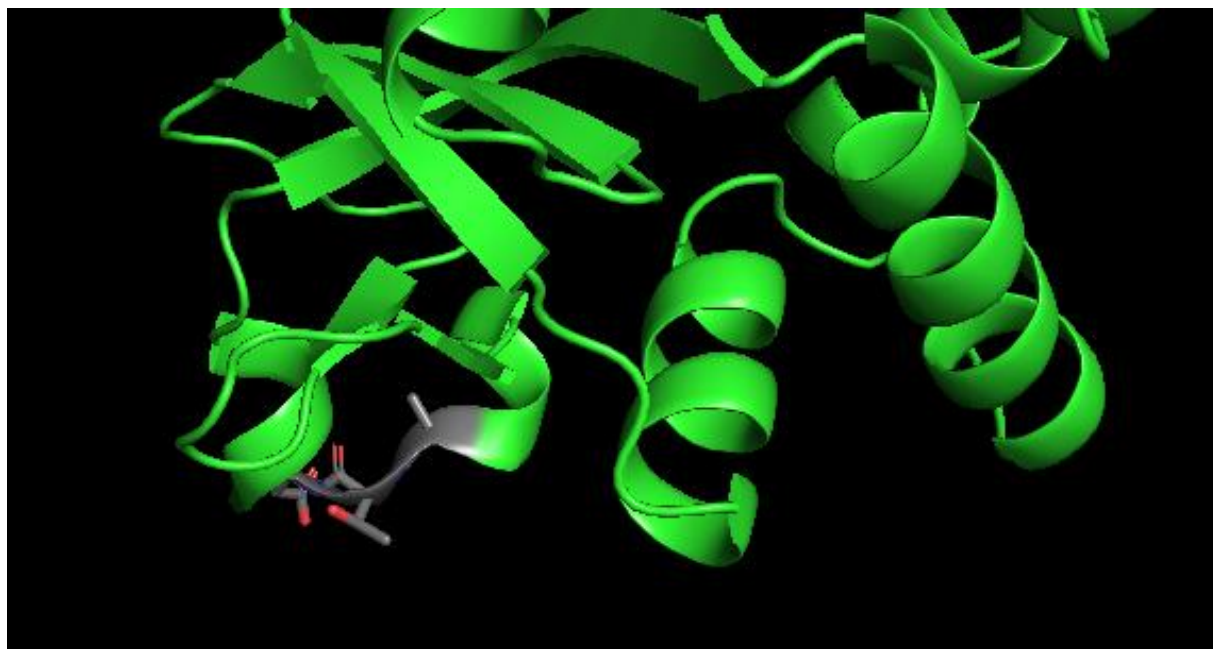


Рис. 12. Положение Thr159 в структуре белка.

Ala147

Маргинален сразу по двум показателям – сервис EDS выявил для данного остатка высокий RSRZ (больше двух), а также цис-конформацию (Рис. 13).

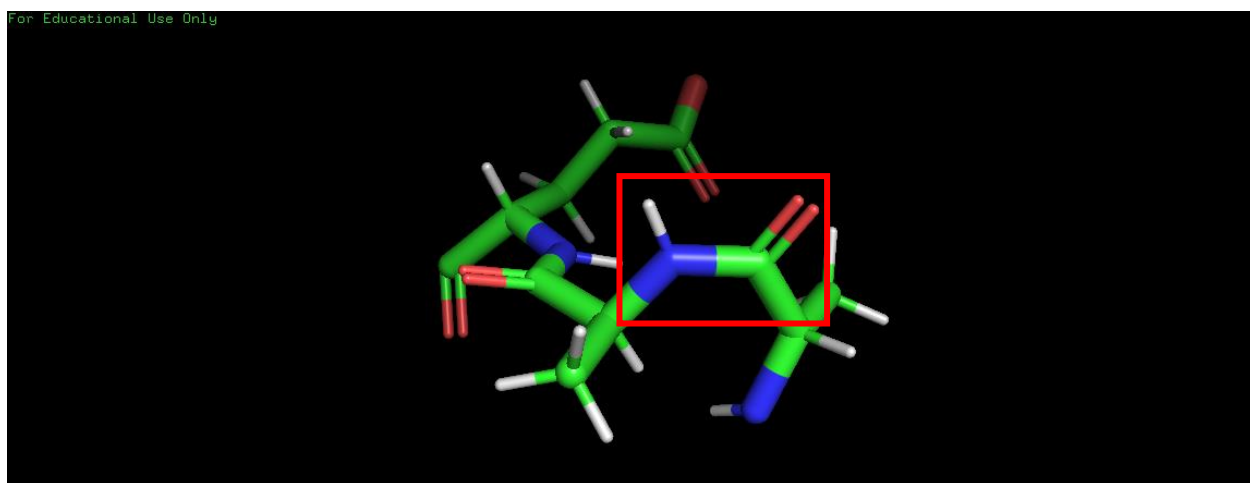


Рис. 13. Ala147 в цис-конформации.

Рис. 14 показывает, что данный остаток очень плохо описывается электронной плотностью.

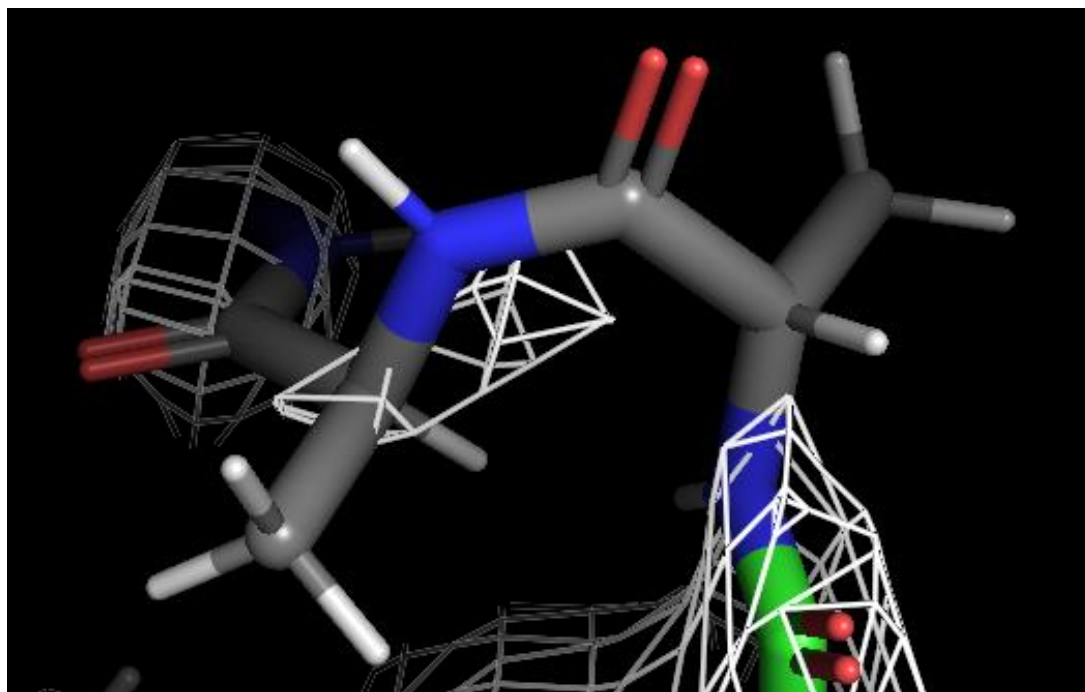


Рис. 14. Остаток Ala147 и его окружение. Электронная плотность показана на уровне подрезки $1,5\sigma$.

Непролиновые цис-конформации в общем случае запрещены. Лишь в очень редких случаях остатки могут действительно находиться в структуре белка в цис-конформации, но в том случае это имеет важное функциональное значение, в частности, такие остатки могут играть ключевую роль в стабилизации активного центра [5]. Но в данном случае это скорее не особенность, а опять же ошибка расшифровки, так как электронная плотность очень плохо описывает остаток, и он располагается в петле на поверхности белка, то есть имеет большую подвижность (Рис. 15)

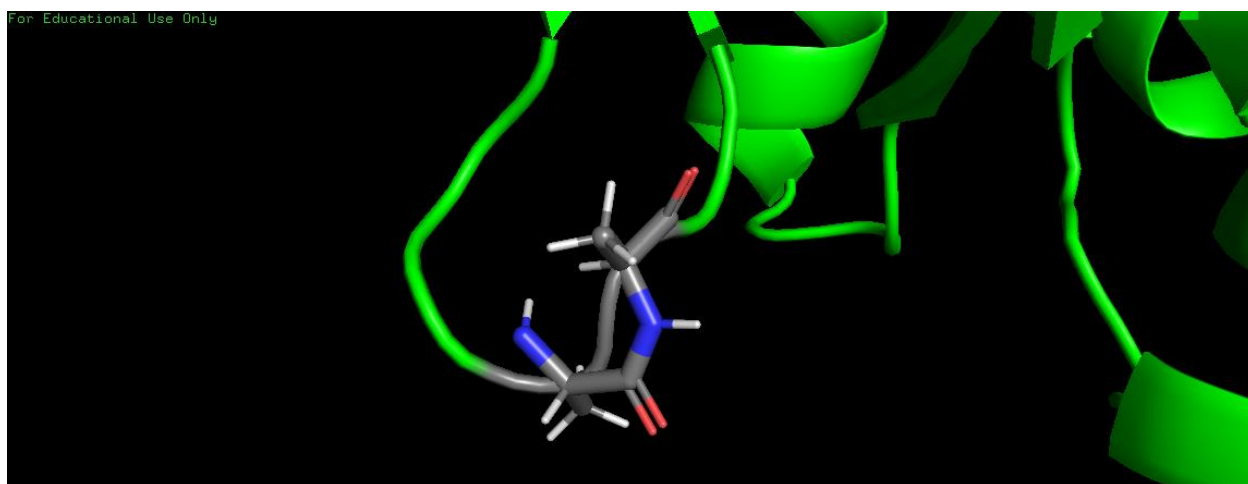


Рис. 15. Расположение остатка Ala147 в структуре 4aaw.

Ala64

Маргинальный признак – положение в недопустимой области карты Рамачандрана (сервис EDS). Такое возможно, если остаток расположен близко с другими остатками белка, например, при контакте между субъединицами, или ионами металла, входящих в структуру белка в качестве кофакторов. Однако в данном случае это ошибка расшифровки, так как остаток плохо описывается электронной плотностью и располагается на поверхности глобулы белка (Рис. 16).

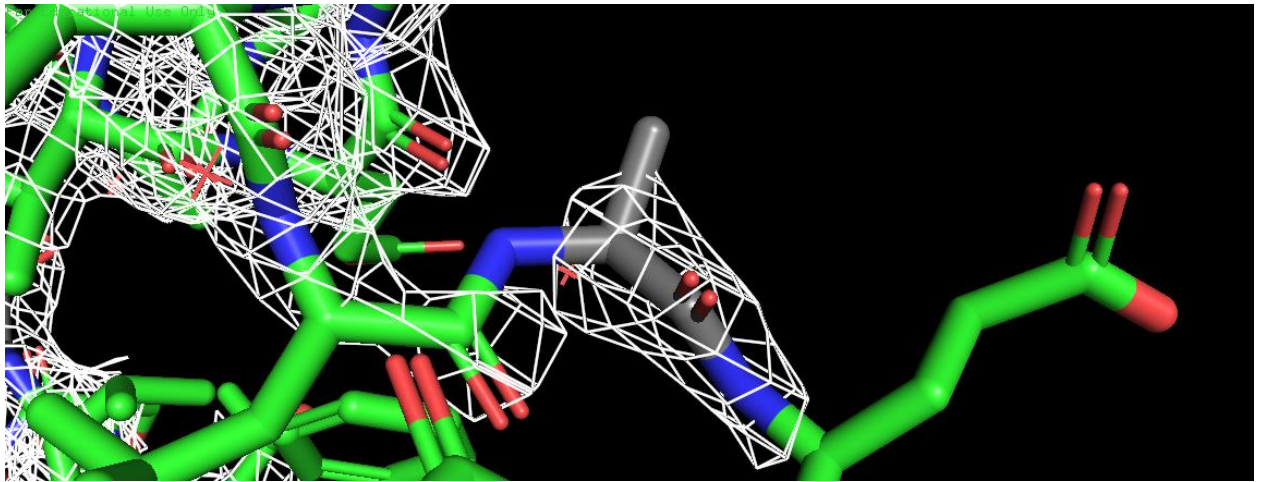


Рис. 16. Остаток Ala62 и его окружение. Электронная плотность показана на уровне подрезки $1,5\sigma$.

Lys389

Маргинальный признак – неpreferred конформация боковой цепи (Sidechain outliers, Рис. 17). Скорее всего это тоже ошибка расшифровки, так как боковая группа данного лизина очень плохо описывается электронной плотностью и этот остаток лежит в петле на поверхности глобулы.

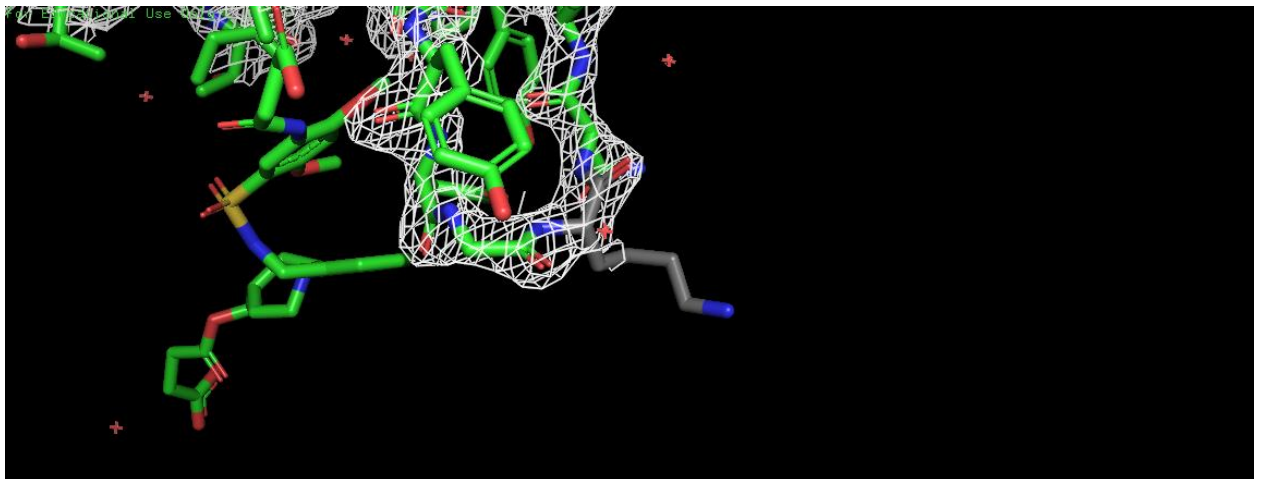


Рис. 17. Остаток Lys389 и его окружение. Электронная плотность показана на уровне подрезки $1,5\sigma$.

Сравнение модели из PDB и PDB_redo

Для проверки того, можно ли улучшить структуру 4AAW, была проанализирована соответствующая запись в PDB_redo [6].

Таблица 3. Сравнение характеристик исходной PDB-структуры и структуры из PDB_redo

<i>Характеристика</i>	<i>4AAW (PDB)</i>	<i>4AAW (PDB_redo)</i>
R-фактор	0.1971	0.1904
R-Free	0.2320	0.2233

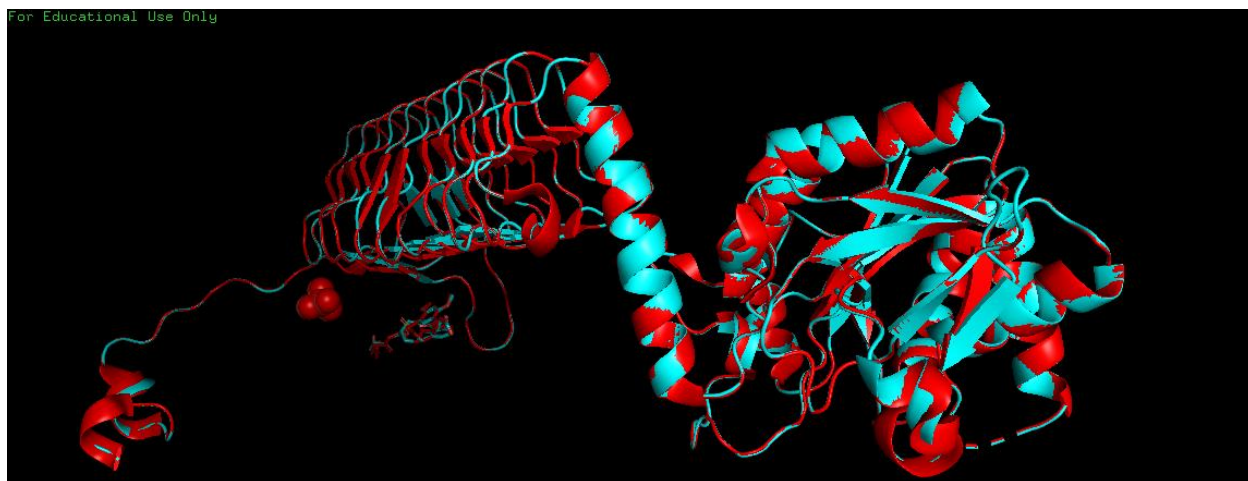


Рис. 18. Наложение исходной структуры 4AAW (красный цвет) и структуры PDB_redo (голубой цвет).

Видим, что значения основных индикаторов качества расшифровки (Табл. 3) в модели PDB_redo немного лучше, чем таковые для исходной структуры. При наложении (Рис. 18) структуры совпадают (практически идентичны)

Заключение

Качество расшифровки структуры 4AAW бифункционального белка GImU из организма *Streptococcus pneumoniae* хорошее. Индикаторы качества в большинстве своем выше, чем у PDB-структур с близким разрешением. Анализ PDB_redo показал, что расшифровка структуры очень близка к оптимальной. В структуре немало маргинальных остатков, но рассмотренные из них являются следствием скорее ошибок расшифровки, чем функциональными особенностями.

Список литературы

- 1 Green OM, McKenzie AR, Shapiro AB, Otterbein L, Ni H, Patten A, Stokes S, Albert R, Kawatkar S, Breed J. Inhibitors of acetyltransferase domain of N-acetylglucosamine-1-phosphate-uridylyltransferase/glucosamine-1-phosphate-acetyltransferase (GImU). Part 1: Hit to lead evaluation of a novel arylsulfonamide series. *Bioorg Med Chem Lett*. 2012 Feb 15;22(4):1510-9.
- 2 https://files.rcsb.org/pub/pdb/validation_reports/aa/4aaw/4aaw_full_validation.pdf
- 3 <http://molprobitry.biochem.duke.edu>
- 4 <http://eds.bmc.uu.se>
- 5 Joseph AP, Srinivasan N, de Brevern AG. Cis-trans peptide variations in structurally similar proteins. *Amino Acids*. 2012 Sep;43(3):1369-81.
- 6 <https://pdb-redo.eu>