

ФАКУЛЬТЕТ БИОИНЖЕНЕРИИ И БИОИНФОРМАТИКИ
Московского Государственного Университета им. М. В. Ломоносова

Отчёт по качеству расшифровки структуры RNase J
из *Bacillus subtilis* (PDB ID 3zq4)

Работу выполнила:
студентка 4 курса
Князева Анастасия

Москва, 2017 г.

Оглавление

Аннотация	3
Введение	3
Результаты и обсуждение	4
Общая информация о модели.....	4
Значения индикаторов качества модели в целом	5
Маргинальные остатки	10
Анализ пяти маргинальных остатков	12
Сравнение модели PDB с моделью PDB_redo.....	14
Заключение	16
Список литературы	17

Аннотация

В данном отчете представлена информация о структуре белка RNase J из организма *Bacillus subtilis*, который имеет идентификатор в базе данных PDB 3zq4. Был проведен анализ качества расшифровки структуры на основе показателей качества расшифровки структуры, таких как параметры кристаллической решетки, R-фактор, R_free, карта Рамачандрана, и других, полученных с помощью различных сервисов и баз данных.

Введение

RNase J является важным ферментом в *Bacillus subtilis* с необычной двойной эндонуклеазной и 5'-3'-экзонуклеазной активностью, которые играют важную роль в созревании и деградации mRNA. Rnase J также является компонентом недавно идентифицированной деградосомы в *B. subtilis*.

В 2011 году была расшифрована кристаллическая структура RNase J1 из *B. subtilis* с разрешением 3,0 Å [1]. Анализ этой структуры показал, что белок находится в открытой конформации, подходящей для связывания субстратной RNA. RNase J является членом семейства β-CASP цинк-зависимых металло-β-лактамаз. Это сходство было использовано при построении модели для RNase J1:RNA комплекса. Структура была расшифрована для анализа стекинга взаимодействий с консервативными ароматическими боковыми цепями, которые обеспечивают молекулярную основу для наблюдаемой активности фермента. Сравнение структуры RNase J из *B. subtilis* с родственными ферментами выявляет ключевые различия, которые дают представление о конформационных изменениях во время катализа и роли С-концевого домена.

Биологическая единица совпадает с ассиметрической ячейкой и представляет из себя гомотетрамер, состоящий 4х одинаковых мономеров. Структура фермента представлена на рис. 1.

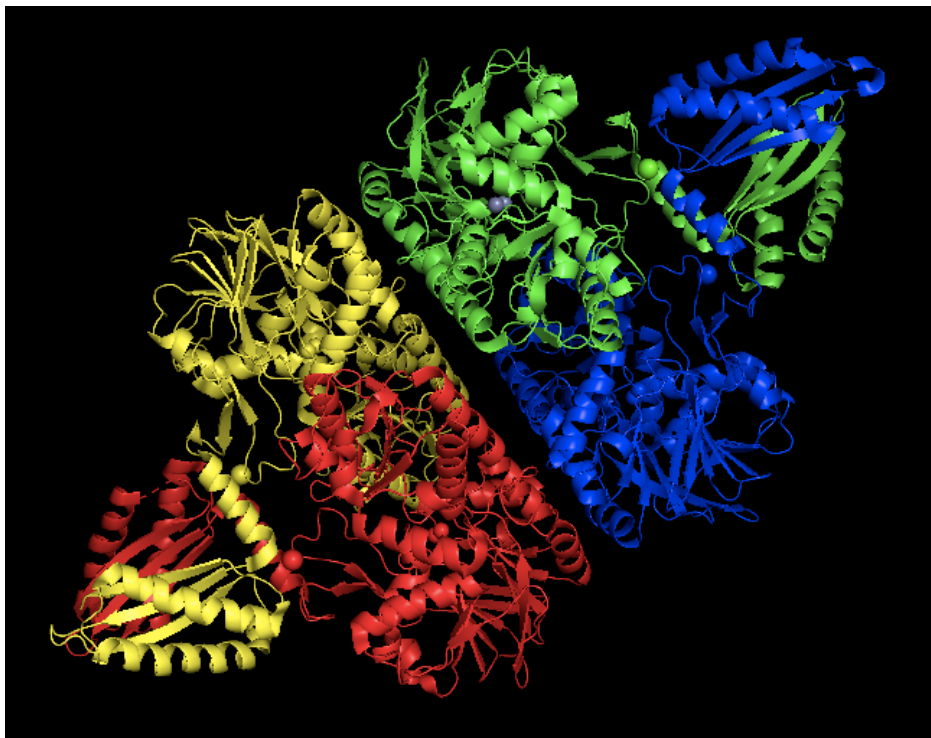


Рис. 1. Изображение структуры белка 3zq4.

Результаты и обсуждение

Общая информация о модели

Статья с расшифровкой структуры RNase J1 из *B. subtilis* вышла 7 сентября 2011 года в журнале *Structure* PMID: 21893285 [1]. Авторами данной статьи являются Joseph A. Newman, Lorraine Hewitt, Cecilia Rodrigues, Alexandra Solovyova, Colin R. Harwood, Richard J. Lewis.

Для получения структуры 3zq4 белок был закристаллизован вместе с ионами цинка (8 Zn²⁺) и кальция (4 Ca²⁺).

Структура RNase J1 из *B. Subtilis* была получена методом рентгено-структурного анализа. В качестве метода решения фазовой проблемы применялся метод молекулярного замещения в программе Phaser [2], используя данные о индивидуальных доменах из структуры RNase J из *T. thermophiles*.

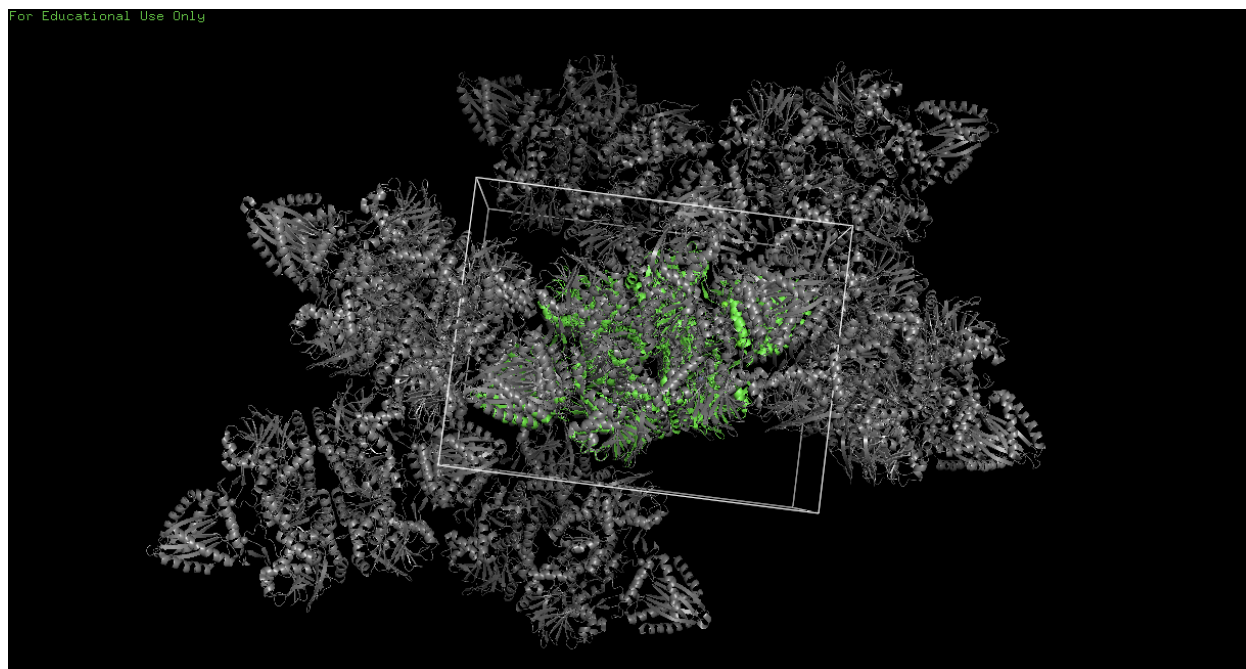
Авторами было получено 55828 рефлексов при полноте 99.2%. Структура имеет разрешение 3.00 Å, диапазон разрешений структурных факторов равен 29.76 – 3.00 Å.

Кристаллографические характеристики из записи PDB (поле CRYST1):

```
CRYST1 177.823 209.430 74.106 90.00 90.00 90.00 P 21 21 2 16
```

Длины направляющих векторов ячейки - (177.823, 209.430 и 74.106). Углы между направляющими векторами ячейки - (90.00, 90.00 и 90.00) - ортогональная система координат. Кристаллографическая группа - P 21 21 2. Число молекул в ячейке - 16. Структура кристалла RNase J1 в разных проекциях представлена на рис. 2.

В ассиметрической ячейке есть некристаллографическая симметрия.



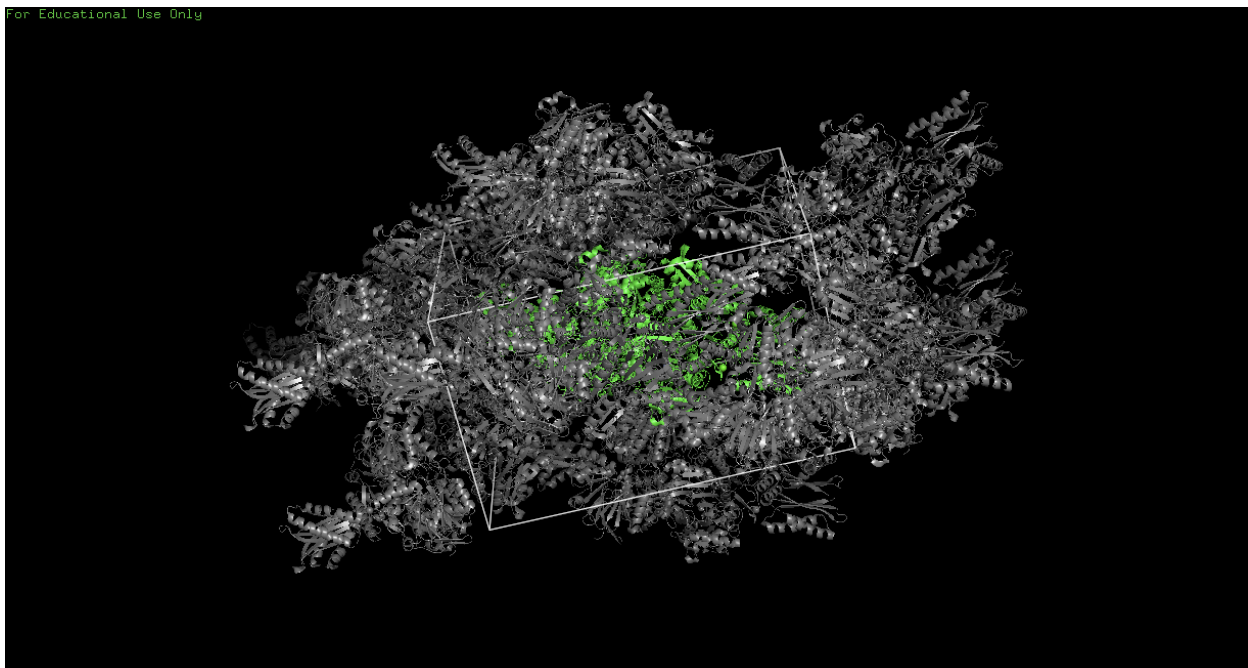


Рис. 2. Структура кристалла RNase J1 в разных проекциях (построена функцией *symexr* с отображением соседних структур на расстоянии 5 Å). Цветом выделена одна асимметрическая единица и изображены границы элементарной ячейки.

Значения индикаторов качества модели в целом

Основные значения различных индикаторов качества модели приведены на сайте PDB в виде графа (рис. 3).

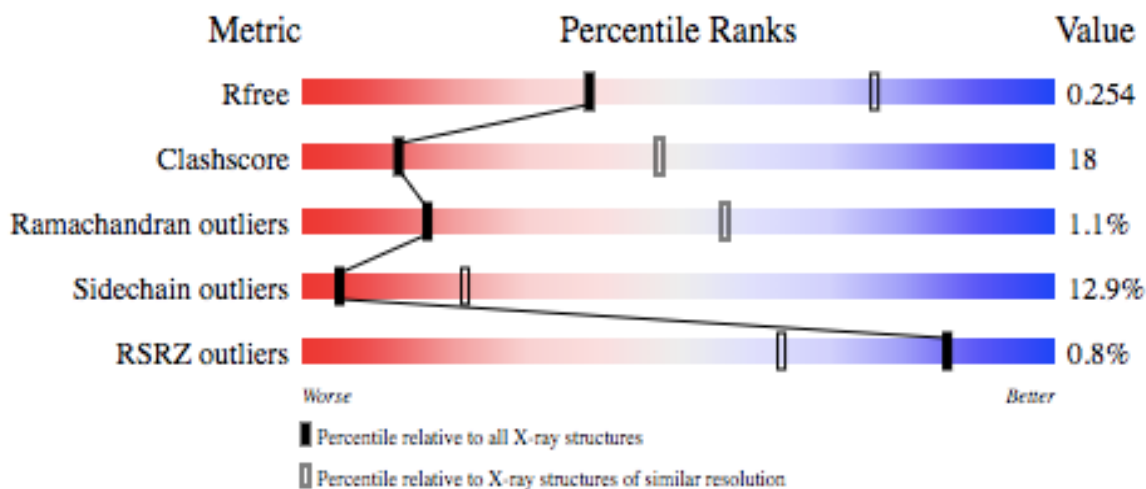


Рис.3. Граф, содержащий основные показатели качества модели 3zq4. Белые риски – процентиль структур с таким же значением показателя, как и 3zq4, и таким же разрешением. Черные риски – процентиль структур с таким же значением показателя, как и 3zq4 по всему банку PDB.

R-фактор характеризует соответствие модели экспериментальным данным – структурным факторам. Рассчитывается он по формуле:

$$R = \frac{\sum_{hkl} |F_{hkl}^{calc} - F_{hkl}^{obs}|}{\sum_{hkl} F_{hkl}^{obs}} * 100\%$$

где F_{hkl}^{calc} – рассчитанные модули структурных факторов, F_{hkl}^{obs} – измеренные в эксперименте модули структурных факторов. При оптимизации модели R-фактор минимизируется: чем меньше R-фактор, тем модель лучше. Значение R-фактора 3zq4 равно 0.214.

Показатель R_{free} нужен для того, чтобы модель не была переоптимизирована только под R-фактор. Рассчитывается R_{free} так же как и R-фактор, но по небольшой группе контрольных рефлексов, которые не участвуют в оптимизации. В данном анализе контрольная группа составляла 5.07% всех рефлексов. Значение R_{free} 3zq4 равно 0.258. Значение $R_{free} - R = 0.258 - 0.214 = 0.044$, что меньше 10%, это показывает, что переоптимизации модели быть не должно. Значение $R_{free} = 0.258$ примерно равно 25%, что является хорошим значением.

Для построения карты Рамачандрана был использован сервер MolProbity [3]. На рис. 4. представлена карта Рамачандрана для структуры 3zq4.

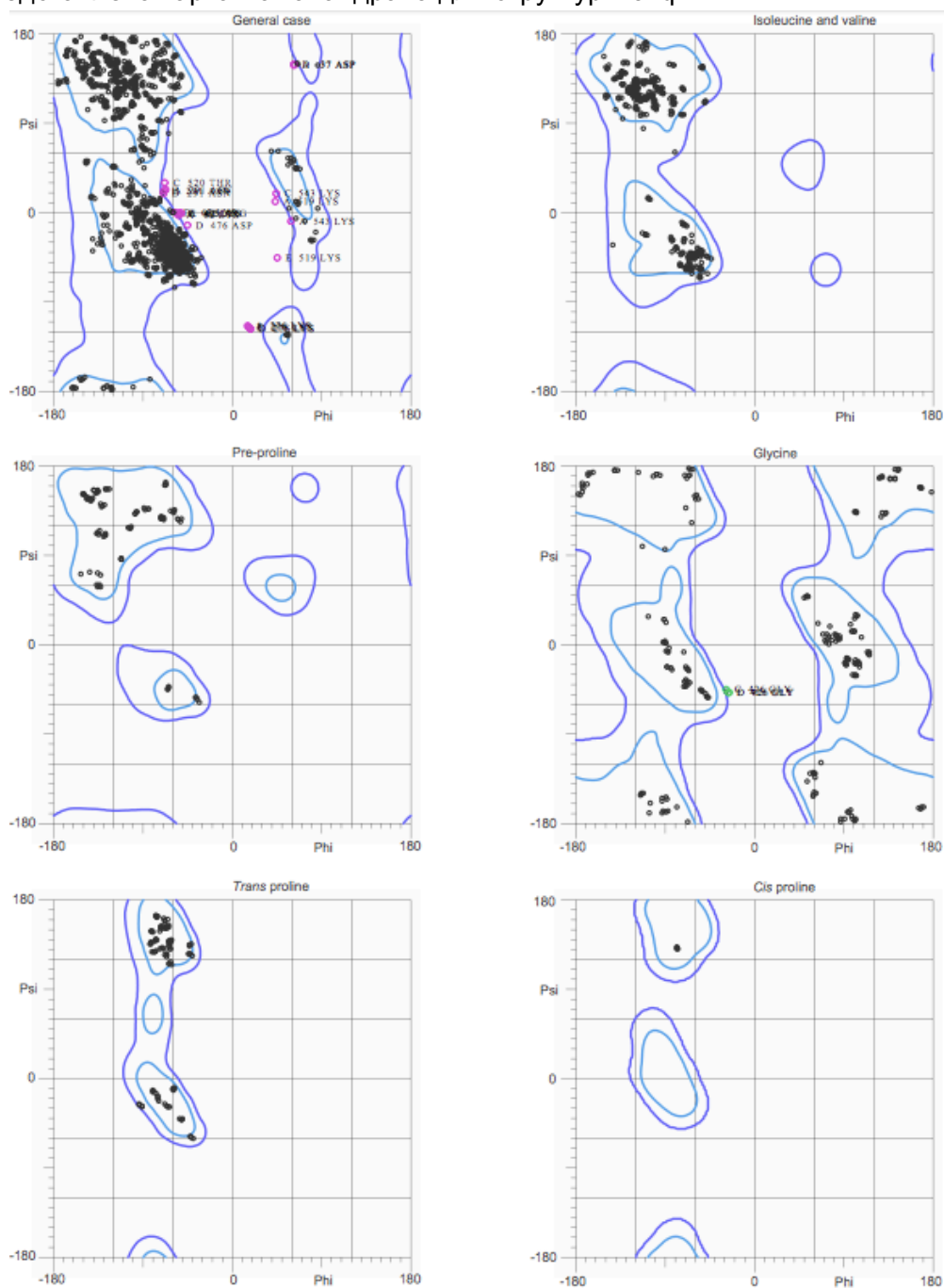


Рис.4. Карта Рамачандрана для 3zq4. В левом верхнем углу Карта Рамачандрана для совокупности всех аминокислот 3zq4. Буквами выделены маргиналы по карте. Разрешенные области обведены синим, благоприятные – голубым.

В структуре 3zq4 присутствует 23 маргинальных по карте Рамачандрана остатка, что составляет 1.07% от всех остатков (ожидаемое значение <0.05% согласно сервису MolProbity). В таблице 1 приведены маргинальные остатки. Другие характеристики качества модели с сервера MolProbity представлены в таблице 2. Видно, что присутствует большое количество значений, выделенных красным, которые совсем не удовлетворяют условиям хорошей модели.

Таблица 1. Маргинальные остатки по карте Рамачандрана.

There were 23 outliers (phi, psi):

A 37 ASP (62.7, 150.9)	D 37 ASP (63.0, 151.5)
A 62 LYS (-55.2, -2.3)	D 62 LYS (-58.4, 1.9)
A 276 LYS (15.6, -114.4)	D 276 LYS (18.6, -117.1)
A 291 ASN (-68.4, 24.4)	D 291 ASN (-70.5, 20.4)
A 495 ARG (-53.1, -0.5)	D 426 GLY (-26.8, -48.5)
A 519 LYS (43.5, 12.9)	D 476 ASP (-46.4, -12.5)
A 543 LYS (59.5, -8.5)	E 62 LYS (-52.4, -0.9)
C 62 LYS (-54.5, 0.7)	E 276 LYS (16.7, -115.6)
C 276 LYS (17.3, -116.4)	E 291 ASN (-70.1, 23.1)
C 426 GLY (-29.7, -45.8)	E 426 GLY (-27.3, -48.8)
C 520 THR (-69.1, 31.0)	E 519 LYS (45.2, -45.5)
C 543 LYS (44.1, 20.4)	

Таблица 2. Критерии качества модели, выданные сервисом MolProbity. Последняя колонка – референсные значения. Зеленым цветом выделены значения, которые удовлетворяют условиям хорошей модели, желтым цветом выделены значения, немного не удовлетворяющие условиям хорошей модели, а красным цветом – значения, которые сильно не удовлетворяющие условиям хорошей модели.

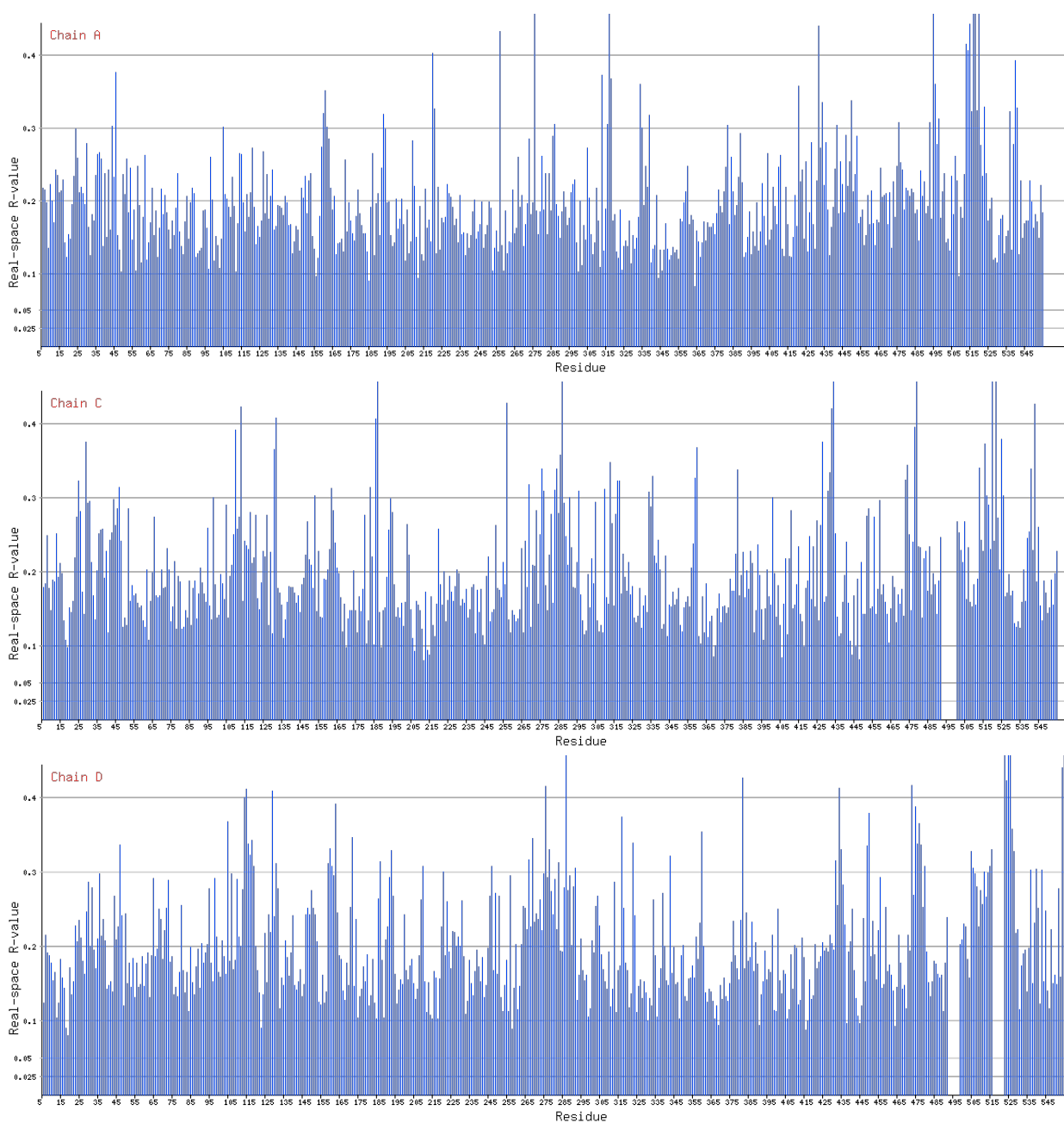
All-Atom Contacts	Clashscore, all atoms:	18.19	96 th percentile* (N=75, 3.00Å ± 0.25Å)
	Clashscore is the number of serious steric overlaps (> 0.4 Å) per 1000 atoms.		
Protein Geometry	Poor rotamers	260	13.96% Goal: <0.3%
	Favored rotamers	1434	77.01% Goal: >98%
	Ramachandran outliers	23	1.07% Goal: <0.05%
	Ramachandran favored	1991	92.26% Goal: >98%
	MolProbity score [^]	3.10	68 th percentile* (N=3130, 3.00Å ± 0.25Å)
	Cβ deviations >0.25Å	12	0.61% Goal: 0
	Bad bonds:	0 / 17196	0.00% Goal: 0%
Bad angles:	4 / 23234	0.02% Goal: <0.1%	
Peptide Omegas	Cis Prolines:	4 / 100	4.00% Expected: ≤1 per chain, or ≤5%
Low-resolution Criteria	CaBLAM outliers	69	3.22% Goal: <1.0%
	CA Geometry outliers	17	0.79% Goal: <0.5%

Показатель качества Clashscore отвечает за количество пар атомов, расположенных слишком близко друг к другу. Выражается в числе таких пар на тысячу атомов. Вычисления Clashscore проводятся на сервисе MolProbity. В данной модели Clashscore равен 18.19, что лежит на уровне 99 перцентили для структур с таким же разрешением.

Пространственный R-фактор позволяет оценить соответствие модели группы атомов “экспериментальной” электронной плотности. Рассчитывается следующим образом:

$$RSR = \frac{\sum_{A \in L} |\rho_{\text{ЭКСП}} - \rho_{\text{МОДЕЛЬ}}|}{\sum_{A \in L} \rho_{\text{ЭКСП}}} [\cdot 100\%]$$

На рис. 5 приведен RSR для остатков цепей структуры 3zq4. Хорошие значения показателя – $RSR < 10\%$. Как видно на предоставленных графиках, довольно много остатков имеют $RSR > 10\%$.



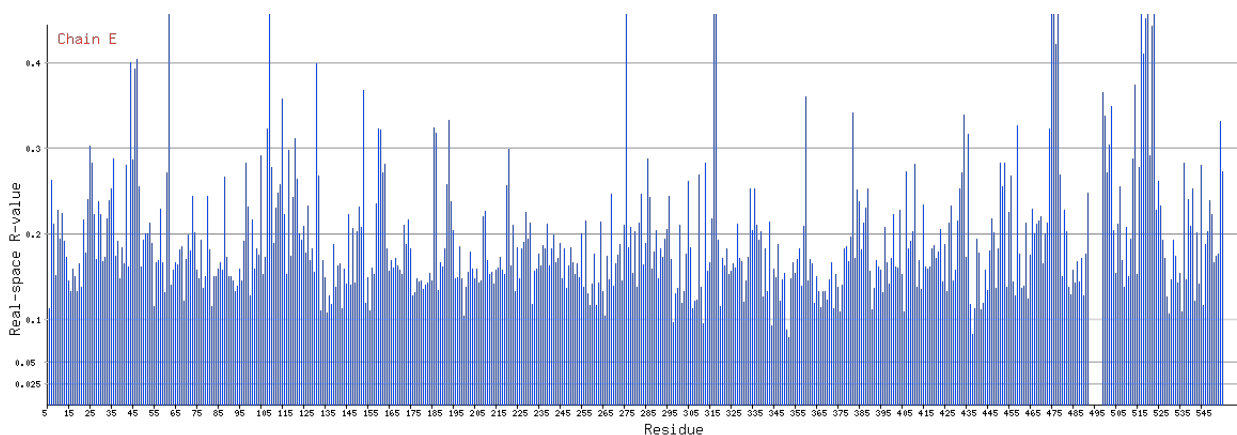


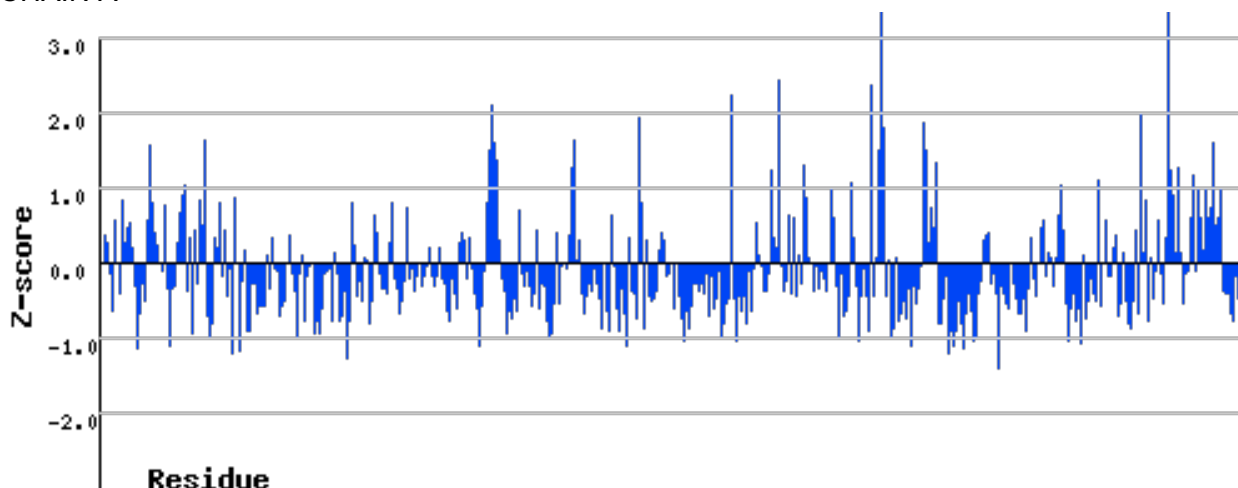
Рис.5. Значения RSR для остатков цепей структуры 3zq4.

RSR-Z – это относительная оценка RSR. Она показывает, насколько RSR для остатка отличается от среднего RSR для такого же типа остатков в структурах PDB со сходным разрешением. Для вычисления RSR-Z остатка его RSR сравнивается со средним RSR для того же типа остатков по выборке из PDB с примерно таким же разрешением, разность делится на $\sigma(\text{RSR})$ по этому типу остатков в таком разрешении.

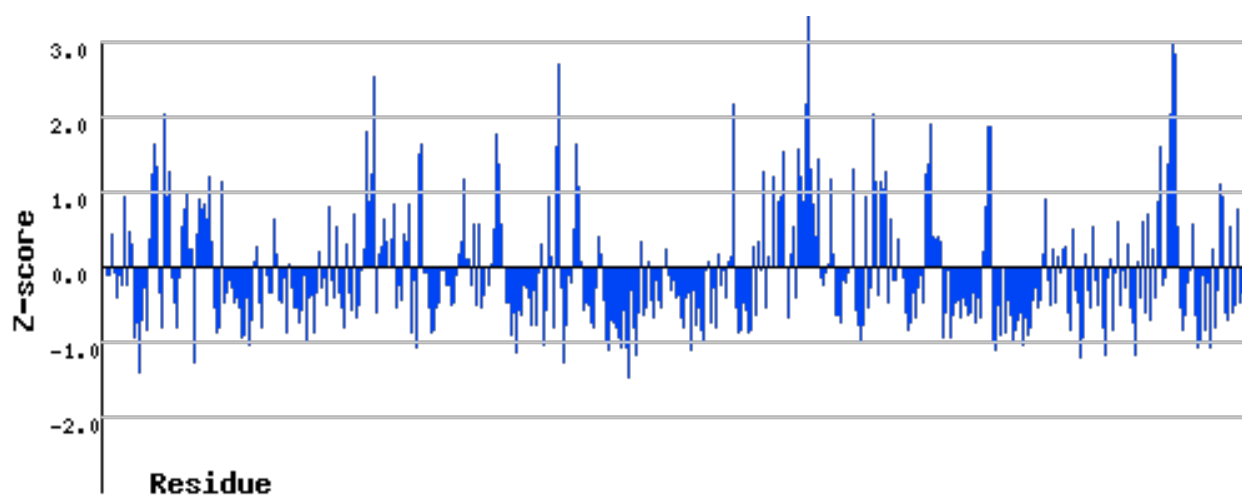
$$Z = (\text{RSR} - \langle \text{RSR}_{\text{resolution}} \rangle) / \text{Sigma}_{\text{resolution}}$$

Если RSR плохой, а RSR-Z – хороший, то значит координаты атомов расшифрованы плохо, но не хуже, чем в других подобных структурах. Высокие положительные значения $\text{RSR-Z} > 2$ свидетельствуют о том, что остаток плохо вписан в электронную плотность по сравнению с другими структурами с тем же разрешением. В целом, $\text{RSR-Z} = 0.8\%$, что выше среднего значения как по всему банку PDB, так и по структурам со сходным разрешением. На рис. 6 приведен RSR-Z для остатков цепей структуры 3zq4. Видно, что очень мало оаастков выходят за пределы значения RSR-Z от -2 до 2.

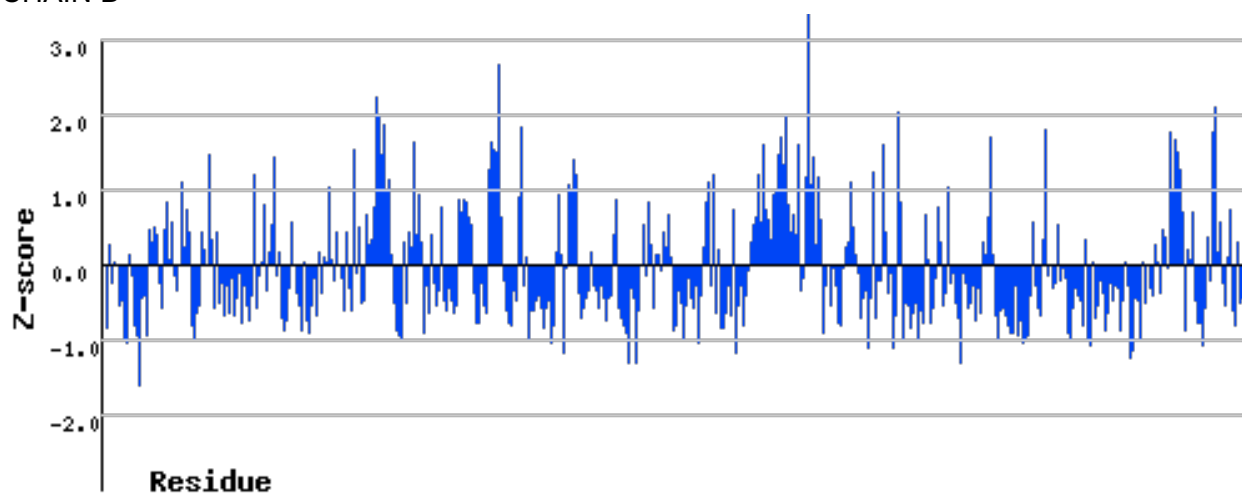
CHAIN A



CHAIN C



CHAIN D



CHAIN E

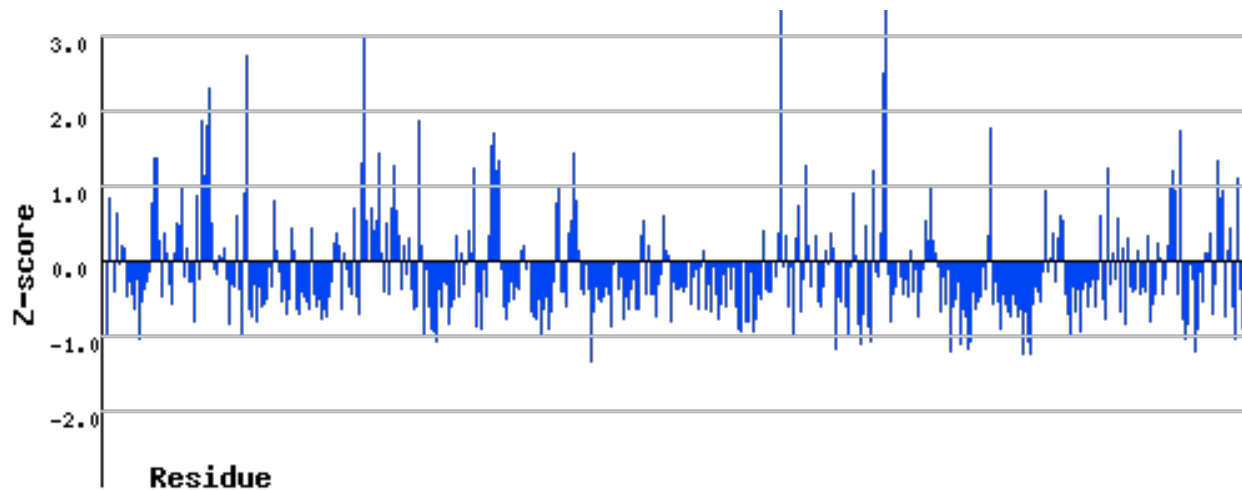


Рис.6. Значения RSR-Z для остатков цепей структуры 3zq4.

Таким образом, кроме очевидно выбивающихся остатков видно, что в целом электронная плотность модели неплохо соответствует экспериментальной.

Маргинальные остатки

Для определения всех маргинальных остатков обратимся к подробному отчету о валидации структуры 3zq4, который можно найти на сайте PDB в поле Full report. В отчете

имеются схемы, представленные на рис. 7 для цепи А в качестве примера, которые показывают, какие остатки обладают отклонениями в своей геометрии по различным показателям и позволяют оценить маргинальность каждого остатка по совокупности показателей. Зеленый цвет остатка в такой схеме означает отсутствие маргинальности по всем показателям, желтый – маргинальность по одному, оранжевый – по двум, красный – по трем и более критериям. Серыми выделены остатки, не представленные в структуре.

Информация о нескольких маргинальных остатках представлена в таблице 3.

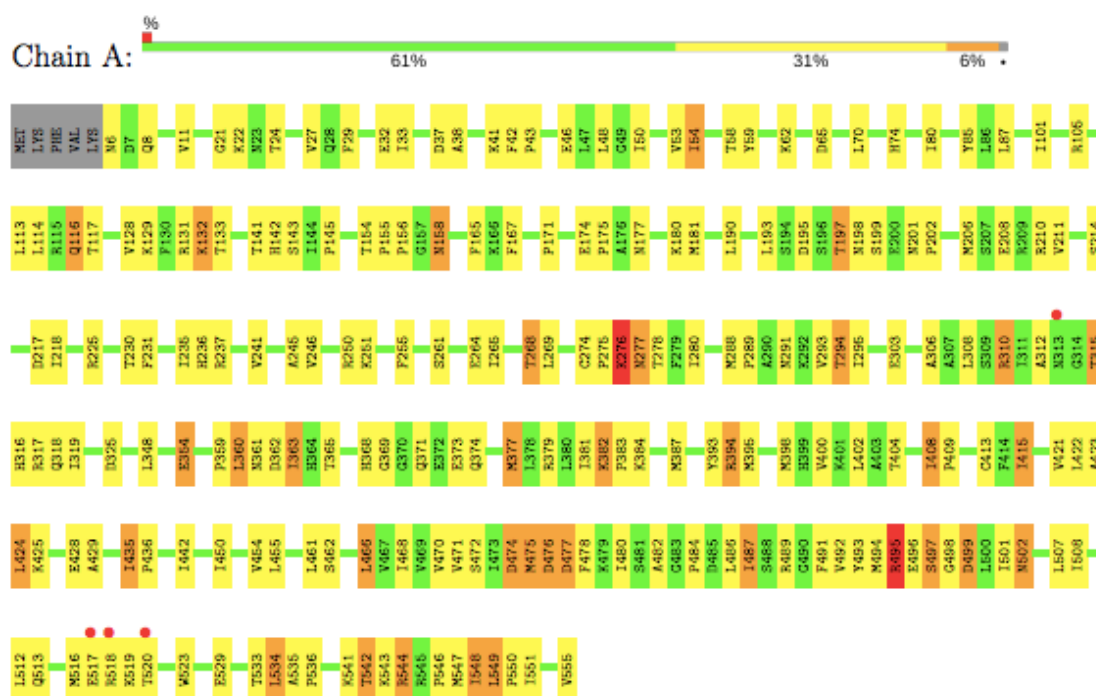


Рис. 7. Карта маргинальности остатков цепи А структуры 3zq4.

Таблица 3. Примеры маргинальных остатков.

№	А. о.	Цепь	Причина маргинальности
1	K276	A	Нахождение в запрещенной области карты Рамачандрана, RSR = 0.487, RSR-Z = 2.41
2	R495	A	Нахождение в запрещенной области карты Рамачандрана, RSR = 0.462
3	D476	E	Нахождение в запрещенной области карты Рамачандрана, слишком близкий контакт внутри одной ассиметрической единицы, RSR = 0.652, RSR-Z = 3.8
4	W523	D	Маргинал по торсионным углам в боковой цепи остатка, RSR = 0.589, RSR-Z = 2.8
5	G426	D	Нахождение в запрещенной области карты Рамачандрана

Анализ пяти маргинальных остатков

Для остатков, представленных в таблице 3, был проведен анализ. В таблице 3 были приведены причины, по которым данный остаток считается маргинальным. На рис. 8 – 12 представлены изображения электронной плотности этих маргинальных остатков.

Как видно из рис. 8, остаток K276 цепи А абсолютно не вписывается в электронную плотность. Вероятно, этот остаток является ошибкой расшифровки.

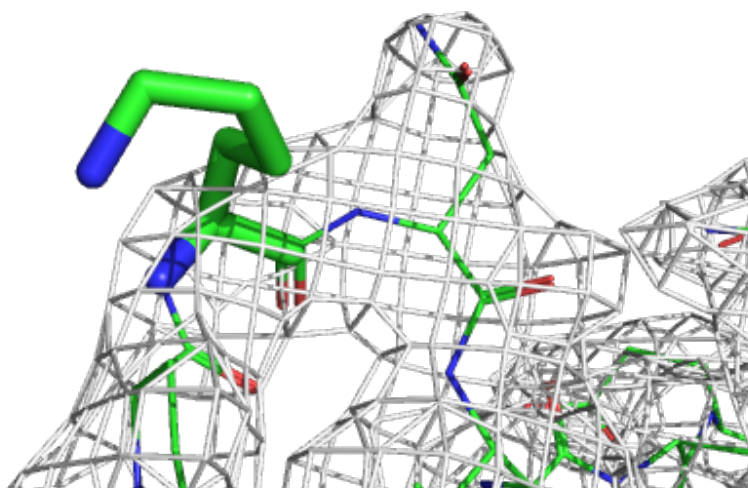


Рис.8. Построение электронной плотности вокруг K276 цепи А с $\sigma=1$.

Изображенный на рис. 9 остаток R495 цепи А только частично покрывается электронной плотностью, причем складывается ощущение, что эта электронная плотность относится к остатку, расположенному в соседнем положении. Если переместить этот остаток в соседнее положение, то кажется, что тогда он хорошо впишется в электронную плотность. По всей видимости, этот остаток является ошибкой расшифровки.



Рис.9. Построение электронной плотности вокруг R495 цепи A с $\sigma=0.5$.

На рис. 10 представлена электронная плотность остатка D476 цепи E. Этот остаток полностью лежит за его электронной плотностью. Скорее всего, это снова ошибка расшифровки.

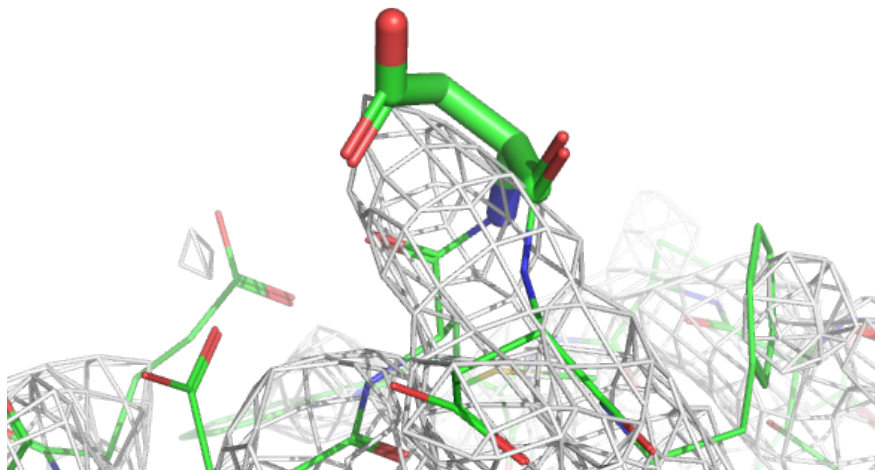


Рис.10. Построение электронной плотности вокруг D476 цепи E с $\sigma=0.5$.

Остаток W523 цепи D достаточно хорошо вписывается в свою электронную плотность, поэтому, скорее всего, маргинальное положение этого остатка это особенность структуры, а не ошибка расшифровки.

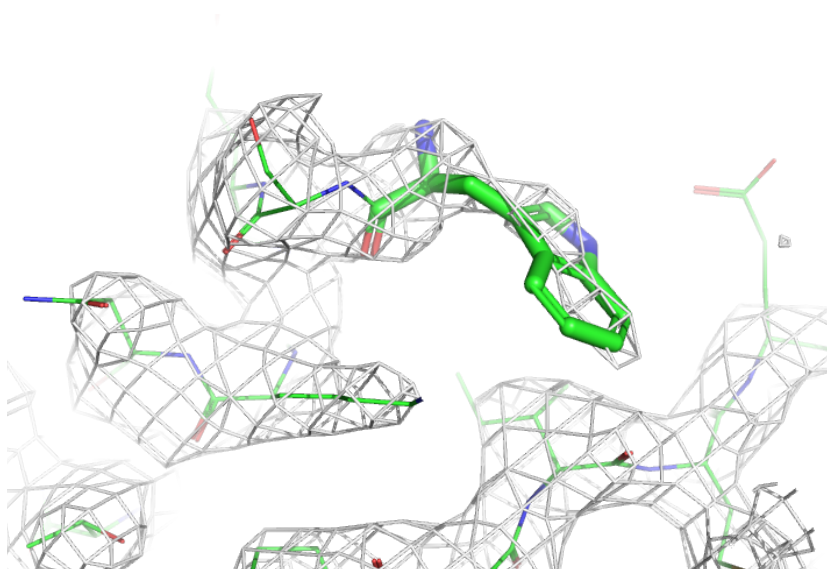


Рис.11. Построение электронной плотности вокруг W523 цепи D с $\sigma=1$.

Остаток G426 цепи D полностью вписывается в электронную плотность. По-видимому, маргинальное положение этого остатка является особенностью структуры, а не ошибкой расшифровки.

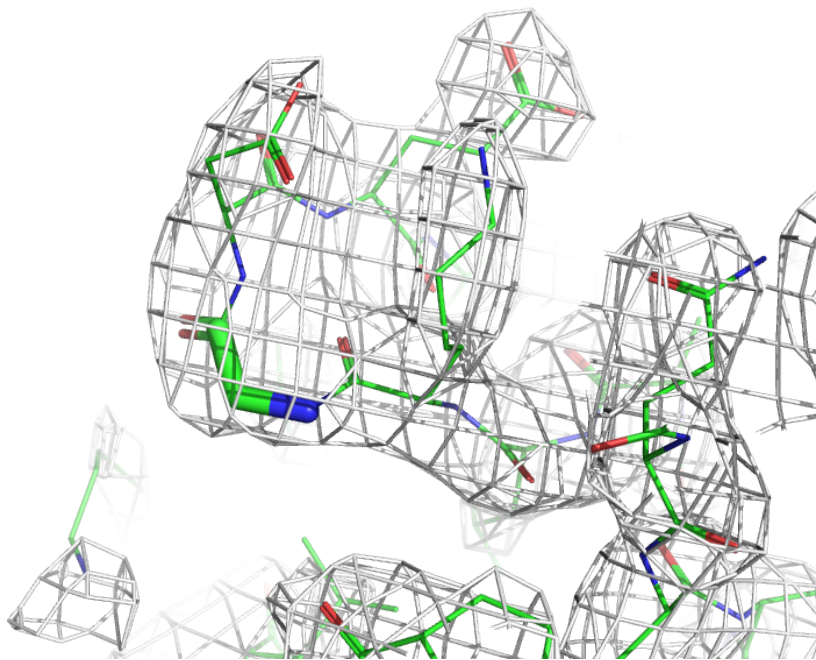


Рис.11. Построение электронной плотности вокруг G426 цепи D с $\sigma=1$.

Сравнение модели PDB с моделью PDB_redo

PDB_redo [4] представляет из себя сервис оптимизации, который позволяет улучшить модель с привлечением ряда параметров, которые используются в процессе улучшения. Общий вид сравнения моделей PDB и PDB_redo представлена на рис. 12. Глобальных изменений в оптимизированной модели не наблюдается, однако есть некоторые различия даже на уровне вторичной структуры. На рис. 13 представлен пример таких различий. В исходной модели петля была представлена в виде β -слоя, хотя на аналогичных участках гомотетрамера эта петля имела вид петли в исходной модели. В модели PDB_redo рассматриваемая петля действительно имеет вид петли, что, кажется, очень разумным. Кроме того, положения других петель немного отличается в исходной модели и модели PDB_redo.

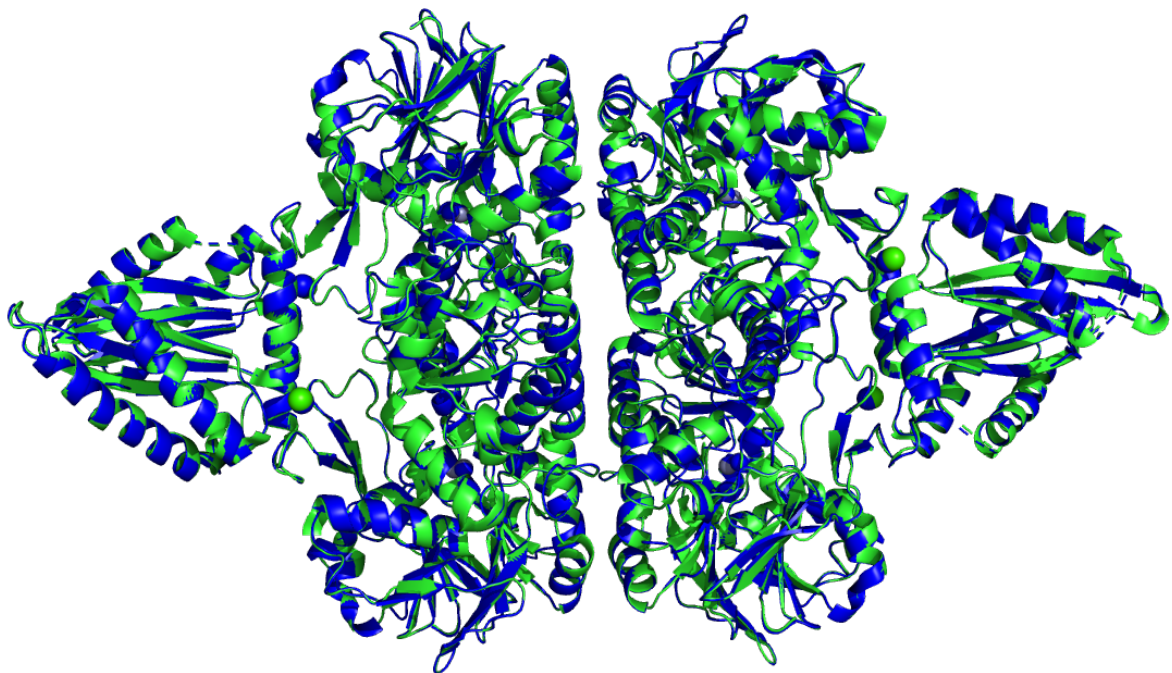


Рис. 12. Сравнение моделей PDB и PDB_redo. PDB модель – зеленая, PDB_redo – синяя.

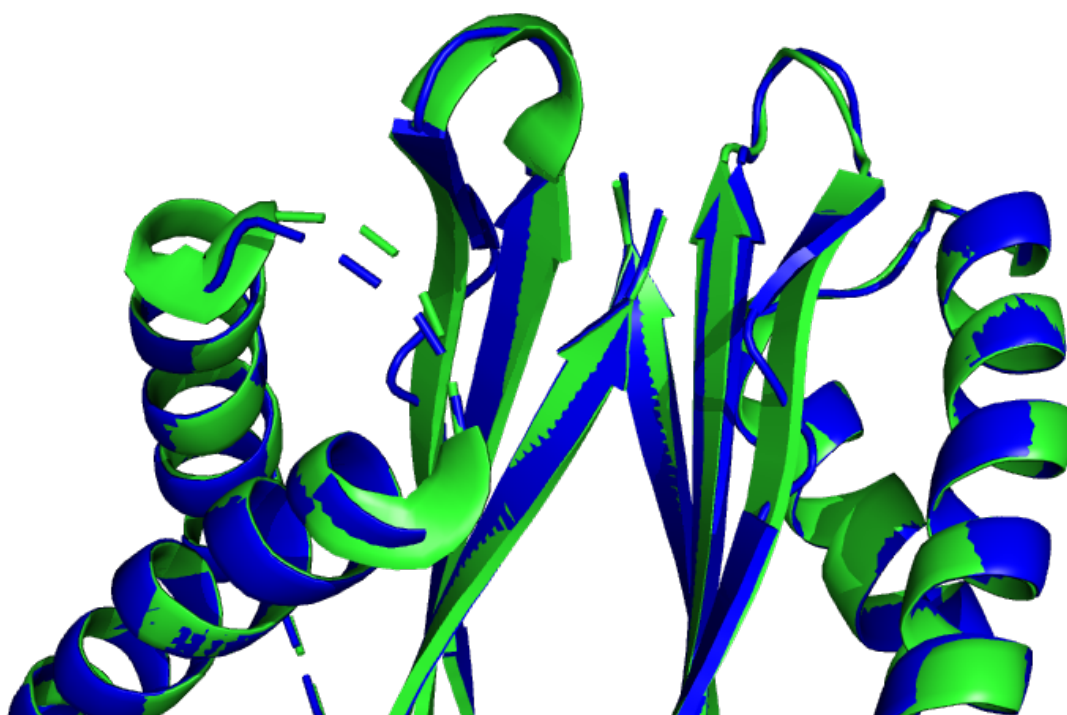


Рис. 13. Частичное сравнение моделей PDB и PDB_redo. PDB модель – зеленая, PDB_redo – синяя.

Для сравнения моделей структура PDB_redo была загружена на сервер MolProbity [3]. Основные характеристики качества модели PDB_redo с сервера MolProbity представлены в таблице 4. Если сравнить эту таблицу с таблицей 2, то можно увидеть, что многие показатели качества модели улучшились, однако даже это не позволило им перейти в зеленую область, т. е. изменения получились незначительными. Единственное значительное улучшение произошло в числе остатков в предпочитаемой области карты Рамачандрана. Количество таких остатков значительно увеличилось, это позволило перейти показателю из красной в желтую область, это не помогло структуре перейти порог в 98%. Это улучшение произошло за счет увеличения неправильных углов в модели, с 0.02% показатель поднялся до 0.16% и перешел из зеленой в красную область.

Таблица 4. Критерии качества модели PDB_redo, выданные сервисом MolProbity.

All-Atom Contacts	Clashscore, all atoms:	5.24		100 th percentile* (N=75, 3.00Å ± 0.25Å)
	Clashscore is the number of serious steric overlaps (> 0.4 Å) per 1000 atoms.			
Protein Geometry	Poor rotamers	147	7.88%	Goal: <0.3%
	Favored rotamers	1579	84.66%	Goal: >98%
	Ramachandran outliers	13	0.60%	Goal: <0.05%
	Ramachandran favored	2060	95.46%	Goal: >98%
	MolProbity score [^]	2.28		98 th percentile* (N=3130, 3.00Å ± 0.25Å)
	Cβ deviations >0.25Å	17	0.86%	Goal: 0
	Bad bonds:	0 / 17203	0.00%	Goal: 0%
	Bad angles:	37 / 23248	0.16%	Goal: <0.1%
Peptide Omegas	Cis Prolines:	4 / 100	4.00%	Expected: ≤1 per chain, or ≤5%
	Twisted Peptides:	1 / 2166	0.05%	Goal: 0
Low-resolution Criteria	CaBLAM outliers	38	1.77%	Goal: <1.0%
	CA Geometry outliers	15	0.70%	Goal: <0.5%

Показатель R-фактор незначительно уменьшился с 0.2197 до 0.2143, а R-free уменьшился с 0.2656 до 0.2467 по данным с сервера PDB_redo.

В целом сервер PDB_redo не смог сильно улучшить модель, все улучшения оказались незначительными.

Заключение

В целом, структура RNase J из организма *Bacillus subtilis*, который имеет идентификатор в базе данных PDB 3zq4, расшифрована достаточно хорошо. Расшифрованный белок большой, а разрешение не очень высокое (3,0 Å), поэтому многие показатели модели оказались не очень хорошими. Однако эти значения являются средними или лучше средних для структур со сходным разрешением. Повторная оптимизация модели на сервере PDB_redo принципиальных изменений в модель не внесла, что говорит о том, что модель достаточно оптимизирована, но не переоптимизирована, потому что разница между R-фактором и R_free лежит в пределах нормы.

Список литературы

1. Newman J.A. и др. Unusual, dual endo- and exonuclease activity in the degradosome explained by crystal structure analysis of RNase J1. // *Structure*. Elsevier, 2011. Т. 19, № 9. С. 1241–1251.
2. McCoy A.J. и др. *Phaser* crystallographic software // *J. Appl. Crystallogr.* 2007. Т. 40, № 4. С. 658–674.
3. Main page - MolProbity [Электронный ресурс]. URL: <http://molprobity.biochem.duke.edu/> (дата обращения: 03.12.2017).
4. Welcome to PDB_REDO [Электронный ресурс]. URL: http://www.cmbi.ru.nl/pdb_redo/ (дата обращения: 04.12.2017).