

**МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ  
имени М.В.ЛОМОНОСОВА**

---

***ФАКУЛЬТЕТ БИОИНЖЕНЕРИИ И БИОИНФОРМАТИКИ***

Отчет по качеству расшифровки структуры белка бета-  
лактамазы 6FM6 методом рентгеноструктурного анализа

**Выполнила:**  
Студентка 4 курса ФББ  
Герасева Елизавета

**Москва  
2018 г.**

## АННОТАЦИЯ

В отчёте рассмотрены некоторые индикаторы качества модели структуры бФМб, а также проведён анализ индикаторов локального качества структуры для ряда аминокислотных остатков. В ходе работы над отчётом удалось познакомиться с возможностями, которые предоставляют серверы для оценки качества структур белковых молекул.

## ВВЕДЕНИЕ

Бета-лактамаза (EC: 3.5.2.6) катализирует гидролиз бета-лактамов с образованием соответствующей бета-аминокислоты (рис. 1). Субстратами бета-лактамаз являются бета-лактамные антибиотики: пенициллины, цефалоспорины, карбапенемы, монобактамы, эффективность гидролиза которых варьирует в зависимости от типа бета-лактамазы и субстрата. Таким образом бета-лактамазы обеспечивают основной механизм устойчивости грам-отрицательных бактерий к бета-лактамным антибиотикам.

### Enzyme Reaction (EC:3.5.2.6)

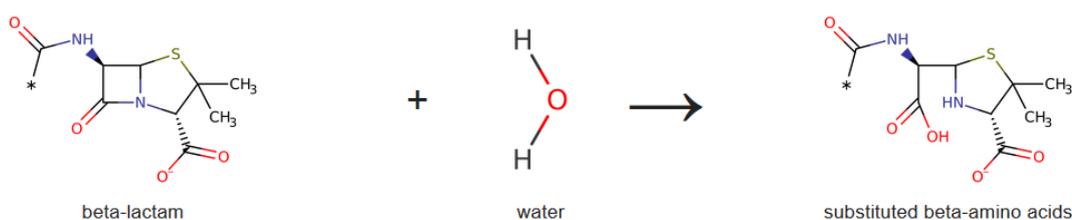


Рисунок 1. Реакция, катализируемая бета-лактамазами.

Бета-лактамазы делятся на два суперсемейства, отличающихся каталитическим механизмом: сериновые бета-лактамазы (включающие классы А, С и D), в которых основную роль играет каталитический остаток серина, и металло-бета-лактамазы (класс В), которым для катализа необходимы два (реже один) иона  $Zn^{+}$ .

Бета-лактамазы, принадлежащие классу С, больше известны как бета-лактамазы типа AmpC. Эти ферменты присутствуют у грам-отрицательных аэробных бактерий, таких как *Enterobacteriaceae*, *Pseudomonas spp.*, и *Acinetobacter spp.* Часто представители AmpC типа возникают у патогенных микроорганизмов, тем самым затрудняя лечение.

Бета-лактамаза TRU-1 из бактерии *Aeromonas enteropelogenes*, структура которой рассматривается в этом отчете, принадлежит к AmpC типу. Она эффективно катализирует пенициллины, цефалоспорины узкого спектра (кроме цефалоридина), цефокситин, менее эффективно – оксиминоцефалоспорины. Структура бФМб (рис. 2) была получена как

контроль для сравнения со структурой бFM7 (тот же белок в комплексе с авибактамом) [1].

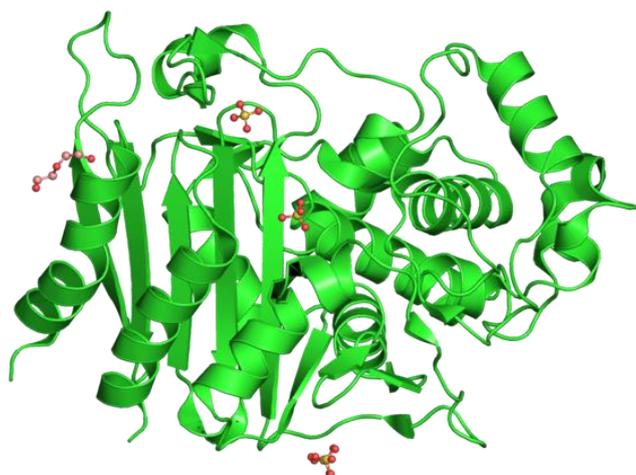


Рисунок 2. Структура бета-лактамазы бFM6.

## РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

### Общие сведения о структуре

Модель бFM6 была получена 4 сентября 2018. Авторы: S. Mangani, C. Pozzi, F. Di Pisa, F. De Luca, M. Benvenuti, J.D. Docquier. DOI публикации - 10.1002/cmdc.201800213.

Модель содержит 1 полипептидную цепь (41,82 кДа) длиной 356 (27-382) аминокислоты, 3 сульфат-иона и 1 молекулу триэтиленгликоля.

Метод решения фазовой проблемы – молекулярное замещение моделью 4GZB (Crystal structure of native AmpC beta-lactamase from *Pseudomonas aeruginosa* PAO1) с использованием программы MOLREP.

Тип кристаллографической симметрии - P 1 21 1. Некристаллографических симметрий нет. Параметры элементарной ячейки – в таблице 1. В одной ячейке находятся две молекулы[2].

Таблица 1. Параметры элементарной ячейки

A	45.885
B	78.345
C	48.190
Alpha	90.00
Beta	106.63
Gamma	90.00

Число измеренных рефлексов – 151680. Полнота набора структурных факторов - 93.3%.  
 Разрешение структуры – 1.05 Å . Диапазон разрешений структурных факторов 1.05 - 26.43 Å (данные получены с помощью сервиса EDS [3]).

### Общие значения индикаторов качества модели

На рисунке 3 представлены основные параметры, рассчитанные авторами при оценке качества полученной модели.

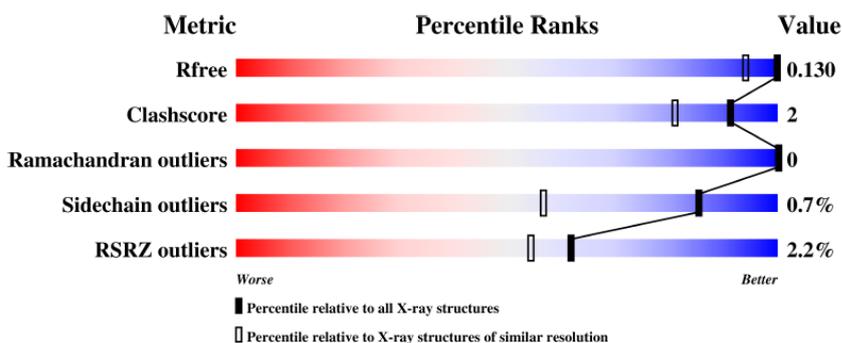


Рисунок 3. Индикаторы качества модели в сравнении со средними показателями для структур того же разрешения.

Как можно заметить, все индикаторы заметно лучше чем средние показатели для других структур. Теперь рассмотрим их подробнее.

Основные параметры для оценки качества модели это R и R-free факторы. Первый характеризует соответствие построенной модели экспериментальным данным. Он рассчитывается по формуле на рисунке 4.

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

Рисунок 4. Формула для расчета R-фактора.  $F^{obs}$  – наблюдаемые структурные факторы,  $F^{calc}$  – рассчитанные по модели амплитуды гармоник ряда Фурье.

R фактор минимизируется в процессе оптимизации модели. Хороший R фактор не превышает 0.25.

R-free характеризует переоптимизацию модели. Он вычисляется как и обычный R фактор, но для небольшой случайной выборки структурных факторов. В случае хорошей модели

R-free должен быть равным или ненамного больше общего R фактора, разница же более чем в 0.1 говорит о переоптимизации.

В данной модели значение R фактора 0.115, что говорит о хорошем соответствии модели экспериментальным данным. Для вычисления R-free использовалась тестовая выборка из 7107 структурных факторов (5% от всех). Значение R-free 0.129, разница с R фактором составляет 0.014, следовательно, нет оснований подозревать переоптимизацию.

Для оценки качества модели также используются карты Рамачандрана. Эти карты отражают конфигурацию торсионных углов остова белка. По ним можно определить наличие маргинальных остатков, то есть остатков с высокоэнергетической, а потому запрещенной конфигурацией остова. В хорошей модели их не более 10%.

С помощью сервиса EDS получена карта Рамачандрана для всех остатков (рис.5). Найденные маргинальные остатки (4 штуки или 1,2%) описаны в таблице 2.

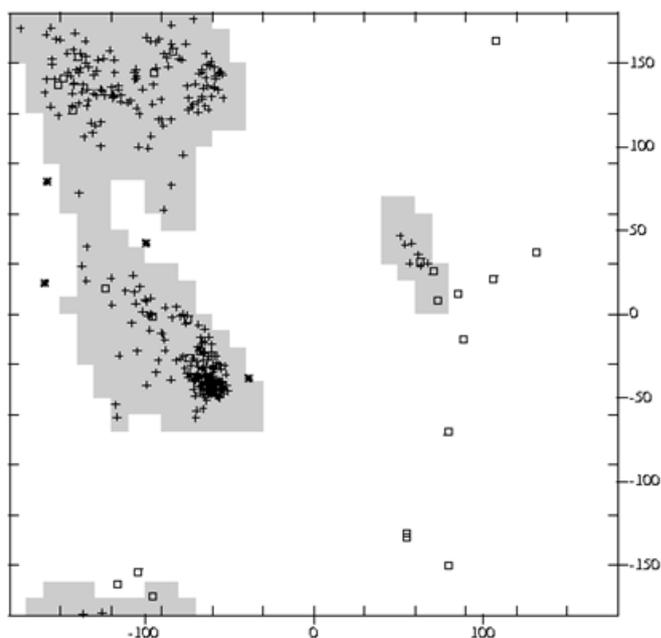


Рисунок 5. Карта Рамачандрана из EDS. Квадраты – глицины, кресты – разрешенные остатки, звездочки – маргиналы. Серым обозначена разрешенная область.

Таблица 2. Маргинальные остатки по EDS.

Residue	Type	Phi Angle	Psi Angle
A91	ALA	-157.8	79.1
A220	TYR	-159.2	18.6
A273	TYR	-38.4	-38.5
A338	ASN	-99.2	42.2

Однако чаще для построения карт Рамачандрана и определения маргинальных остатков используется программа MolProbity [4]. С ее помощью были построены карты (рис.6-7).

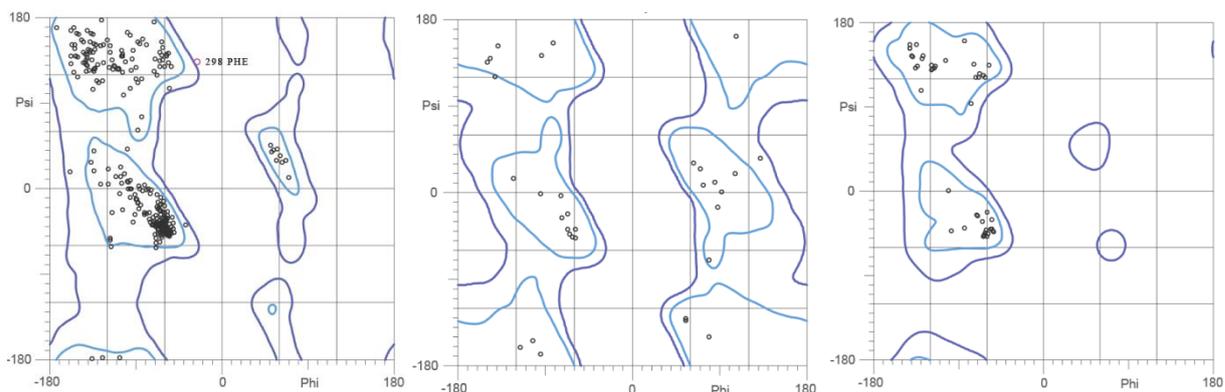


Рисунок 6. Карты Рамачандрана для bfm6, построенные MolProbity. Слева - для всех остатков кроме глицина, валина, изолейцина и пролина. В центре – для глицина. Справа – для валина и изолейцина. Голубые линии ограничивают предпочтительные области, синии – допустимые области.

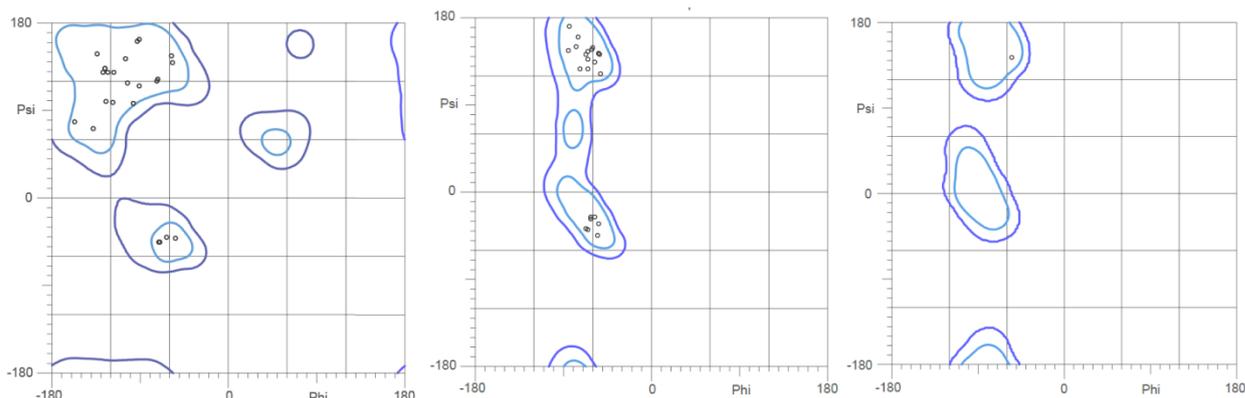


Рисунок 7. Карты Рамачандрана для bfm6, построенные MolProbity. Слева – пре-пролин, в центре – транс-пролин, справа – цис-пролин.

Как можно заметить, программа нашла только 1 маргинальный остаток (298 PHE). Еще 8 остатков расположены в допустимой области, остальные же – в предпочтительной.

Что интересно, при валидации модели авторами не было найдено ни одного маргинального остатка. Однако, несмотря на это, количество маргинальных остатков, находимых разными программами, незначительно и говорит о высоком качестве структуры (правда, MolProbity выставляет порог в 0,05%).

Помимо карт Рамачандрана, MolProbity рассчитывает и другие характеристики качества структуры (таблица 3).

Таблица 3. Индикаторы качества модели, рассчитанные MolProbity.

All-Atom Contacts	Clashscore, all atoms:	1.79	97 <sup>th</sup> percentile* (N=208, 1.05Å ± 0.25Å)	
	Clashscore is the number of serious steric overlaps (> 0.4 Å) per 1000 atoms.			
Protein Geometry	Poor rotamers	4	1.25%	Goal: <0.3%
	Favored rotamers	310	96.88%	Goal: >98%
	Ramachandran outliers	1	0.28%	Goal: <0.05%
	Ramachandran favored	345	97.46%	Goal: >98%
	MolProbity score <sup>^</sup>	1.09	96 <sup>th</sup> percentile* (N=830, 1.05Å ± 0.25Å)	
	Cβ deviations >0.25Å	0	0.00%	Goal: 0
	Bad bonds:	0 / 3167	0.00%	Goal: 0%
Bad angles:	1 / 4358	0.02%	Goal: <0.1%	
Peptide Omegas	Cis Prolines:	1 / 21	4.76%	Expected: ≤1 per chain, or ≤5%

Clashscore представляет собой число недопустимых наложений на 1000 атомов. Как можно заметить, эта величина является 97% перцентилем, то есть лучше, чем у 97% структур.

Также программа находит остатки, маргинальные по отклонению от ротамеров боковыми цепями (poor rotamers). Их было найдено 4 (1,25%), и программа, видимо, решила, что это слишком много. Еще был найден 1 «плохой угол» и 1 цис-пролин, но их количество в пределах нормы. Прочих отклонений от нормы найдено не было. В таблице 4 все маргинальные остатки.

Таблица 4. Маргинальные остатки по MolProbity.

Residue	Resname	Type	RSR
A32	LYS	Bad rotamer	0.130
A210	VAL	Bad rotamer	0.224
A274	PRO	Cis proline	0.061
A298	PHE	Ramachandran outlier	0.119
A311	ASN	Bad rotamer+bad angle	0.050
A316	THR	Bad rotamer	0.125

Что интересно, в области большинства этих маргиналов наблюдаются пики в Real-Space-R факторе (смотри рисунок 8). Эта величина (сокращенно RSR) характеризует, насколько

модель атома или группы атомов (в данном случае, 1 остатка) соответствует экспериментальной электронной плотности. Его величина должна быть не больше 0,2.

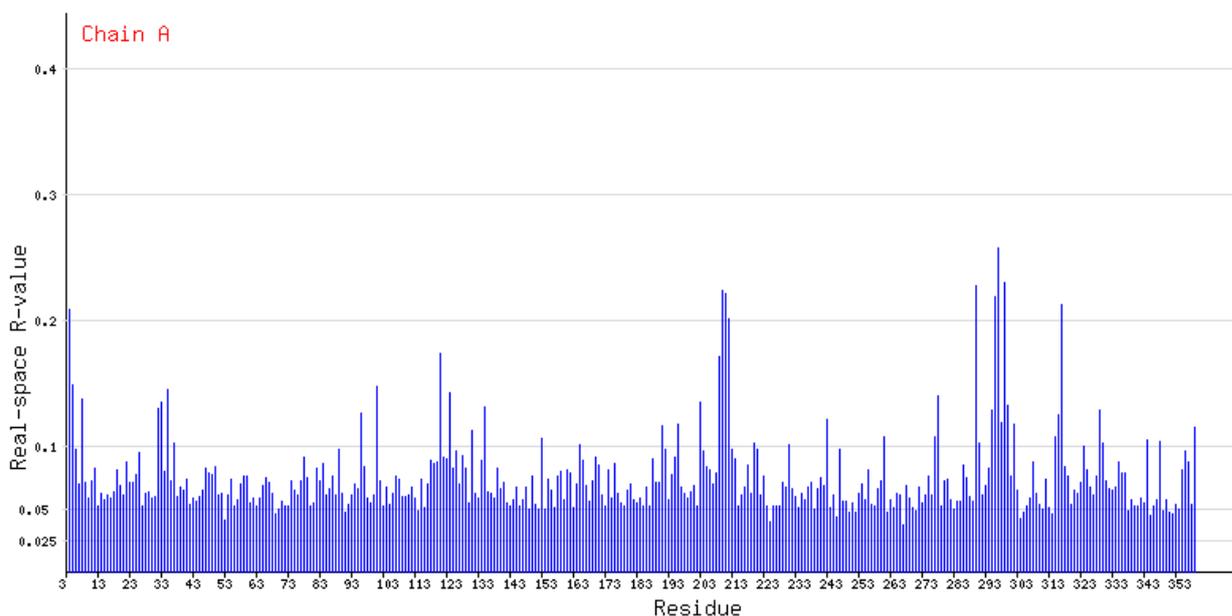


Рисунок 8. Распределение RSR по длине цепи (EDS).

В данном случае высокий RSR может являться следствием попытки подогнать конформации остатков к относительно допустимым значениям с ухудшением их соответствия электронной плотности. С другой стороны, что более вероятно, электронная плотность в этих областях просто плохо определилась, тем более что пики охватывают несколько остатков, а не один. В таблице 5 остатки, маргинальные по RSR (то есть с  $RSR > 0.2$ ).

Таблица 5. Маргинальные по RSR остатки (EDS).

Residue	Resname	RSR
A4	GLU	0.209
A210	VAL	0.224
A211	ASN	0.221
A212	PRO	0.201
A290	PHE	0.228
A296	LYS	0.219
A297	PRO	0.258
A299	VAL	0.230
A317	ASN	0.213

Однако о качестве расшифровки отдельных остатков обычно судят не по RSR, а по Z-score, который является нормализованным RSR и показывает соответствие остатка его электронной плотности в сравнении с аналогичными остатками в структурах с тем же разрешением. Хороший Z-score не больше 2. На рисунке 9 представлено распределение Z-score по цепи.

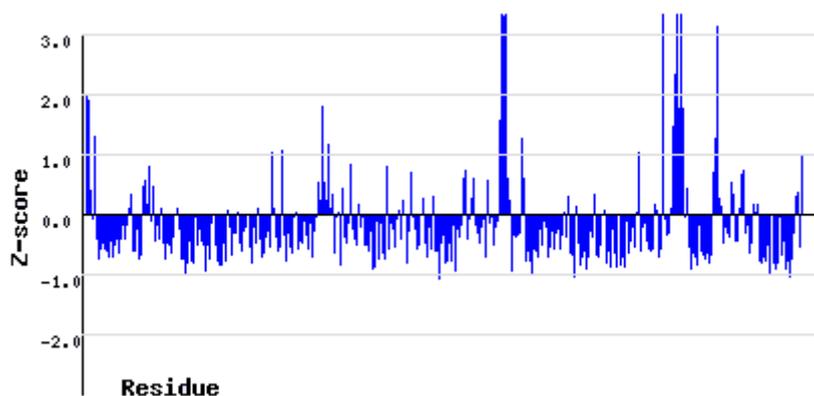


Рисунок 9. Распределение Z-score по длине цепи (EDS).

Видно, что в целом качество структуры хорошее ( $Z\text{-score} < 0$ ), однако есть и пики, причем довольно большие. В таблице 6 маргиналы по Z-score.

Таблица 6. Маргинальные по Z-score остатки (EDS).

Residue	Resname	Z-score
A210	VAL	4.47
A211	ASN	3.31
A212	PRO	3.32
A290	PHE	5.82
A296	LYS	2.32
A297	PRO	4.86
A299	VAL	4.66
A317	ASN	3.12

Как можно видеть, это те же маргинальные остатки, что были маргинальными по RSR. Стоит заметить, что авторы тоже проводили валидацию и рассчитывали Z-score, но у них значения получились немного другие, хотя список маргиналов практически совпал с полученным EDS.

Рассмотрим подробнее некоторые маргинальные остатки.

## Анализ отдельных маргинальных остатков

### VAL 210

Остаток, маргинальный по отклонению от ротамеров боковой цепью (bad rotamer). Угол  $\chi$  равен 150.4, вероятность встретить такой ротамер – 0,02% (согласно MolProbity). Остаток имеет высокий RSR (0.224) и самый высокий Z-score (4.47). Также он отличается довольно высоким B фактором (26.22 при среднем 16.05). Рассмотрим расположение остатка в модели (рисунок 10-11).

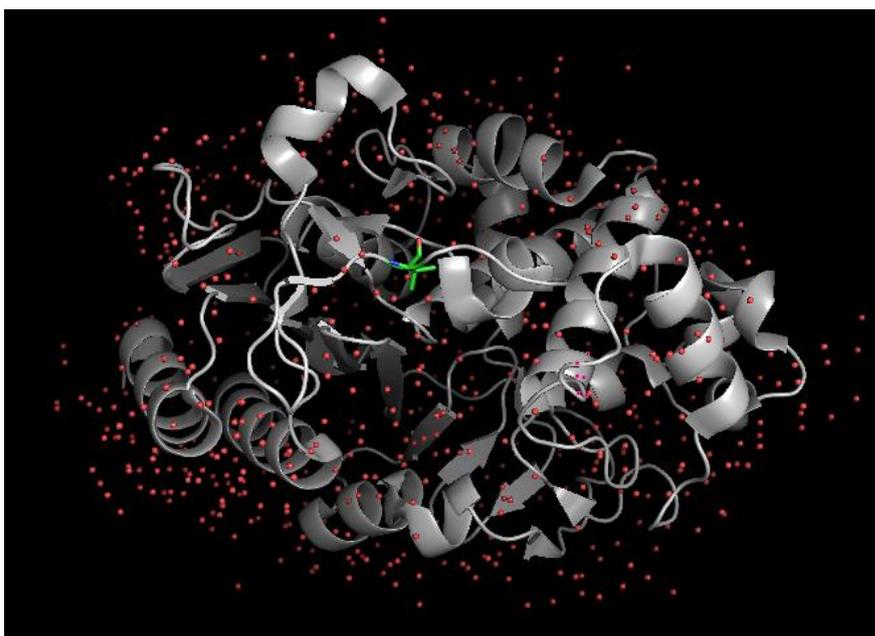


Рисунок 10. Расположение VAL210 в модели белка.

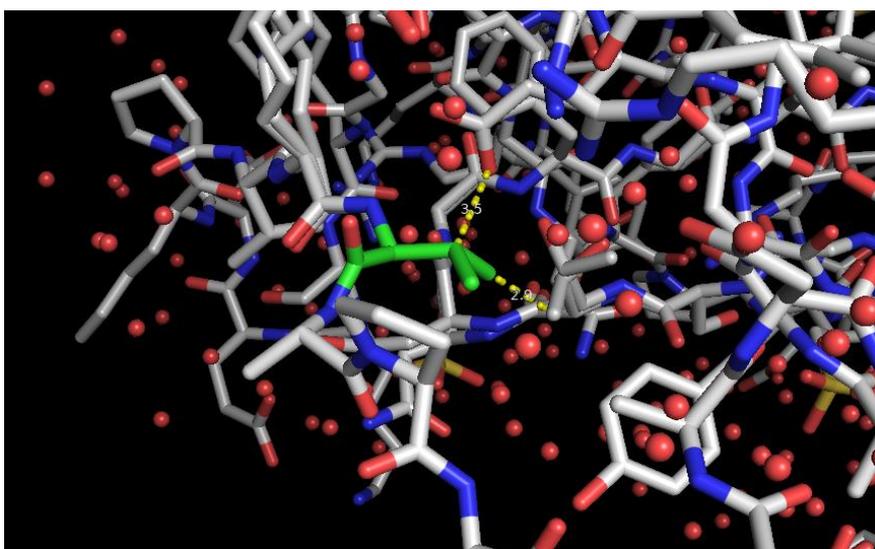


Рисунок 11. Расположение VAL210 в модели белка. Зеленый – VAL210, желтый – расстояния до ближайших остатков, красные шарики – вода.

Как можно заметить, остаток расположен на поверхности белковой глобулы, однако боковая цепь смотрит внутрь. То, что он не располагается в стабильных вторичных структурах, может говорить о высокой подвижности этого участка цепи, и, как следствие, сложностям в расшифровке. Неясно, почему этот ротамер является запрещенным, возможно из-за того, что конформация заслоненная по одной из групп боковой цепи. Такое положение остатка можно объяснить его окружением, а именно остатком глутамата сверху и несколькими гидрофобными остатками, стабилизирующими эту конформацию. Теперь рассмотрим электронную плотность (рис. 12).

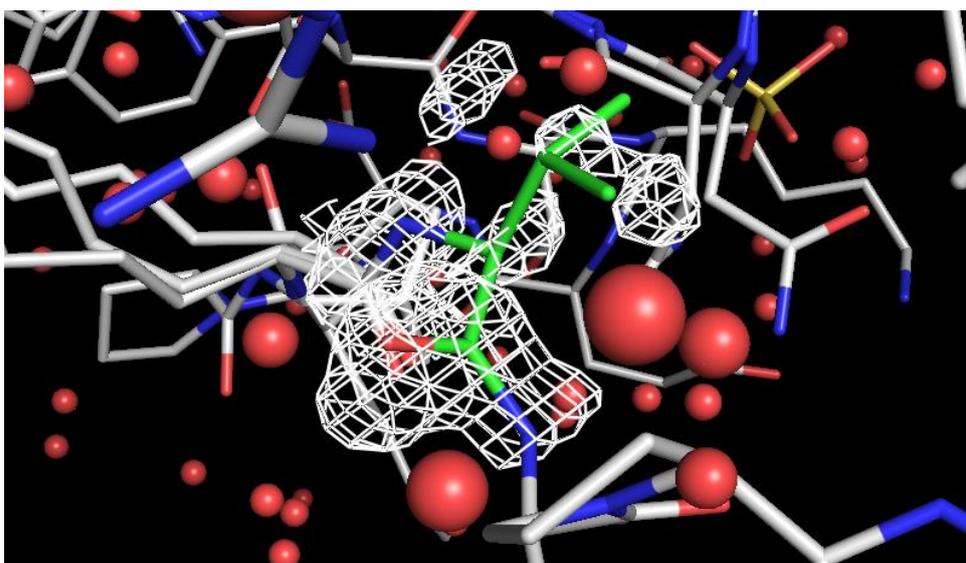


Рисунок 12. VAL210 и его электронная плотность. Уровень подрезки  $Z=1.0$ .

Как можно заметить, электронная плотность не слишком располагает к правильной расшифровке этого остатка. Вокруг одного из углеродов боковой цепи электронная плотность настолько низкая, что при уровне подрезки 1 ее не удастся увидеть, а вокруг главной цепи она наоборот высокая. Как мне кажется, конформацию боковой цепи расшифровали неправильно, ведь она почти не совпадает с электронной плотностью (поэтому RSR и Z-score такие высокие). Над остатком расположен пик электронной плотности, который, по моему мнению, должен соответствовать истинному положению углерода боковой цепи. Хотя, также возможно, что низкая электронная плотность свидетельствует о подвижности боковой цепи и изменчивости конформации.

## 211 ASN

Причина выбора этого остатка в том, что в нем отсутствуют 3 атома: CG, OD1 и ND2. Он, кстати, сосед VAL210, который я уже рассматривала. Его характеристики: B-value= 40.48 (максимальный), RSR=0.221, Z-score=3.31 (высокие). Расположен на поверхности

глобулы, боковая цепь смотрит в раствор (смотри рисунок 11, чуть ниже и левее VAL210, выглядит как аланин). Такое расположение говорит о его большой подвижности, так что неудивительно, что не удалось его расшифровать. Рассмотрим электронную плотность (рисунок 13).

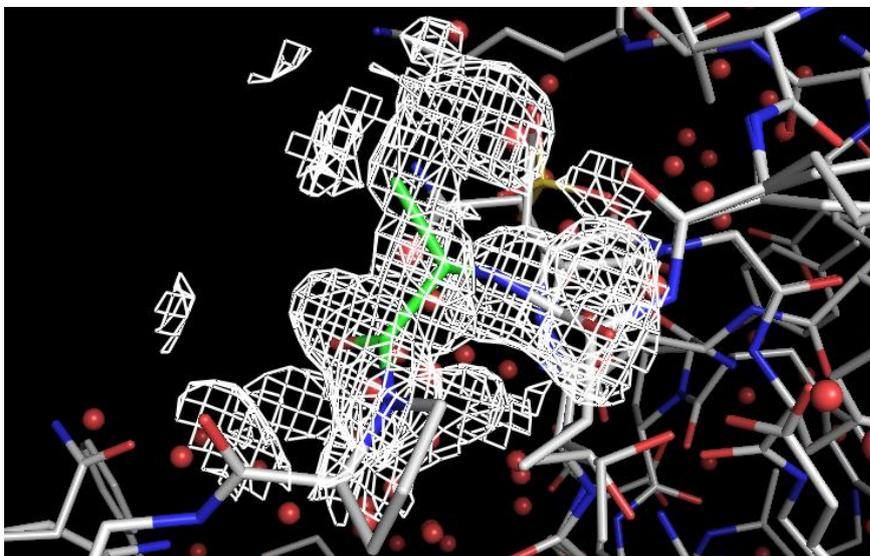


Рисунок 13. Электронная плотность ASN211. Уровень подрезки  $Z=0.5$ .

Очевидно, что по такой электронной плотности сложно определить структуру. Однако, как мне кажется, даже те атомы, которые расшифровать удалось, следовало поместить не там, где они есть. В принципе, место, где должны быть отсутствующие атомы, легко заметить по треугольнику из делокализованной электронной плотности. Если СА сдвинуть ниже, СВ поместить на место СА, а СG на место СВ, то оставшиеся гетероатомы легко бы вписались. Правда, тогда бы скорее всего возникли недопустимые углы и расстояния между атомами, или запрещенные конформации. В любом случае, этот участок, судя по всему, очень подвижный, так что в принципе можно было бы поместить туда остаток в любой конформации.

### **PRO 274**

Этот пролин имеет не самую популярную цис-конформацию, угол  $\omega=0.43$ . Однако прочие его характеристики достаточно убедительные: В-фактор равен 7.3, что говорит о низкой подвижности, RSR=0.061, Z-score равен -0.47, то есть соответствие электронной плотности весьма хорошее. Рассмотрим остаток в модели (рис. 14).

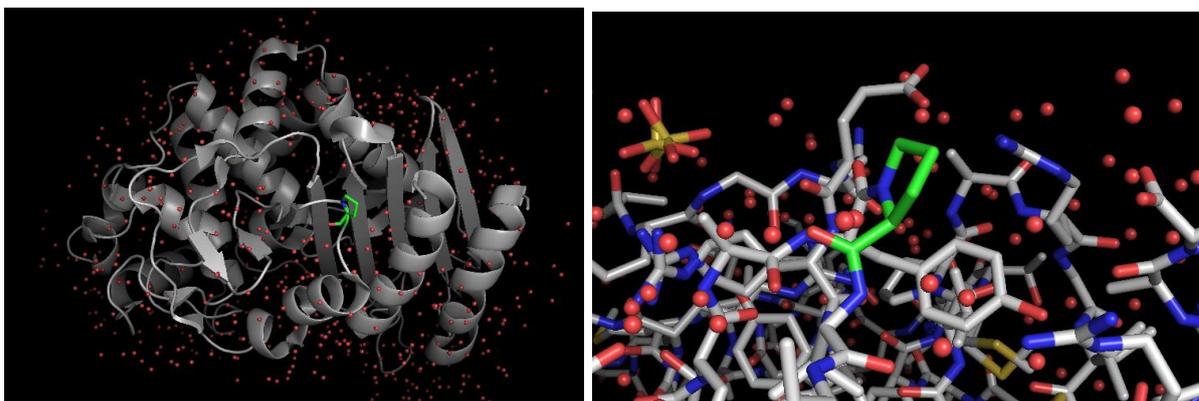


Рисунок 14. Расположение PRO274 в модели белка. Справа – фрагмент.

Как и предыдущий остаток, этот расположен на поверхности глобулы, но на этот раз, как ни парадоксально, его гидрофобная циклическая часть смотрит в сторону раствора и ничем не стабилизирована, что весьма подозрительно. Однако его электронная плотность весьма однозначно свидетельствует в пользу именно этой конформации. (рис. 15).

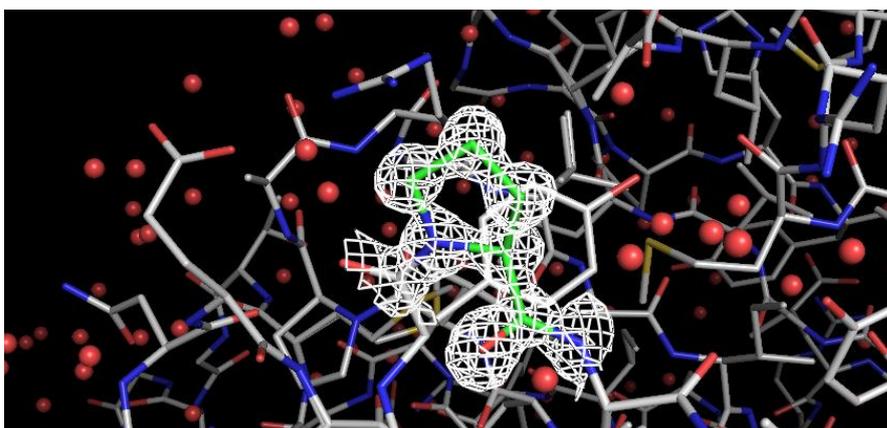


Рисунок 15. PRO274 и его электронная плотность. Уровень подрезки  $Z=1.0$ .

Судя по всему, такая странная конформация, как и невыгодное окружение, обусловлены взаимодействием с соседней молекулой в кристалле. Доказательство – на рисунке 16.

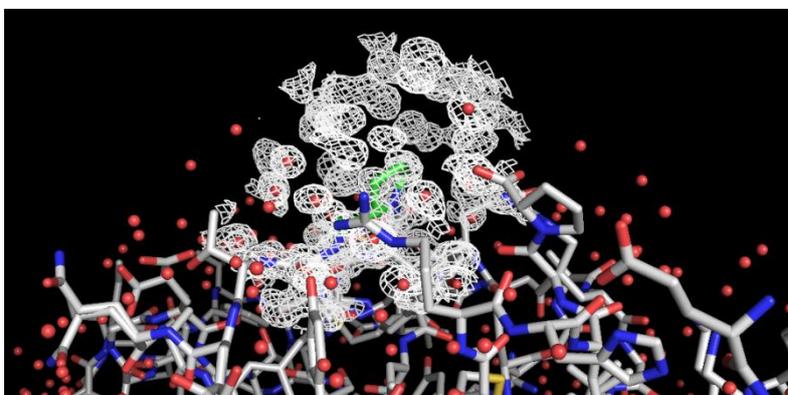


Рисунок 16. PRO274 и электронная плотность его окружения на расстоянии 6 ангстрем. Уровень подрезки  $Z=1.0$ .

Можно заметить электронную плотность соседней цепи, которая, судя по расстоянию, взаимодействует с рассматриваемым остатком и стабилизирует его боковую цепь. Как бы повел себя этот пролин в растворе – неизвестно. В такие моменты проникаешься недоверием к рентгеноструктурному анализу.

## 298PHE

Следующим рассмотрим остаток, маргинальный по Рамачандрану, то есть по конформации остова. Сразу скажу, что в структуре два варианта конформации, и только одна из них запрещенная. Угол  $\phi = -26.2$ , угол  $\psi = 134.4$ , вероятность встретить такую конформацию 0.02%. Прочие характеристики: RSR=0.12, Z-score=1.75 (не очень хорошие, но в пределах нормы), B-фактор равен 18.61 (чуть выше среднего). Стоит отметить, что соседние с данным остатком обладают высокими RSR и Z-score, то есть плохо соответствуют электронной плотности. Теперь посмотрим на его расположение и окружение (рис 17).

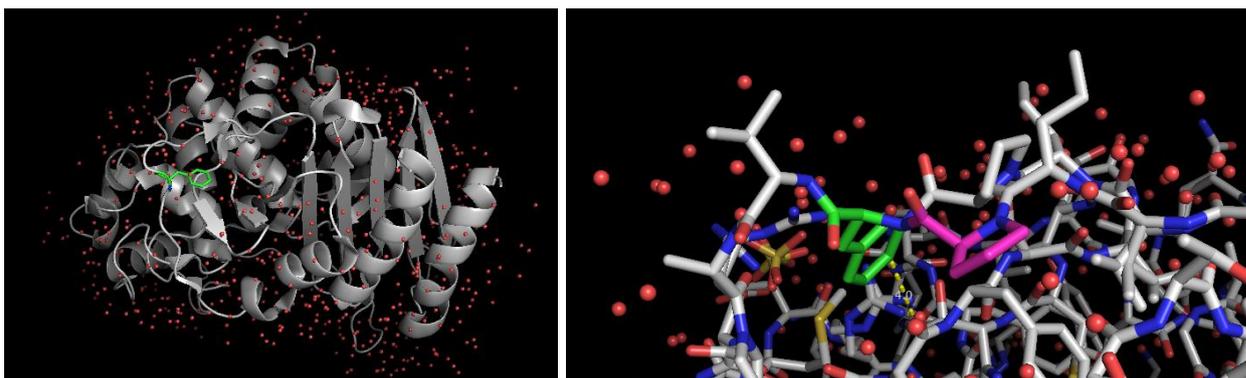


Рисунок 17. Расположение PHE298 (зеленый) в модели белка. Справа – фрагмент. Фиолетовым отмечен вариант расположения цепи с запрещенной конформацией.

И снова остаток расположен на поверхности, а боковой цепью смотрит внутрь глобулы. Если посмотреть на соседние остатки, можно заметить что гидрофобные боковые цепи смотрят в раствор, видимо, снова имеет место взаимодействие с соседней молекулой. Теперь смотрим на электронную плотность (рис. 18).

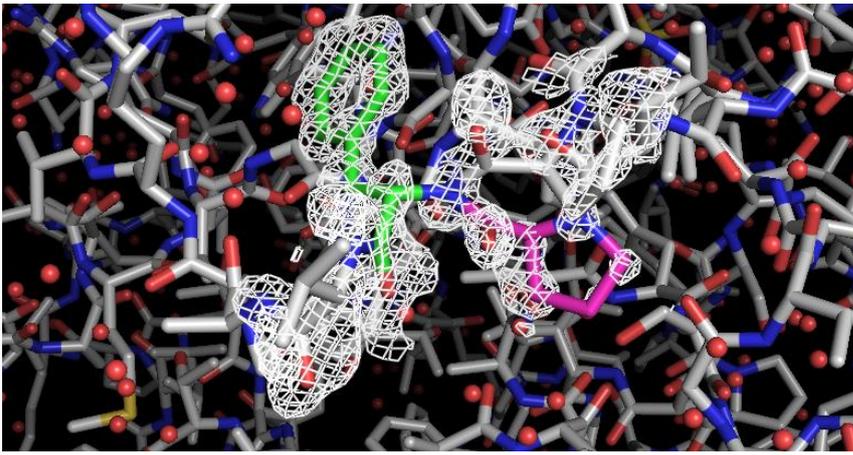


Рисунок 18. Электронная плотность PHE298 и его соседей. Уровень подрезки  $Z=1.0$ .

Как мы видим, два варианта конформации действительно имеют место, судя по всему, соседний пролин довольно подвижен. Однако, как оказалось, никакой стабилизации соседней молекулой здесь нет, так что пролин действительно остается незащищенным от молекул растворителя, однако в случае запрещенной Рамачандраном конформации он глубже погружен в глобулу, что может обеспечивать ему некоторую защиту. Вероятно происходит переключение между невыгодной по Рамачандрану и невыгодной по окружению конформациями, из-за чего возникают альтернативные положения цепи.

### 311 ASN

Этот остаток имеет запрещенную конформацию боковой цепи ( $\chi_1=81.1$ ,  $\chi_2=164.8$ , вероятность 0%). Помимо этого имеет плохой угол CA-CB-CG ( $117.1^\circ$ ). Остальные его параметры: RSR=0.05, Z-score=-0.78 (хорошие), B-value= 4.6 (намного ниже среднего, что говорит о малой подвижности). Теперь посмотрим на расположение в структуре (рис. 19).

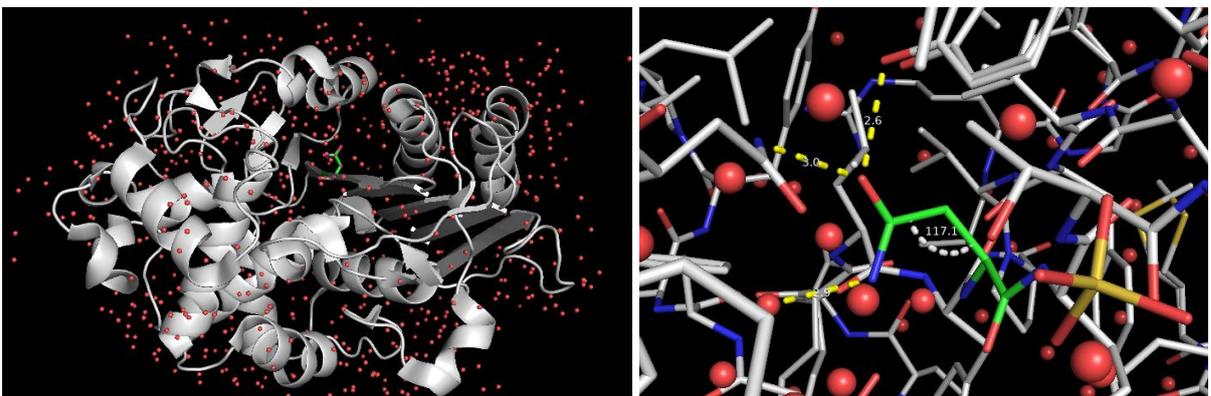


Рисунок 19. Расположение ASN311 (зеленый) в модели белка. Справа – фрагмент.

В отличие от предыдущих случаев, этот остаток расположен в центре структуры, стабилизирован водородными связями. Как можно заметить, конформация заслоненная по  $\chi_2$ , а один из углов слишком большой для  $\text{sp}^3$ . На рисунке 20 – электронная плотность.

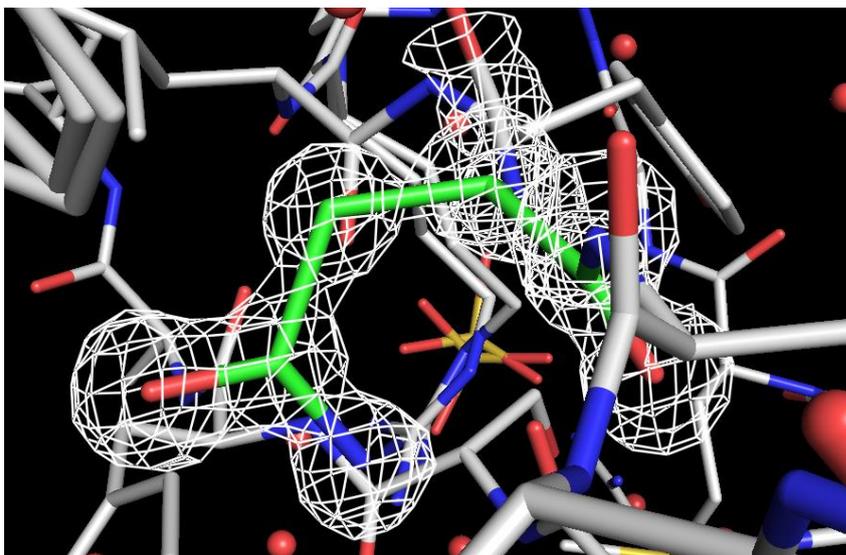


Рисунок 20. Электронная плотность ASN311. Уровень подрезки  $Z=2.0$ .

Расположение всех атомов в остатке определяется достаточно однозначно. В итоге, если заслоненную конформацию можно объяснить влиянием окружения, то причин для аномального угла не наблюдается. Можно было бы предположить отщепление протона от СВ и образование енолят-иона, между СВ-СG даже наблюдается какая-то делокализация электронной плотности. Тем более что остаток находится вблизи активного центра [5]. Однако это объяснение слишком экзотическое.

### PDB-redo

База данных PDB\_redo [6] содержит улучшенные версии моделей, существующих в PDB. Улучшенная bfm6 там тоже есть. В таблице 7 статистика изменений. В таблице 8 – основные индикаторы качества улучшенной модели в сравнении с исходной.

Таблица 7. Статистика изменений по структуре bfm6 в PDB-redo.

Description	Count
Rotamers changed	0
Side chains flipped	3
Waters removed	58
Peptides flipped	0
Chiralities fixed	0
Residues fitting density better	75
Residues fitting density worse	3

Помимо изменений, указанных в таблице, в новом PDB дорисованы все отсутствующие атомы в боковых цепях (например, в ASN211). Как можно заметить, после небольшого

количества изменений 75 остатков улучшили свой RSR (из них 19 это вода, 1 – лиганд), 3 – ухудшили (из них 1 вода, второй – это LYS206, которому дорисовали боковую цепь).

С помощью MolProbity был проведен анализ новой структуры на предмет маргинальных остатков. Результаты сравнили с результатами по структуре из PDB (тоже таблица 8).

Таблица 8. Сравнение индикаторов качества структуры бfmб в PDB и PDB redo.

	<b>PDB</b>	<b>PDB-REDO</b>
R	0,1154	0,1058
R-free	0,1301	0,1217
Bond length RMS Z-score	0,362	0,421
Bond angle RMS Z-score	0,721	0,851
<b>MolProbity statistics</b>		
Clashscore, all atoms:	1.79	3.72
Poor rotamers	4	3
Favored rotamers	310	311
Ramachandran outliers	1	1
Ramachandran favored	345	345
C $\beta$ deviations >0.25Å	0	0
Bad bonds:	0 / 3167	1 / 3173
Bad angles:	1 / 4358	8 / 4365
Cis Prolines:	1 / 21	1 / 21
Twisted Peptides:	0 / 371	1 / 371

Как мы видим, улучшилось соответствие электронной плотности и исчез 1 плохой ротамер, прочие же параметры остались без изменений или даже ухудшились (плохие длины связей и углы, clash-score). Так что, можно сказать, что улучшения достаточно условные, и еще неизвестно, какая из структур лучше.

## **ВЫВОДЫ**

Проанализировав качество структуры бfmб, можно сделать следующие выводы: данная структура имеет высокое разрешение и хорошее качество расшифровки. Однако имеются некоторые локальные проблемы с расшифровкой, связанные прежде всего с особенностями рентгеноструктурного анализа как метода, а именно: подвижность отдельных участков цепи и, как следствие, плохое разрешение электронной плотности; необходимость получения кристалла и, как следствие, несоответствие полученной модели структуре в растворе. Оптимизация структуры в PDB\_redo, как мне кажется, не имеет

смысла, так как улучшение соответствия экспериментальным данным происходит за счет химической достоверности.

## **ИСТОЧНИКИ**

- [1] Pozzi C. et al (2018) Atomic-Resolution Structure of a Class C  $\beta$ -Lactamase and Its Complex with Avibactam, ChemMedChem, 18;13(14):1437-1446, doi: [10.1002/cmdc.201800213](https://doi.org/10.1002/cmdc.201800213).
- [2] <https://www.rcsb.org/structure/6fm6>
- [3] <http://eds.bmc.uu.se/eds/>
- [4] <http://molprobitry.biochem.duke.edu/>
- [5] <https://www.ebi.ac.uk/thornton-srv/m-csa/entry/257/>
- [6] <https://pdb-redo.eu/db/6fm6>