

МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ  
имени М.В. ЛОМОНОСОВА

ФАКУЛЬТЕТ БИОИНЖЕНЕРИИ И БИОИНФОРМАТИКИ

**Отчет о качестве расшифровки структуры белка 193I  
методом рентгеноструктурного анализа**

Работу выполнила студентка 4 курса  
Софронова Алина Андреевна

Москва, 2017

## **Аннотация**

В данной работе произведена оценка индикаторов качества структуры белка 193I, расшифрованной методом рентгеноструктурного анализа, и проанализированы некоторые маргинальные остатки структуры. Произведено сравнение структуры с моделью из PDB\_REDO.

## 1. Введение

Структура 193I, расшифрованная в 1996 году М. С. Vaneу, S. Maignan, M. Ries-Kautt, A. Ducruix, принадлежит лизоциму из яичного белка *Gallus gallus*. Лизоцим (ЕС 3.2.1.17) обладает бактериолитической активностью. Лизоцим катализирует гидролиз В-1,4 гликозидной связи между N-ацетилмурамовой кислотой и N-ацетилглюкозамином в пептидогликане клеточной стенки бактерий<sup>1</sup>.

Ранее структура лизоцима уже была расшифрована<sup>2-4</sup>. Особенностью именно этой структуры было использование нового метода кристаллизации (Advanced Protein Crystallization Facility) в обычных условиях и в условиях микрогравитации. Было проведено сравнение структур при указанных условиях<sup>5</sup>. Полученные структуры имеют высокое разрешение (особенно для 1996 года): 1,33Å – в условиях микрогравитации (та, которая представлена в базе данных PDB<sup>6</sup>), и 1,4 Å – в других<sup>5</sup>. Полученная структура позволила исследовать конформацию петли Ser60-Leu75. Кроме того, сравнение структур показало, что метод кристаллизации с пониженной гравитацией имеет преимущества над методом кристаллизации в обычных «земных» условиях.

## **2. Результаты и обсуждение**

### **Общая информация о модели**

Лизоцим был закристаллизован при pH 4,3. Структура содержит 1001 атом, принадлежащий белку, 2 гетероатома ( $\text{Na}^+$  и  $\text{Cl}^-$ ) и 142 атома воды.

Кристалл принадлежит к тетрагональной сингонии ( $a=78.54\text{\AA}$ ,  $b=78.54\text{\AA}$ ,  $c=37.77\text{\AA}$ ,  $\alpha=90.00$ ,  $\beta=90.00$ ,  $\gamma=90.00$ ). Структура 193I имеет примитивную кристаллическую решетку и относится к пространственной группе  $4_22_12$ . Биологическая единица совпадает с асимметрической единицей и является мономером. Некристаллографических симметрий асимметрической единицы не наблюдается.

Нет данных о методе решения фазовой проблемы.

Минимальное разрешение структуры 193I равно  $1,33\text{\AA}$ . Максимальное разрешение структуры 193I равно  $15\text{\AA}$ . Полнота набора рефлексов, использованных для расшифровки структуры, равна 87,1%.

### **Индикаторы качества**

#### ***R-value и R-free***

В рентгеновском эксперименте было измерено 24111. Значения R-free и R-value равны 22.6% и 18.4% соответственно. Полученные значения R-value и R-free отличаются друг от друга незначительно (на 4.2%), что позволяет исключить вероятность перепогонки структуры, однако данные значения R-value и R-free свидетельствуют о том, что расшифровка структуры находится на грани среднего и низкого качества.

## Сводная статистика сервиса MolProbity

С помощью сервиса MolProbity к структуре 193I были добавлены атомы водорода, затем были проанализированы контакты атомов и геометрия структуры. Сводная статистика анализа структуры приведена в Таблице 1.

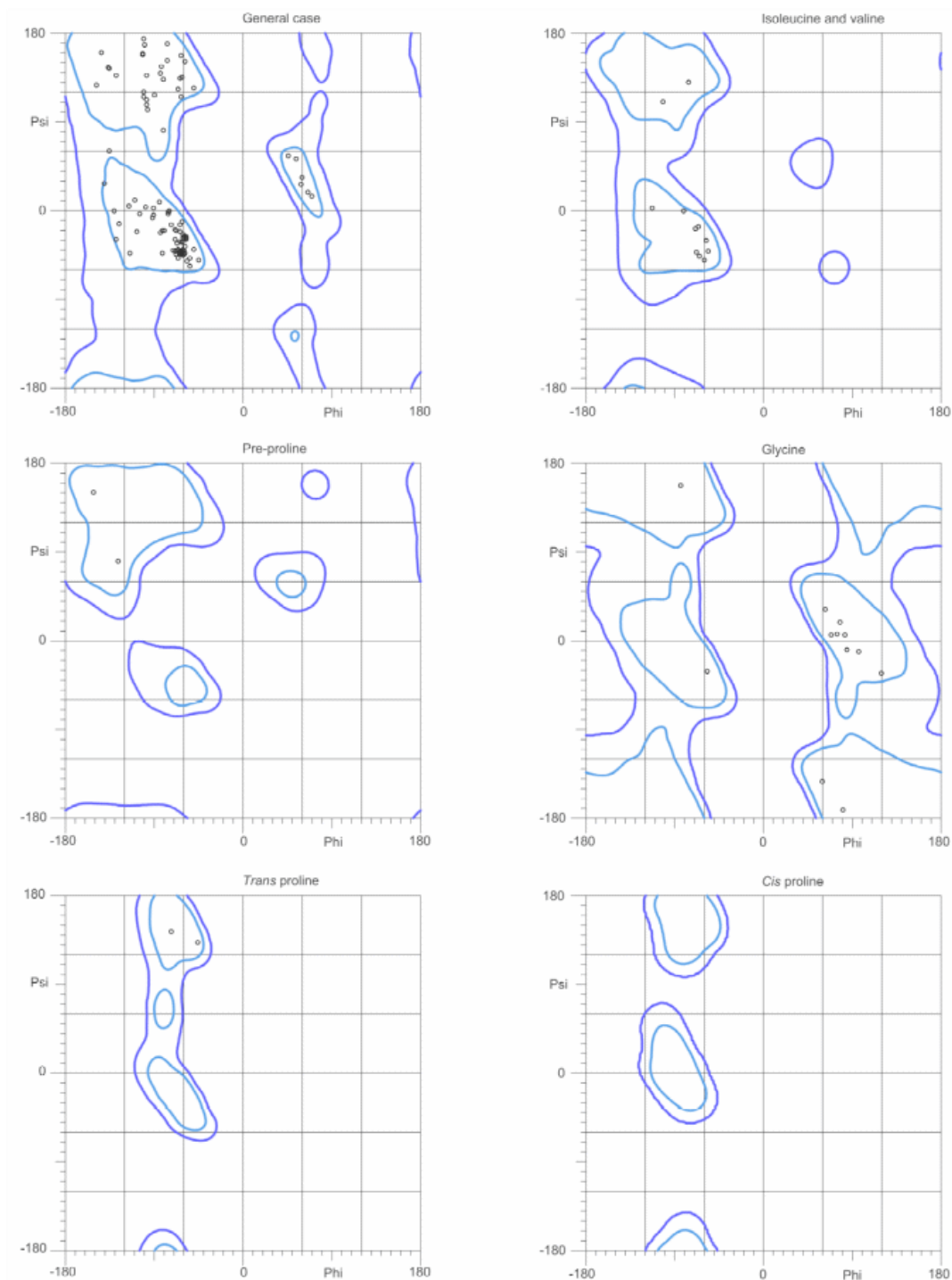
Таким образом, Clashscore структуры 193I равен 6,02 на 1000 атомов структуры, т.е. лишь 0.602% атомов структуры имеют недопустимые наложения (т.е. их Ван-дер-Ваальсовы радиусы перекрываются больше, чем на 0.4 Å). Структура 193I имеет 80<sup>ю</sup> перцентиль, т.е. 80% структур подобного разрешения имеют больший Clashscore, из чего следует, что относительно мало атомов структуры 193I имеют недопустимые наложения. Структура 193I содержит 2 (1,83%) остатка (Arg14, Leu129), конформация боковых цепей которых существенно отличается от ротамеров (в идеале < 0,3% остатков), 2 остатка принадлежат допустимой конформации (Asn77, Arg125) и 105 остатков (96,33%) находятся в предпочитаемых конформациях (в идеале > 98% остатков). Структура 193I не содержит остатков, находящихся в запрещенной области на карте Рамачандрана, 2 остатка, находящихся в допустимой области (Arg68, Asn74) и 125 (98.43%) остатков, находящихся в предпочитаемой области на карте.

**Таблица 1.** Сводная статистика анализа структуры 193I с помощью сервиса MolProbity (добавлены атомы водорода). Красным помечены отрицательные параметры модели, желтым - средние, зеленым – хорошие.

All-Atom Contacts	Clashscore, all atoms:	6.02		80 <sup>th</sup> percentile* (N=382, 1.33Å ± 0.25Å)
	Clashscore is the number of serious steric overlaps (> 0.4 Å) per 1000 atoms.			
Protein Geometry	Poor rotamers	2	1.83%	Goal: <0.3%
	Favored rotamers	105	96.33%	Goal: >98%
	Ramachandran outliers	0	0.00%	Goal: <0.05%
	Ramachandran favored	125	98.43%	Goal: >98%
	MolProbity score <sup>^</sup>	1.53		76 <sup>th</sup> percentile* (N=2319, 1.33Å ± 0.25Å)
	Cβ deviations >0.25Å	0	0.00%	Goal: 0
	Bad bonds:	0 / 1036	0.00%	Goal: 0%
	Bad angles:	0 / 1404	0.00%	Goal: <0.1%
Peptide Omegas	Cis Prolines:	0 / 2	0.00%	Expected: ≤1 per chain, or ≤5%

## Карта Рамачандрана

С помощью сервиса MolProbity была построена карта Рамачандрана структуры 193I (предварительно к структуре были добавлены атомы водорода). Построенная карта Рамачандрана приведена на Рисунке 1.



**Рисунок 1.** Карта Рамачандрана структуры 193I (добавлены атомы водорода). Представлены карты для остатков в общем случае, для валина и изолейцина, для аминокислоты перед пролином, для глицина, для транс и цис пролина.

Структура 193I не содержит остатков, находящихся в запрещенной области на карте Рамачандрана. Только 2 остатка находятся в разрешенной зоне (Arg68, Asn74). Остальные остатки находятся в предпочитаемой области.

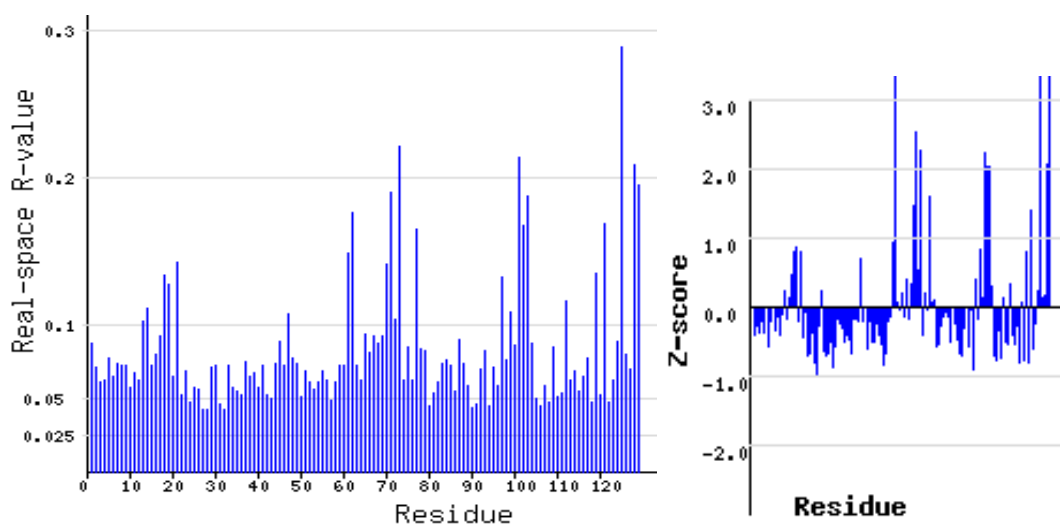
### **RSR-value u RSR Z-score**

Real Space R (RSR, пространственный R-фактор) структуры равен 8,3%, что говорит о хорошем соответствии «экспериментальной» электронной плотности и электронной плотности модели.

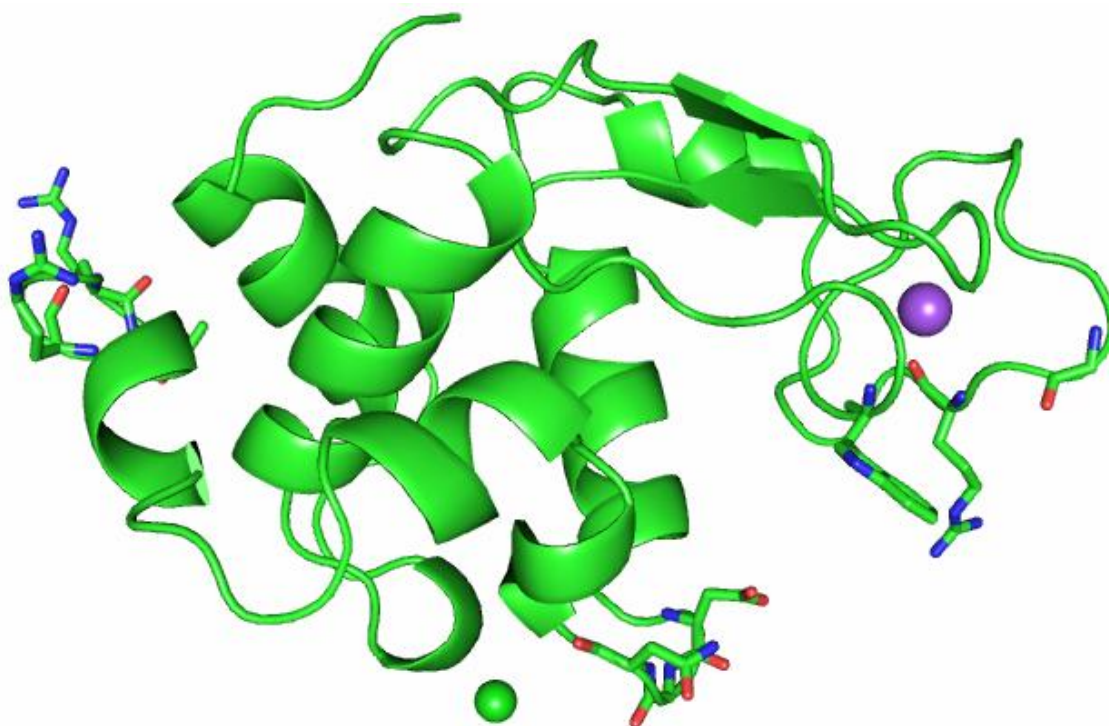
С помощью сервера EDS в структуре 193I было проанализировано качество восстановления электронной плотности. На рисунке 2 приведены RSR-value и RSR Z-score ( $Z\text{-score} = (RSR - \langle RSR \rangle) / \sigma$ ) для каждого остатка структуры 193I.

RSR-value позволяет оценить соответствие восстановленной по модели электронной плотности и «экспериментальной» электронной плотности. Чем больше RSR-value, тем хуже расшифрованы координаты атомов в модели. RSR Z-score позволяет сравнить RSR-value остатка со средним RSR-value остатков того же типа в структурах подобного разрешения. Чем больше RSR Z-score, тем хуже расшифрованы координаты атомов в модели по сравнению с координатами атомов остатков того же типа в структурах подобного разрешения.

Маргиналами (остатки, которые плохо вписаны в электронную плотность) будем считать те остатки, которые имеют  $RSR > 20\%$  или  $Z\text{-score} > 2$ . Среди них Trp62, Gly71, Arg73, Asp101, Gly102, Asn103, Arg125, Agr128, Leu129. Приведенные остатки располагаются в неструктурированных петлях на поверхности глобулы лизоцима (Рисунок 3).



**Рисунок 2.** RSR-value и RSR Z-score для каждого остатка структуры 193I.



**Рисунок 3.** Структура 193I. Белок представлен в виде «cartoon», маргинальные остатки (см. в тексте) в проволочной («stick») модели, сферами показаны атомы  $\text{Na}^+$  и  $\text{Cl}^-$ .

Сводная информация об RSR Z-score для структуры 193I представлена в Таблице 2. Таким образом, структура содержит 6.98% остатков, которые имеют RSR Z-score > 2.

**Таблица 2.** Сводная информация о Z-score для структуры 193I.

Chain	Nres	<Z>	Sigma	%<-1	%<0	%>1	%>2
A	129	0.01	0.90	0.00	66.67	9.30	6.98

### Маргинальные остатки

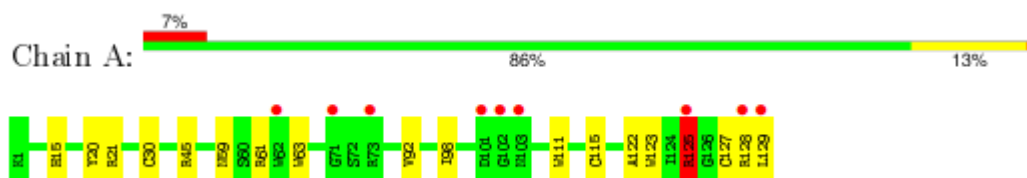
При поиске маргинальных остатков в структуре 193I была проанализирована выдача сервиса MolProbity, выдача сервера EDS, а также полный отчет о валидации структуры 193I, представленный на сайте PDB. Некоторые найденные маргинальные остатки приведены в Таблице 3, с указанием критерия маргинальности.

В первую очередь в Таблицу 3 были внесены остатки, чьи боковые цепи находятся в запрещенной или допустимой конформации (по данным сервиса MolProbity - Arg14, Leu129, Asn77, Arg125). Затем в Таблицу 4 были внесены остатки,



находящие в допустимой области на карте Рамачандрана (Arg68, Asn74). Кроме того, на основе RSR и RSR Z-score (RSR > 20%, Z-score > 2) были внесены следующие остатки Trp62, Gly71, Arg73, Asp101, Gly102, Asn103, Arg125, Agr128, Leu129.

На Рисунке 4 представлена карта маргинальности остатков в структуре 193I. Данная карта позволяет определить маргинальность остатка по совокупности параметров (отклонение геометрии, соответствие электронной плотности). Остатки, выделенные красным, имеют отклонения по трем критериям, желтые по одному, зеленый цвет означает отсутствие маргинальности по всем критериям. Красные точки означают, что соответствующие остатки имеют RSR Z-score > 2.



**Рисунок 4.** Карта маргинальности остатков структуры 193I.

Кроме того, согласно отчету по валидации нет остатков с неправильной длиной связи, углами. Все остатки необходимой хиральности. Выявлено отклонение от плоскости в гуанидиновой группе Arg125. Для уменьшения перекрытия атомов и образования водородной связи следуют перевернуть Asn44.

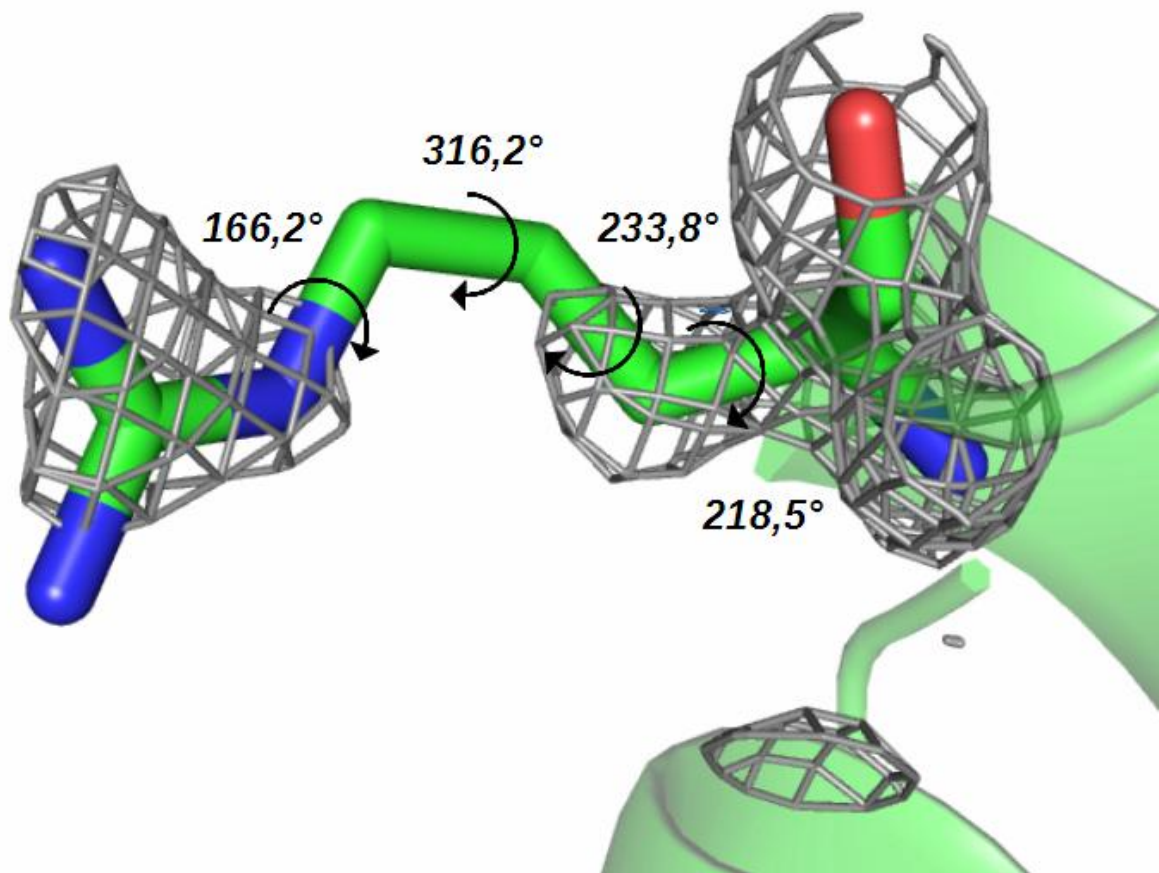
На основе сказанного выше, были выбраны следующие 5 остатков (в Таблице 4 подсвечены розовым): Arg14, Leu129, Arg125, Arg128 и Asn44.

**Таблица 3.** Некоторые маргинальные остатки в структуре 193I. Розовым фоном выделены остатки для дальнейшего рассмотрения.

Маргинал	Критерий
Arg14	Имеет боковую цепь в запрещенной конформации
Leu129	Имеет боковую цепь в запрещенной конформации, Z-score>2, перекрытие CD2 атома с HB2 атомом Cys127
Asn77	Имеет боковую цепь в допустимой конформации
Arg125	Имеет боковую цепь в допустимой конформации, Z-score>2, перекрытие атома NH11 с атомами HG3 и CG внутри аргинина, и

	атома HD3 с O в Ala122, отклонение от плоскости гуанидиновой группы
<b>Arg68</b>	Лежит в допустимой области на карте Рамачандрана
<b>Asn74</b>	Лежит в допустимой области на карте Рамачандрана
<b>Trp62</b>	Z-score>2
<b>Gly71</b>	Z-score>2
<b>Arg73</b>	Z-score>2
<b>Asp101</b>	Z-score>2
<b>Gly102</b>	Z-score>2
<b>Asn103</b>	Z-score>2
<b>Agr128</b>	Z-score>2, перекрывание атомов HD3 и O в HOH248
<b>Asn44</b>	Нуждается в инверсии боковой цепи

### Arg14



**Рисунок 5.** 14 остаток аргинина. Электронная плотность с подрезкой 1,5σ

Аргинин 14 является маргинальным остатком, так как его боковая цепь находится в запрещенной области. Как показано на Рисунке 4 торсионные углы его боковой цепи равны (с точностью до 360) 218.5°, 233.8°, 316.2°, 166.2°. Ниже (Таблица 4) представлена таблица ротамеров аргинина <sup>7</sup>. Видно, что 14 остаток аргинина действительно находится не в комфортной для него конформации. Электронная плотность получена неточно (кое-где отсутствует полностью), из-за чего не удалось правильно вписать этот остаток аргинина.

**Таблица 5. Таблица ротамеров аргинина**

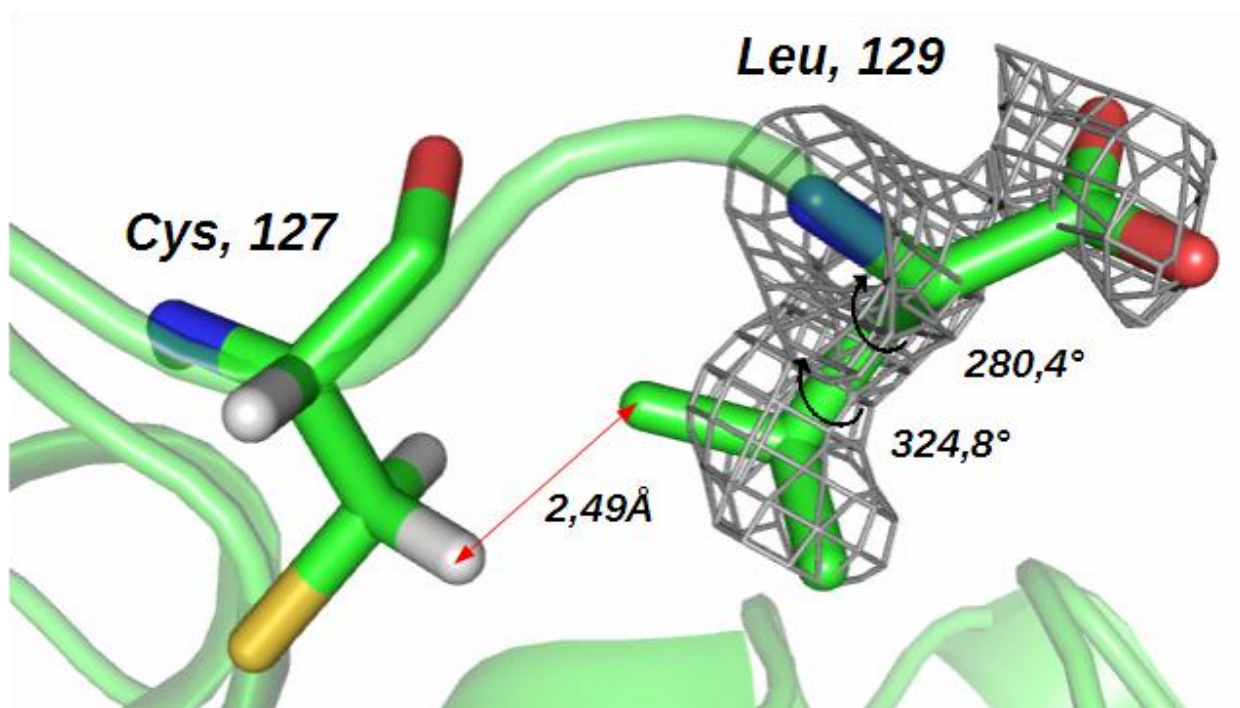
Chi(1)	Chi(2)	Chi(3)	Chi(4)
62	180	65	85
62	180	65	-175
62	180	180	85
62	180	180	180
62	180	180	-85
62	180	-65	175
62	180	-65	-85
-177	65	65	85
-177	65	65	-175
-177	65	180	85
-177	65	180	180
-177	180	65	85
-177	180	65	-175
-177	180	65	-105
-177	180	180	85
-177	180	180	180
-177	180	180	-85
-177	180	-65	105
-177	180	-65	175
-177	180	-65	-85
-67	180	65	85
-67	180	65	-175
-67	180	65	-105
-67	180	180	85
-67	180	180	180
-67	180	180	-85
-67	180	-65	105
-67	180	-65	175
-67	-167	-65	-85
-62	-68	180	85
-62	-68	180	180
-62	-68	180	-85
-62	-68	-65	175
-62	-68	-65	-85

### Leu129

129 остаток лейцина является маргинальным сразу по нескольким параметрам. Во-первых, конформация его боковой цепи не соответствует не одному ротимеру (Таблица 5). Кроме того, RSR Z-score = 4, значит остаток плохо вписан в электронную плотность. Более того, перекрывание ван-дер-ваальсовых радиусов CD2 атома Leu129 с HB2 атомом Cys127 составляет 0.42Å, что должно приводить к сильному отталкиванию этих атомов (Рисунок 6).

**Таблица 5. Таблица ротамеров лейцина**

Chi(1)	Chi(2)
62	80
-177	65
-172	145
-85	65
-65	175

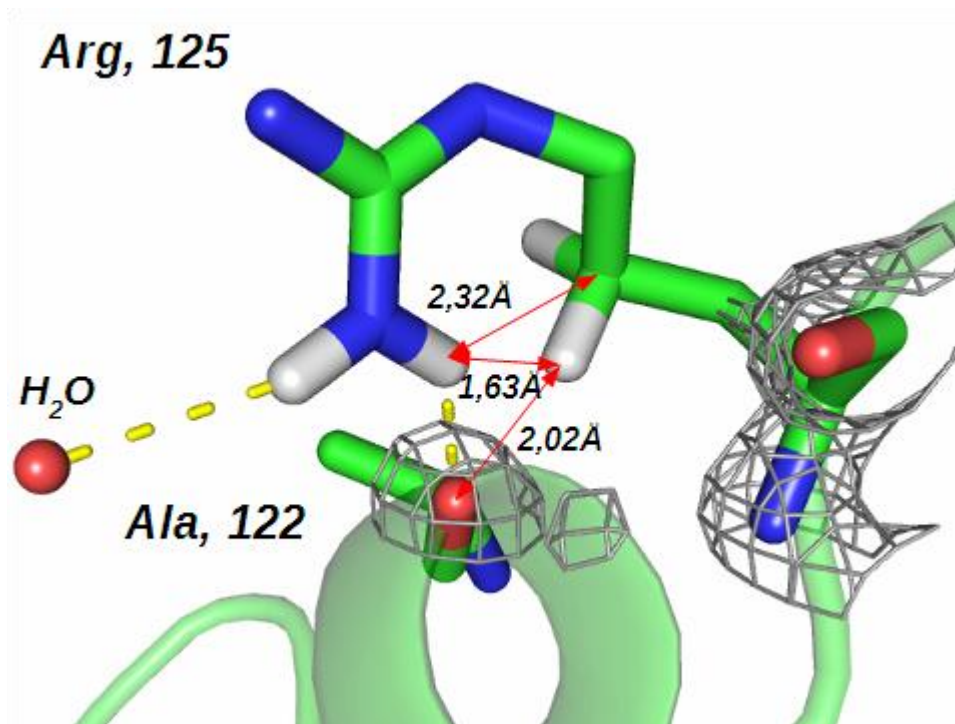


**Рисунок 6.** Остаток 129 лейцина. Электронная плотность с подрезкой  $1\sigma$

На Рисунке 6 видно, что остаток лейцина действительно плохо вписан в электронную плотность. Из-за этого неточно определено положение остатка, а значит и его углы. При небольшом сдвиге лейцина в модели (на  $0,02\text{\AA}$ ) пропадет проблема перекрывания ван-дер-ваальсовых радиусов.

### Arg125

Аргинин является самым «проблемным» остатком в структуре (Рисунок 7). Во-первых, он имеет боковую цепь в допустимой конформации. Кроме того, его Z-score=3,7. Атомы пересекаются своими ван-дер-ваальсовыми радиусами как внутри аргинина (атом NH11 с атомами HG3 и CG на 0,65 и 0,4 $\text{\AA}$  соответственно), так и с 122 остатков аланина (атом HG3 аргинина с атомом O аланина).

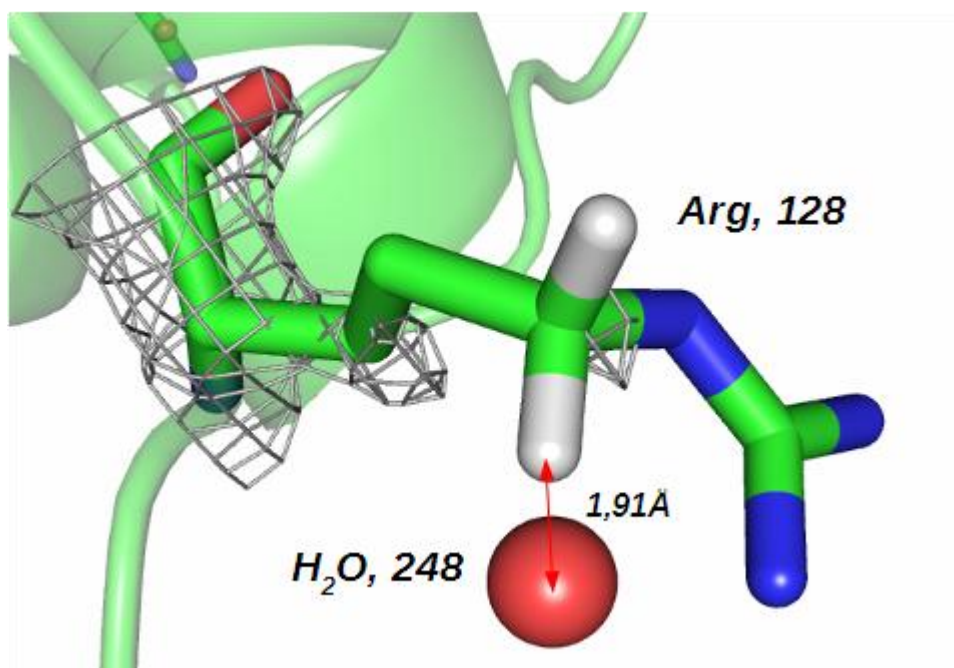


*Рисунок 7. 125 остаток аргинина. Электронная плотность с подрезкой 1,5σ*

Для 125 аргинина полностью отсутствует электронная плотность боковой цепи, из-за этого неправильно расшифровано пространственное положение остатка. Вероятно, допустимая конформация боковой цепи и столкновение атомов связано с образованием водородной связи между аргинином и аланином. Но более вероятно, что данный остаток плохо оптимизирован.

### Arg 128

RSR Z-score данного остатка 2,1, что говорит лишь о небольшом отклонении от электронной плотности (Рисунок 8). Зато очень большое перекрытие ванн-дер-ваальсовых радиусов атома HD3 аргинина и атома O в НОН248 (0,7 Å).

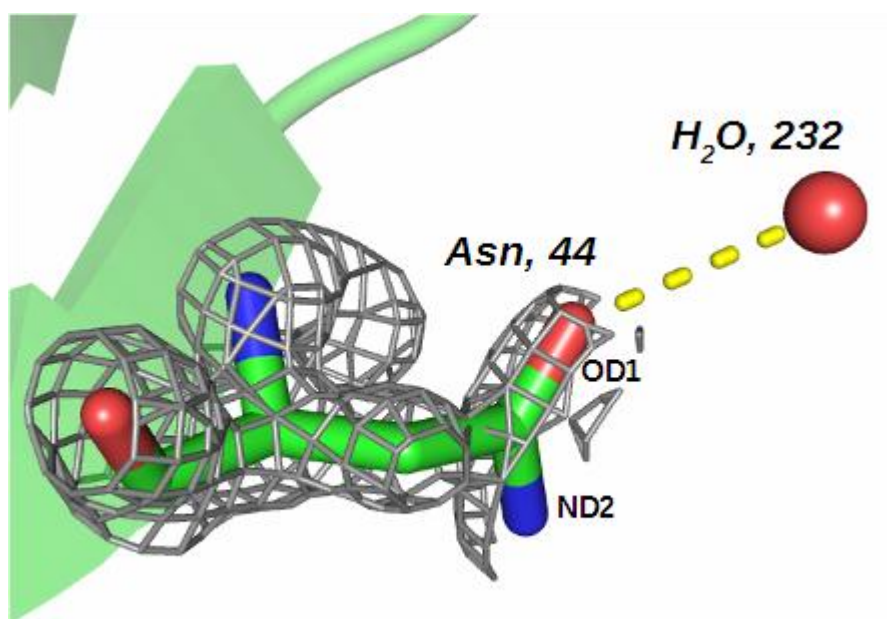


**Рисунок 8.** 128 остаток аргинина. Электронная плотность с подрезкой  $1,5\sigma$

Как и в предыдущем случае, очень плохо прописана электронная плотность. По ней нельзя с точностью определить положение боковой цепи, что и привело к описанным выше ошибкам модели.

### **Asn44**

Атомы N, C, O плохо различаются по электронной плотности, поэтому часто для таких остатков как Asn, His, Gln приходится делать инверсию боковой цепи. Рассмотрим 44 остаток аспарагина (Рисунок 9).



**Рисунок 9.** 44 остаток аспарагина. Электронная плотность с подрезкой 1,5 $\sigma$

На Рисунке 9 видно, что OD1 атом боковой цепи аспарагина образует водородную связь с водой. Скорее всего, OD1 нужно поменять с ND2, так лишь ND2 из этих двух атомов способен образовывать водородную связь.

### Сравнение с моделью из PDB\_REDO

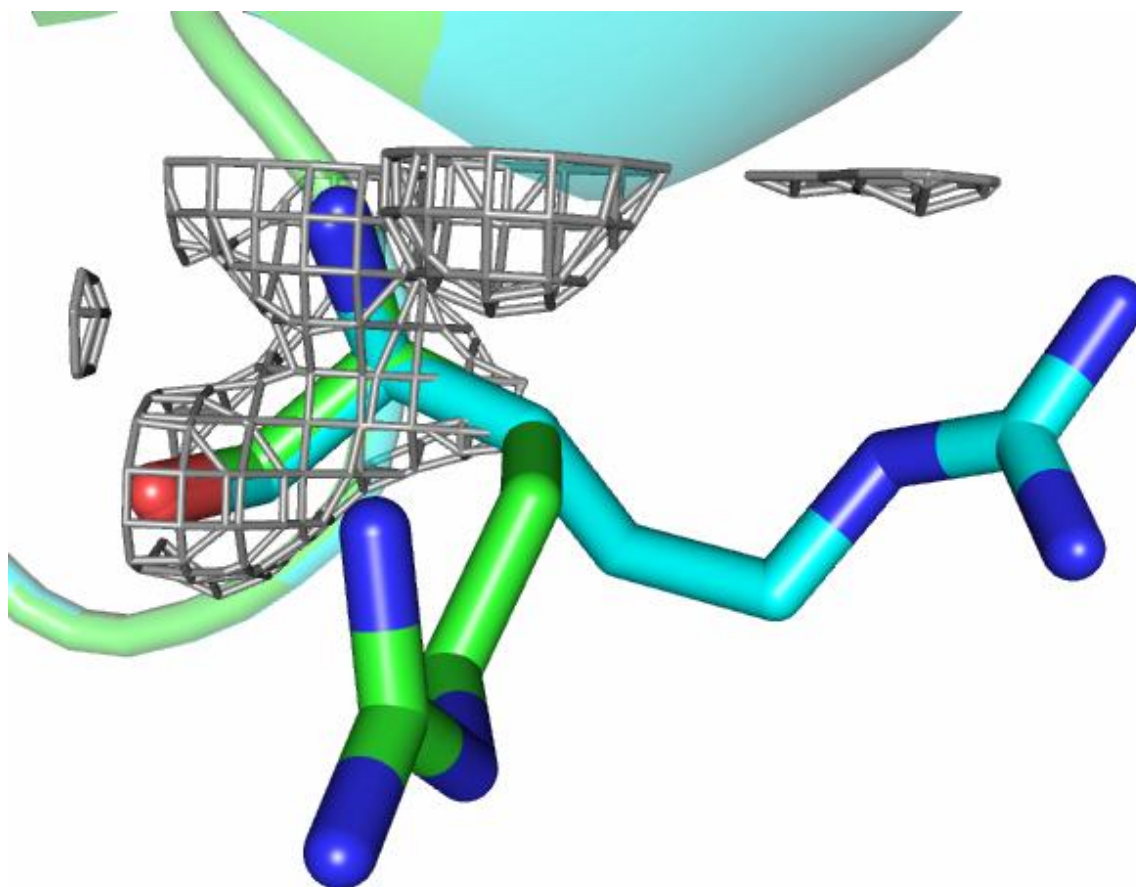
В базе данных PDB\_REDO была найдена стандартно оптимизированная по экспериментальным данным структура 193I. В результате «оптимизации» R-value и R-free структуры стали равны 0,1234 и 0,1655 соответственно (до оптимизации 0,184 и 0,226), что говорит об ухудшении модели после оптимизации. Судя по протоколу WHAT\_CHECK, были несколько улучшены некоторые параметры структуры: число недопустимых столкновений ван-дер-ваальсовых радиусов атомов уменьшилось с 12 до 6, RSR Z-score длин связей и углов уменьшилось, остатки стали еще лучше ложиться на карту Рамачандрана. Но число доноров/акцепторов водородной связи, не образующих ее, увеличилось с 2 до 3, ухудшилась конформация основной цепи.

Полученная оптимизированная структура была дана на вход серверу Molprobity. Таким образом, Clashscore оптимизированной структуры 193I равен 1 на 1000 атомов структуры, т.е. лишь 0.001% атомов структуры имеют недопустимые наложения (т.е. их Ван-дер-Ваальсовы радиусы перекрываются больше, чем на 0.4 Å). Структура имеет 99<sup>th</sup> перцентиль, что значительно лучше значения до оптимизации (80%). Оптимизированная структура 193I тоже содержит 2 остатка (1,83%) (Asn77, Arg112), конформация боковых цепей которых существенно отличается от ротамеров (ранее



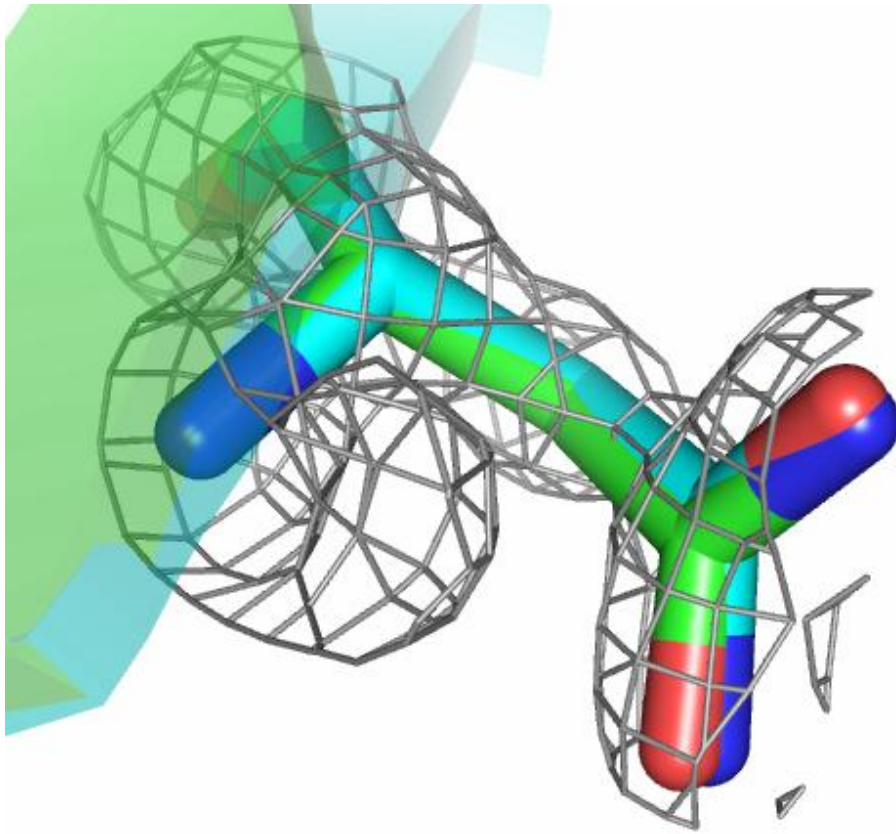
остатков тоже было 2, но у Arg112 была предпочитаемая конформация, у Asn77 допустимая), 1 остаток (Val99) принадлежат допустимой конформации и 106 остатков (97,25%) находятся в предпочитаемых конформациях. Лишь 1 остаток (Asn74, как и до оптимизации) лежит в допустимой области на карте Рамачандрана, остальные находятся в предпочитаемой области.

Рассмотрим как изменились маргинальные остатки, рассматриваемые нами ранее, после оптимизации. Наиболее существенное изменение наблюдалось там, где отсутствовала электронная плотность. Например, полностью поменялось положение боковой цепи 125 остатка аргинина (самого «проблемного» остатка), из-за чего его конформация стала принадлежать предпочитаемой области.



**Рисунок 10.** Конформация 125 остатка аргинина до (зеленая структура) и после оптимизации (голубая структура). Электронная плотность показана с подрезкой  $1,5\sigma$

Интересным оказалось то, что в 44 остатке аспарагина (где требовалась инверсия цепи), в результате оптимизации действительно была проведена инверсия (Рисунок 11).



**Рисунок 11.** Конформация 125 остатка аргинина до (зеленая структура) и после оптимизации (голубая структура). Электронная плотность показана с подрезкой 1,5 $\sigma$

В результате, после оптимизации модель улучшилась. Прежде всего, это наблюдается по значениям R-value и R-free, количеству столкновений ван-дер-ваальсовых радиусов. Кроме того, мы увидели изменение отобранных нами маргинальных остатков. Остальные параметры изменились незначительно. В целом, можно считать, что после оптимизации модель стала лучше.

### 3. Заключение

Структура 193I с высоким разрешением 1,33 Å расшифрована с некоторыми ошибками. В основном ошибки связаны с отсутствием электронной плотности. По сравнению с другими структурами такого же разрешения структура 193I решена на хорошем уровне (80<sup>ая</sup> перцентиль).

Структура содержит достаточное количество маргинальных остатков, однако большинство из них находится на периферии глобулы и экспонированы в раствор.

Полная оптимизация с помощью PDB\_REDO позволила улучшить качество модели: уменьшились R-value и R-free, количество столкновений ван-дер-ваальсовых радиусов. Кроме того, мы увидели изменение отображенных нами маргинальных остатков. Остальные параметры изменились незначительно.

#### 4. Список литературы

1. Wetter, L. R. & Deutsch, Hf. Immunological studies on egg white proteins IV. Immunochemical and physical studies of lysozyme. *J. Biol. Chem.* **192**, 237–242 (1951).
2. Yang, Y. & Hamaguchi, K. Hydrolysis of 4-methylumbelliferyl N-acetyl-chitotetraoside catalyzed by hen lysozyme. *J. Biochem. (Tokyo)* **88**, 829–836 (1980).
3. Blake, C. C. *et al.* Structure of hen egg-white lysozyme. A three-dimensional Fourier synthesis at 2 Angstrom resolution. *Nature* **206**, 757–761 (1965).
4. Blake, C. C., Fenn, R. H., North, A. C., Phillips, D. C. & Poljak, R. J. Structure of lysozyme. A Fourier map of the electron density at 6 angstrom resolution obtained by x-ray diffraction. *Nature* **196**, 1173–1176 (1962).
5. Vaney, M. C., Maignan, S., Ries-Kautt, M. & Ducruix, A. High-resolution structure (1.33 Å) of a HEW lysozyme tetragonal crystal grown in the APCF apparatus. Data and structural comparison with a crystal grown under microgravity from SpaceHab-01 mission. *Acta Crystallogr. D Biol. Crystallogr.* **52**, 505–517 (1996).
6. RCSB Protein Data Bank - RCSB PDB. Available at: <http://www.rcsb.org/pdb/home/home.do>. (Accessed: 8th January 2017)
7. CCP4 Program Suite. Available at: <http://www.ccp4.ac.uk/html/rotamer.html>. (Accessed: 11th January 2017)