

МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ  
имени М.В.ЛОМОНОСОВА

---

*ФАКУЛЬТЕТ БИОИНЖЕНЕРИИ И БИОИНФОРМАТИКИ*

Отчёт по качеству расшифровки панкреатической  
рибонуклеазы A *Bos taurus* (PDB ID 1A5P)

Выполнила студентка 4-го курса  
ФББ МГУ  
Швецова Екатерина Тимуровна

Москва, 2017 г.

## Оглавление

Аннотация	3
Введение	3
Результаты и обсуждение	4
1. Общая информация о модели	4
2. Значения индикаторов качества модели в целом	4
3. Маргинальные остатки	9
4. Анализ маргинальных остатков	10
Ser21	10
Lys37	11
Lys91	11
Pro93	12
Asn113	12
Сравнение модели из PDB с моделью из PDB_redo	13
Заключение о качестве расшифровки структуры	14
Список литературы и сервисов	15

## Аннотация

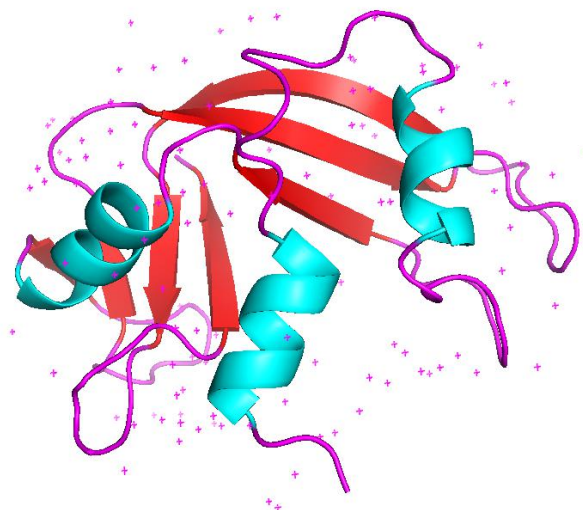
Данный отчет представляет собой анализ качества расшифровки структуры 1A5P на основе определенных показателей, таких как параметры кристаллической решетки, R-фактор, R\_free, карта Рамачандрана, и др., которые могут быть получены с помощью различных сервисов.

## Введение

Рибонуклеаза А (РНКаза А) - белок, функцией которого является расщепление РНК. РНКаза А разрезает 3'-конец неспаренных цитидиловых и уридилиловых нуклеотидов, оставляя 3'-фосфорилированный продукт в виде 2',3'-циклического монофосфата. Панкреатическая рибонуклеаза быка является одним из наиболее широко используемых в лабораторной практике ферментов.

Расшифровка структуры 1A5P была произведена в рамках изучения того, какие связи влияют на фолдинг РНКазы А. Авторы внесли в структуру 2 мутации Cys40Ala и Cys95Ala, чтобы избавиться от дисульфидной связи между остатками Cys40 и Cys95 [1]. Авторы хотели подтвердить ранее существовавшие данные ЯМР-спектроскопии о том, что удаление данной S-S связи влечет за собой появление неструктурированности в участке, где эта связь должна быть в нативном белке, но не влияет на структуру основной части белка. [2] Как и ожидалось, в полученной структуре наблюдались участки неструктурированности (в этих участках отсутствовала четкая электронная плотность) и авторы отметили их с статье: 36-40 и 90-95. [1]

На **рис. 1** можно видеть визуализацию структуры 1A5P.



**Рис. 1.** Визуализация структуры 1A5P. Белковая цепь отображена в модели cartoon, покраска по типу вторичной структуры. Крестиками отмечены молекулы воды. Изображение получено при помощи программы PyMol.

# Результаты и обсуждение

## 1. Общая информация о модели

Структура 1A5P была расшифрована авторами Pearson M.A., Karplus P.A., Dodge R.W., Laity J.H., Scheraga H.A. в 1998 году. Фазовая проблема была решена методом молекулярного замещения, за основу была взята структура 1SRN. Структура содержит белковую цепь, молекулы воды и сульфат-ион.

Рассмотрим некоторые характеристики модели в целом. Их можно узнать из файла структурных факторов, который можно найти на сайте RSCB PDB [3] либо на сервере EDS. [4]

Общее количество измеренных рефлексов - 20673, количество измеренных структурных факторов - 20448, 18906 (92,5%) из них имеют силу сигнала  $\geq 3$ , 19444 (95,1%) были использованы для оптимизации модели. Разрешение структуры довольно хорошее - 1.60 Å, при этом диапазон разрешений структурных факторов равен 22.94 - 1.60 Å.

Параметры кристаллической решетки можно найти в поле CRYST1 pdb-файла:

```
CRYST1 64.600 64.600 65.200 90.00 90.00 120.00 P 32 2 1 6
```

Эта строка обозначает следующее:

Длины направляющих векторов кристалла в Å:  $a=64.600$ ,  $b=64.600$ ,  $c=65.200$

Углы между направляющими векторами:  $\alpha=90$ ,  $\beta=90$ ,  $\gamma=120$

Кристаллографическая группа: P 32 2 1

Число молекул в ячейке кристалла: 6

Разрешены следующие операции симметрии:

X,Y,Z

-Y,X-Y,Z+2/3

-X+Y,-X,Z+1/3

Y,X,-Z

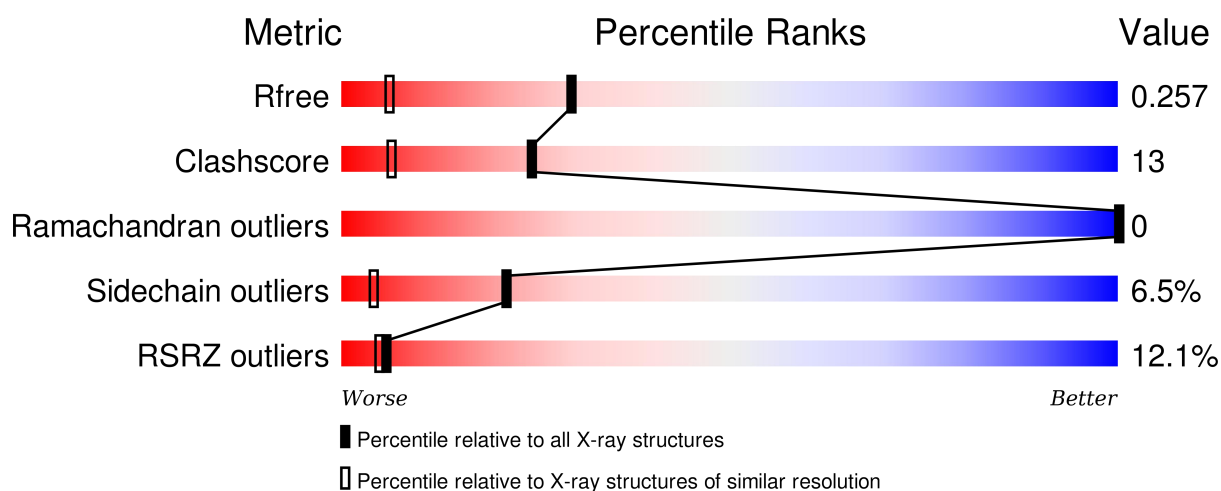
X-Y,-Y,-Z+1/3

-X,-X+Y,-Z+2/3

Некристаллической симметрии в ассиметрической ячейке кристалла не наблюдается.

## 2. Значения индикаторов качества модели в целом

Основные индикаторы качества модели приведены на сайте PDB в виде графа (см. **рис. 2**).



**Рис. 2.** Граф, содержащий основные показатели качества модели 1A5P. Белые риски – процентиль структур с таким же значением показателя, как и в 1A5P, и таким же разрешением. Черные риски – процентиль структур с таким же значением показателя, как и 1A5P по всему банку PDB.

**R-фактор** характеризует соответствие модели экспериментальным данным – структурным факторам и рассчитывается он следующим образом:

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

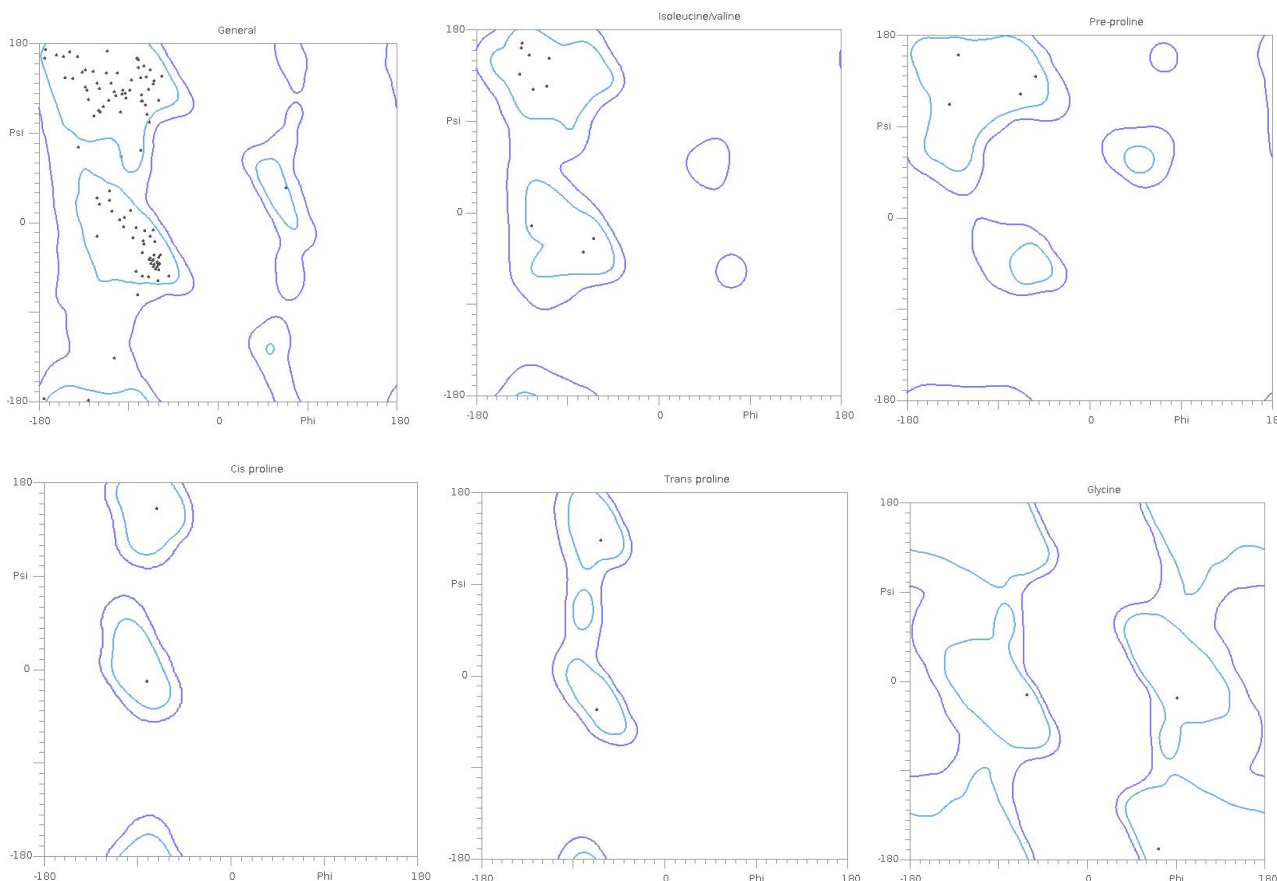
где  $F_{hkl}^{calc}$  (h,k,l) – рассчитанные модули структурных факторов,  $F_{hkl}^{obs}$  (h,k,l) – измеренные в эксперименте. При оптимизации модели R-фактор минимизируется, чем меньше R-фактор, тем модель лучше. Для данной модели он равен 0.223 (меньше 0.25, что говорит о хорошем соответствии модели и эксперимента). Более точно положение дел отражает такой показатель, как **R-free**, который рассчитывается тем же образом, но только по небольшой группе контрольных рефлексов. Его значение составляет 0.272 (то есть больше, чем R-фактор). Разница между упомянутыми выше факторами меньше 0.1, следовательно нет доводов подозревать авторов в переоптимизации данной модели.

Граф из **рис.2** показывает, что данное значение R-free хуже, чем среднее среди структур pdb в целом, а среди структур с похожим разрешением оно чуть хуже, чем среднее.

Другой набор характеристик модели можно узнать с помощью сервиса MolProbity [5].

**Clashscore** отображает число неблагоприятных стерических перекрытий атомов  $>0.4 \text{ \AA}$  на 1000 атомов и для данной структуры он равняется 13 (MolProbity характеризует такой показатель как средний - не самый плохой, но и не самый хороший вариант (среди структур со сходных разрешением), рассчитанный перцентиль - 45th). Всего таких перекрытий внутри асимметрической единицы 24 [5] [6], ни одно из них не связано с преобразованиями симметрии.

Далее, на **рис.3** представлены **карты Рамачандрана** для всех аминокислотных остатков, а также отдельно для некоторых из них. 95.08% (116 из 124, в идеале должно быть >98%) расположены в благоприятных областях, 100.0% (124 из 124) – в разрешенных. Выбросов нет. Число остатков с маргинальными по отклонению от ротамеров боковыми цепями (roo rotamers) равно 9 (8.41%, в идеале <0.3%); число и процент ротамеров в предпочитаемой области - 84.11% (в идеале >98%). Видно, что по этим показателям модель проигрывает.



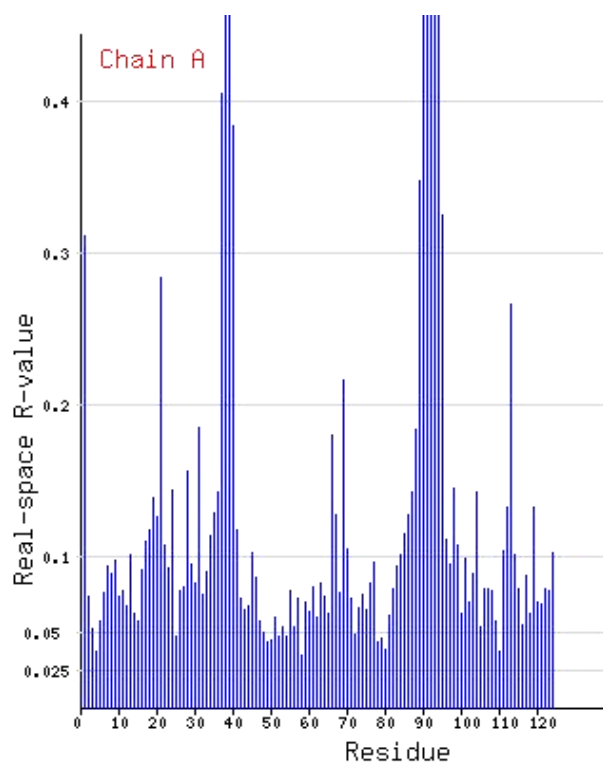
**Рис. 3.** Карты Рамачандрана для 1) основной массы остатков 2) изолейцинов/лейцинов 3) пре-пролинов 4) цис-пролинов 5) транс-пролинов 6) глицинов структуры 1A5P.

**C  $\beta$  deviations** > 0.25Å - число C  $\beta$  атомов с неприемлемым отклонением от ожидаемого положения – равняется 0 (соответствует идеально ситуации), **число плохих валентных углов**, существенно отклоняющихся от теории, равно 4 (это составляет 0.30%, а в идеале должно быть <0.1%), а **число плохих валентных связей** равняется 0. В отчете на сайте PDB представлена 1 валентная связь, имеющая длину, сильно отклоняющуюся от теории, и не указано ни одного некорректного валентного угла. [6]

**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. 4**, полученном с помощью сервера EDS [4], приведен RSR для остатков белка.



**Рис. 4.** Значения RSR для остатков структуры 1A5P

“Идеальные” значения RSR должны быть <0.1%. Остатки, имеющие RSR>0.3% в большой уверенностью можно считать маргинальными по этому параметру. На рис. 4 видно, что таких остатков в структуре 12. Например, Lys37 (0,405%), Asp (0,546%), Ser90 (0,506%), Tyr92 (0,781%). Видно, что самые большие значения RSR наблюдаются в районах 37-40 и 89-95, которые совпадают с участками неструктурированности, охарактерезованными авторами структуры [1], но и некоторые остатки вне этих областей также имеют достаточно большой RSR: Lys1 (0,311%), Ser21 (0,284%), Asn113 (0,266%).

К таким же результатам можно придти, если анализировать **significant regions**, где отображаются наиболее «рискованные» остатки (с  $RSR > (\langle RSR \rangle + 2.00 \cdot \sigma)$ ) - см. рис. 5. EDS детектировал 15 выбросов по этому параметру.

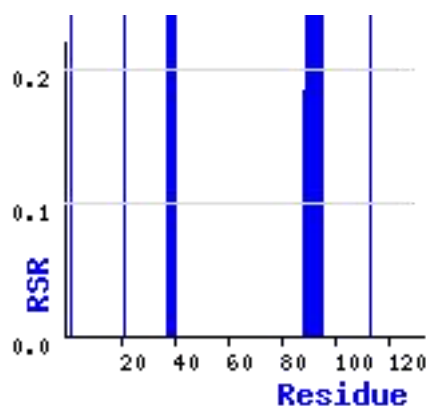


Рис. 5. Significant regions структуры 1A5P

**RSR-Z** – это относительная оценка RSR для отдельных остатков. Для вычисления Z-score остатка (напр. Lys37) его RSR сравнивается со средним RSR для того же типа остатков (Lys) по выборке из PDB с примерно таким же разрешением (в нашем случае 1.40-1.60Å). Z-score вычисляется по формуле  $Z\text{-score} = (RSR - \langle RSR \rangle) / \sigma$ . Высокие положительные значения  $Z > 2$  остатка свидетельствуют о том, что он плохо вписан в электронную плотность по сравнению с другими структурами с тем же разрешением, поэтому его можно рассматривать как маргинальный. RSR-Z для остатков нашей структуры можно увидеть на **рис. 6**.

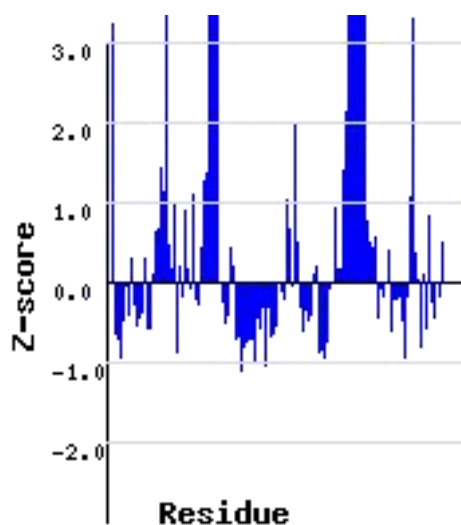


Рис. 6. Значения RSR-Z для остатков структуры 1A5P

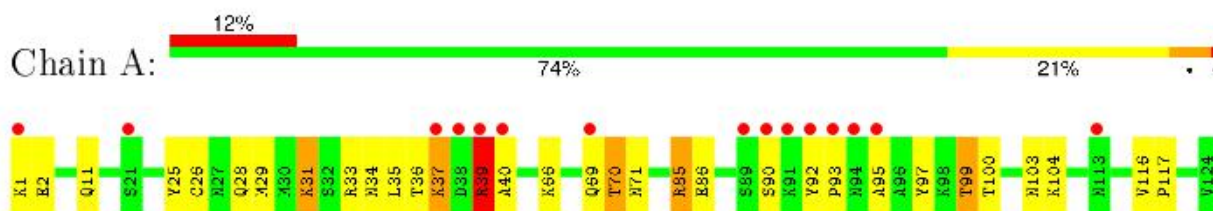
По **рис. 6** видно, что выбросы по RSR-Z (с  $RSR-Z > 2$ ) согласуются с картиной, рассмотренной ранее для выбросов по RSR.

Затем с помощью сайта WHAT\_CHECK [7] опцией Fine Packing Quality Control была оценена **комфортность окружения** для всех остатков. Не найдено ни одного остатка с параметром меньше -5, что соответствовало бы маргиналам по окружению. В основном все значения  $> -2$ . Среднее значение  $Z\text{-score} = -0.27$ , что соответствует хорошей структуре. Все оценочные суждения основаны на инструкции в конце выдачи программы.



### 3. Маргинальные остатки

Для определения всех маргинальных остатков обратимся к подробному отчету о валидации структуры 1A5P [6]. Имеющаяся в нем схема (см **рис.7**) показывает, 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. Также на схеме отмечены остатки с  $RSR-Z > 2$  (красными точками).



**Рис. 7.** Карта маргинальности остатков структуры 1A5P

Видно, что маргинальных остатков в структуре много. В **таблице 1** представлены некоторые примеры маргинальных остатков структуры с указанием показателей их маргинальности.

**Таблица 1.** Примеры маргинальных остатков структуры 1A5P

№ маргинала	Аминокислотный остаток	Причина маргинальности
1	Ser21	Плохо вписан в электронную плотность (outlier по $RSRZ=4.13$ , указан как outlier в significant regions, $RSR=0.284$ )
2	Lys31	Перекрытие (clash) с Lys37 ( $0.42\text{\AA}$ ), маргинал по торсионным углам в боковой цепи остатка (poor rotamer)
3	Lys37	Перекрытие с Asn34 ( $0.54\text{\AA}$ ) и Lys31 ( $0.42\text{\AA}$ ) а также перекрытия внутри самого остатка и с молекулами воды, маргинал по торсионным углам в боковой цепи остатка (poor rotamer), плохо вписан в электронную плотность ( $RSRZ=4.6$ , $RSR=0.405$ , указан в significant regions)
4	Arg39	Маргинал по длине валентной связи между атомами CZ-NH1, маргинал по торсионным углам в боковой цепи остатка (poor rotamer), плохо вписан в электронную плотность ( $RSRZ=6.4$ , $RSR=0.471$ , указан в significant regions)

5	Thr70	Перекрывание атомов O и HG23 (0.52Å), маргинал по торсионным углам в боковой цепи остатка (poor rotamer)
6	Ile81	Маргинал по торсионным углам в боковой цепи остатка (poor rotamer)
7	Lys91	Плохо вписан в электронную плотность (RSRZ=9.2, RSR=0.791, указан в significant regions)
8	Pro93	Плохо вписан в электронную плотность (RSRZ=14.5, RSR=0.836, указан в significant regions), перекрытие с Tyr92 (0.45Å)
9	Thr99	Перекрывание с Cys26 (0.73Å), маргинал по торсионным углам в боковой цепи остатка (poor rotamer)
10	Asn113	Плохо вписан в электронную плотность (RSRZ=3.3, RSR=0.266, указан в significant regions)

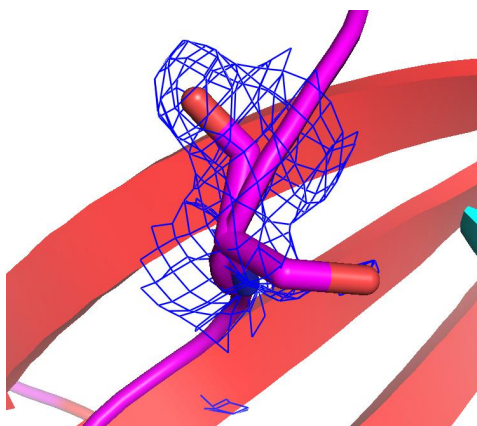
#### 4. Анализ маргинальных остатков

Рассмотрим некоторые маргинальные остатки из **таблицы 1** подробнее.

##### Ser21 (рис. 8)

Маргинальность: плохо вписан в электронную плотность (outlier по RSRZ=4.13, указан как outlier в significant regions, RSR=0.284).

По рисунку видно, что боковая цепь серина не покрыта электронной плотностью, что связано с тем фактом, что в данном участке функция электронной плотности расшифрована плохо. Это согласуется с показателями маргинальности данного остатка.

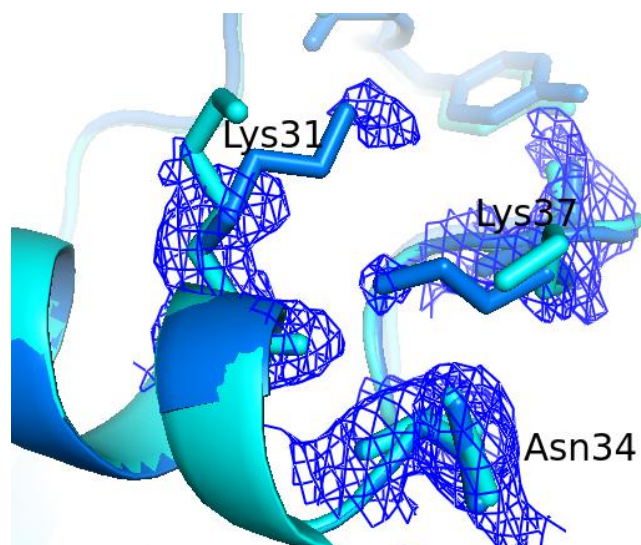


**Рис. 8.** Визуализация электронной плотности вокруг Ser21 структуры 1A5P ( $\sigma=1$ ).

## Lys37 (рис. 9)

Маргинальность: перекрытие с Asn34 (0.54Å) и Lys31 (0.42Å) а также перекрытия внутри самого остатка и с молекулами воды, маргинал по торсионным углам в боковой цепи остатка (poor rotamer), плохо вписан в электронную плотность (RSRZ=4.6, RSR=0.405, указан в significant regions).

На **рисунке 9** показан Lys37, Lys31 и Asn34 (которые находится близко к Lys37, что послужило причиной для включения этих остатков в clashcore). Также для сравнения на рисунке приведена структура, оптимизированная PDB\_redo. Видно, что после оптимизации остатки лизинов приняли более благоприятные положения и перекрытие было ликвидировано. Однако, модель PDB\_redo никак не учитывает небольшие сгустки электронной плотности, в которые вписаны концевые атомы боковых цепей основной модели.

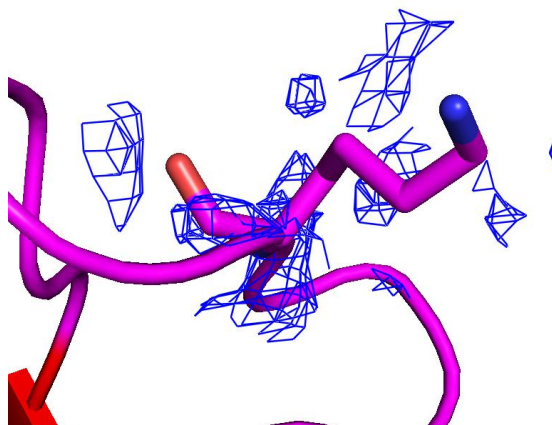


**Рис. 9.** Визуализация электронной плотности вокруг Lys37, Asn34, Lys31 структуры 1A5P ( $\sigma=0.5$ ). Синим показана оригинальная структура 1A5P, бирюзовым - структура, восстановленная PDB\_redo.

## Lys91 (рис. 10)

Маргинальность: плохо вписан в электронную плотность (RSRZ=9.2, RSR=0.791, указан в significant regions)

Электронная плотность в данном участке практически полностью отсутствует (увидеть что-либо можно только на очень низком уровне позрезки). Вероятно, это связано с тем, что данный остаток входит в состав неструктурированного участка, отмеченного авторами структуры.

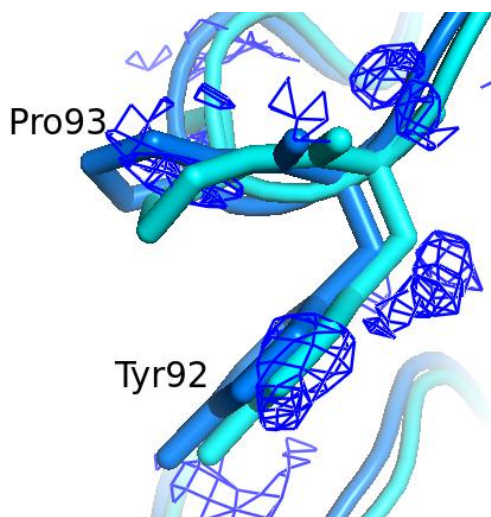


**Рис. 10.** Визуализация электронной плотности вокруг Lys91 структуры 1A5P ( $\sigma=0.5$ ).

### Pro93 (рис. 11)

Маргинальность: плохо вписан в электронную плотность (RSRZ=14.5, RSR=0.836, указан в significant regions), перекрытие с Tyr92 (0.45Å)

На **рисунке 11** показан Pro93 и Tyr92 (который находится близко к пролину, это послужило причиной для включения этих остатков в clashcore). Также для сравнения на рисунке приведена структура, оптимизированная PDB\_redo. Видно, что расположение остатков практически не поменялось после оптимизации. В данном участке почти нет сгустков электронной плотности, поэтому сложно хорошо вписать в неё аминокислотные остатки и PDB\_redo не особенно улучшает ситуацию.

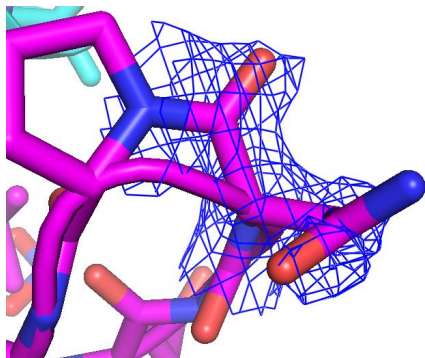


**Рис. 11.** Визуализация электронной плотности вокруг Pro93, Tyr92 структуры 1A5P ( $\sigma=0.5$ ). Синим показана оригинальная структура 1A5P, бирюзовым - структура, восстановленная PDB\_redo.

### Asn113 (рис. 12)

Маргинальность: плохо вписан в электронную плотность (RSRZ=3.3, RSR=0.266, указан в significant regions)

Как и в случае Ser21, боковая цепь Asn113 не полностью покрыта электронной плотностью, что связано с тем фактом, что в данном участке функция электронной плотности расширена плохо. Это согласуется с показателями маргинальности данного остатка.



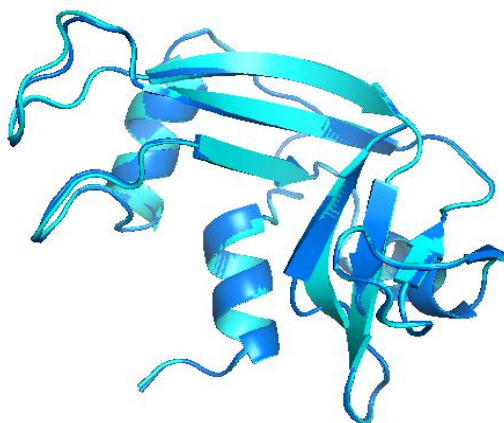
**Рис. 12.** Визуализация электронной плотности вокруг Asn113 структуры 1A5P ( $\sigma=1$ ).

Все приведенные выше остатки можно считать маргиналами и основной причиной их маргинальности можно считать недостатки функции электронной плотности в районе этих остатков.

## Сравнение модели из PDB с моделью из PDB\_redo

PDB\_redo позволяет оптимизировать модель, используя различные параметры уточнения, перестраивая конформации ротамеров, вращая боковые цепи остатков для оптимизации сети водородных связей, пересматривая молекулы воды и конформации лигандов.

Некоторые примеры изменений конкретных остатков в результате оптимизации приводились в предыдущем разделе. Рассмотрим глобальное изменение конформаций белковой цепи (см. **рис. 13**).



**Рис.13.** Сравнение моделей PDB и PDB\_redo. PDB – синяя, PDB\_redo – бирюзовая.

Как видно из **рисунка 13**, не наблюдается каких-либо глобальных изменений в структуре остова модели, есть лишь небольшие отличия в положениях некоторых петель.

Однако, были улучшены некоторые показатели: конформация скелета, вид карты Рамачандрана (меньше остатков в сравнительно неблагоприятных областях), оптимизированы длины и углы связей, уменьшено число столкновений (bumps/clashes) атомов. Практически неизменным осталось количество водородных связей (немного увеличилось).

PDB\_redo улучшил показатели для 74 остатков структуры и ни для каких остатков не ухудшил показателей.

Также улучшились значения R и R\_free (R до = 0.223, R после = 0.1806; R\_free до = 0.272, R\_free после = 0.2048).[8]

## **Заключение о качестве расшифровки структуры**

Исходя из вышеприведённой информации, можно сделать вывод, что, не смотря на высокое разрешение (1.6 ангстрем), качество расшифровки структуры панкреатической рибонуклеазы A *Bos taurus* (PDB ID 1A5P) не самое лучшее и вероятно это в основном связано с тем, что в данном мутанте белка присутствуют неструктурированные участки, что затрудняет расшифровку функции электронной плотности в этих участках. Основные параметры качества структуры хуже средних показателей как среди структур со сходным разрешением, так и по всему банку PDB, маргинальных остатков довольно много.

Повторная оптимизация модели (PDB\_redo) улучшила ряд параметров и показатели для большинства остатков структуры, но не внесла глобальных изменений в структуру. Это может быть сигналом того, что помимо недостатков в самой электронной плотности, полученной в эксперименте, оптимизация модели также произведена не наилучшим образом.

В целом, расшифровка данной структуры были сделана не с целью получить идеальную структуру РНКазы, а с целью изучения неструктурированности, которая появляется в белке в результате мутации, поэтому, можно заключить, что качество расшифровки данной структуры достаточно для достижения целей авторов, однако её лучше не использовать для более общих целей. Этот пример лишний раз напоминает нам о том, что при выборе структуры для исследования нужно обращать внимание не только на разрешение, но и на другие показатели качества расшифровки.

## Список литературы и сервисов

- [1] - Crystal structures of two mutants that have implications for the folding of bovine pancreatic ribonuclease A. Pearson, M.A., Karplus, P.A., Dodge, R.W., Laity, J.H., Scheraga, H.A.(1998) Protein Sci. 7: 1255-1258
- [2]The refined crystal structure of a fully active semisynthetic ribonuclease at 1.8-Å resolution. Martin, P.D., Doscher, M.S., Edwards, B.F. (1987) J.Biol.Chem. 262: 15930-15938
- [3] - «RCSB PDB,» [В Интернете]. Available: <http://www.rcsb.org/pdb/explore/explore.do?structureId=1a5p>
- [4] - «EDS,» [В Интернете]. Available: <http://eds.bmc.uu.se/eds/>
- [5] - «MolProbity,» [В Интернете]. Available: <http://molprobity.biochem.duke.edu>
- [6] - «wwPDB Validation report на сайте PDB со структурой 1A5P,» [В Интернете]. Available: <http://www.rcsb.org/pdb/ngl/ngl.do?pdbid=1A5P&preset=validationReport>
- [7] - «WHAT\_CHECK,» [В Интернете]. Available: <http://swift.cmbi.ru.nl/servers/html/index.html>
- [8] - «Выдача pdb\_redo,» [В Интернете]. Available: [http://www.cmbi.ru.nl/pdb\\_redo/a5/1a5p/index.html](http://www.cmbi.ru.nl/pdb_redo/a5/1a5p/index.html)
- [9] - PyMol