

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

**ФАКУЛЬТЕТ БИОИНЖЕНЕРИИ И БИОИНФОРМАТИКИ**

**Отчёт по качеству RSA расшифровки структуры белка**

**Отчет студентки 4-го курса,  
Мироновой Екатерины Андреевны**

**Москва 2019 г.**

## Содержание

<b>1. Аннотация.....</b>	<b>3</b>
<b>2. Введение.....</b>	<b>3</b>
<b>3. Результаты и обсуждения.....</b>	<b>4</b>
3.1. Общая информация о модели.....	4
3.2. Значения индикаторов качества модели в целом.....	5
3.3. Таблица маргинальных остатков.....	9
3.4. Анализ маргинальных остатков.....	10
3.5. Сравнение PDB с PDB_redo.....	14
<b>4. Заключение.....</b>	<b>16</b>
<b>5. Список литературы.....</b>	<b>17</b>

## 1. Аннотация

В данной работе был произведен анализ качества расшифровки структуры белка 6OM5 — гемофилина, вырабатываемого *Haemophilus haemolyticus*. Были описаны основных показатели качества модели и их соответствие общепринятым значениям и рассмотрены несколько маргинальных аминокислотных остатков.

## 2. Введение

Бактерии-комменсалы служат важной линией защиты от патогенов, но основные молекулярные механизмы остаются малоизученными. В бактерии *Haemophilus haemolyticus* был найден гемофин — гем-связывающий белок, ингибирующий рост патогена — нетипируемого *Haemophilus influenzae* (NTHi) в культуре. Бактерии с нокаутом гемофилина обладают ограниченной способностью использовать свободный гем для роста. Гемофин является гемофором, и ингибирование NTHi происходит при недостатке гема, что повышает вероятность того, что конкуренция со стороны *H. haemolyticus*, продуцирующего гемофин, может противодействовать колонизации NTHi в дыхательных путях. [1]



Рис. 1. Структура гемофилина.

Также РСА гемофилина показал, что белок не принадлежит ни к одной из

известных гем-связывающих укладок белка, что позволяет предположить независимую эволюцию. [1]

### **3. Результаты и обсуждения**

#### 3.1 Общая информация о модели

Состав комплекса: комплекс состоит из 1 субъединицы гемофора и лигандов — сульфат-иона, иона хлора, глицерина, гема. Всего в белке 252 остатка.

Год: 2018.

Авторы структуры: Torrado, M.; Walshe, J.L.; Mackay, J.P.; Guss, J.M.; Gell, D.A.

Статья: Latham RD, Torrado M, Atto B, Walshe JL, Wilson R, Guss JM, Mackay JP, Tristram S, Gell DA. A heme-binding protein produced by *Haemophilus haemolyticus* inhibits non-typeable *Haemophilus influenzae*. // *Mol Microbiol.* . 2019 Nov 19.

Метод разрешения фазовой проблемы: SAD.

Число измеренных рефлексов: 58934.

Разрешение: 1.6 Å.

Полнота набора структурных факторов: 99.99 %.

Диапазон разрешений структурных факторов: 1.6 Å – 41.49 Å.

Параметры кристаллографической ячейки и кристаллографическая группа: длины направляющих векторов 91.043, 91.043, 97.484 Å. Углы между ними: 90°, 90°, 120°.

Наличие некристаллографических симметрий в асимметрической ячейке: 6 молекул в ячейке, симметрия винтовая. [2]

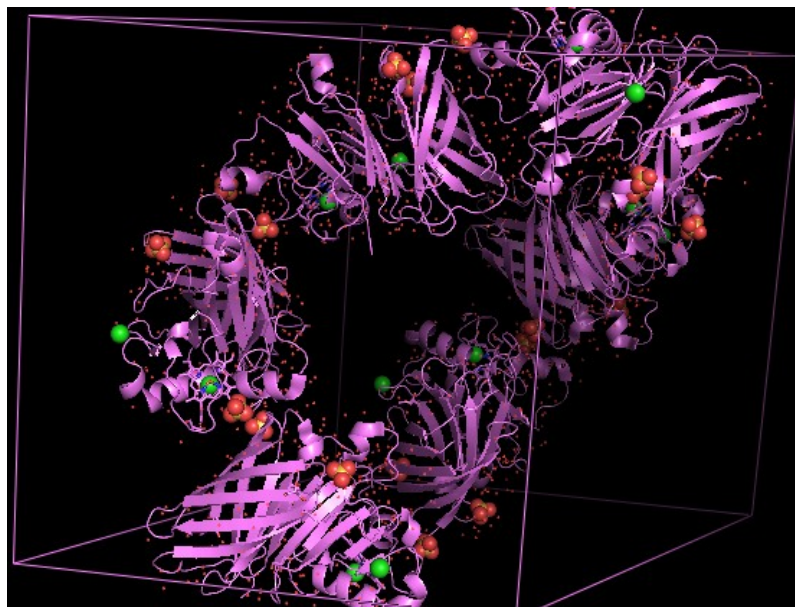


Рис. 2. Ассиметрическая ячейка кристалла.

### 3.2 Значения индикаторов качества модели в целом

Основные показатели качества рентгеноструктурной модели: разрешение, R-фактор, R-free, карты рамачандрана, ротамеры, RSR (пространственный R-фактор), комфортность окружения.

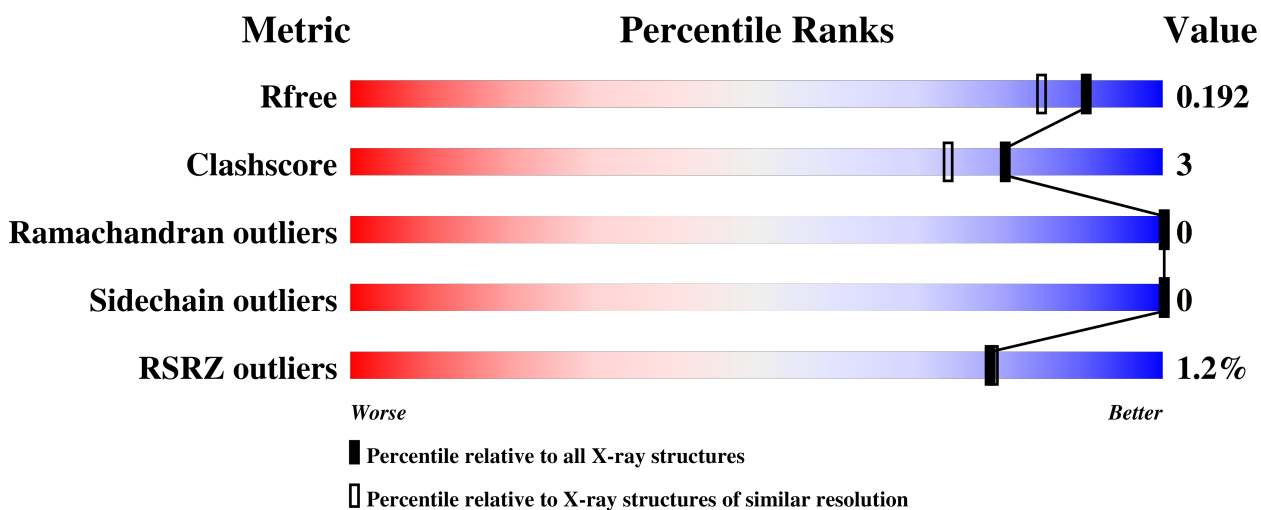


Рис. 3. Основные индикаторы качества модели относительно банка PDB. Черная метка - перцентиль относительно всех PCA структур базы данных, белая – перцентиль относительно структур, близких по разрешению [2].

#### R-фактор и R\_free

R-фактор — характеризует соответствие модели ее экспериментальным

значениям (модулям структурных факторов). Хорошие значения:  $< 25\%$ . Плохие:  $> 40\%$ .

R-free — контроль переоптимизации. Нужен для того, чтобы модель не была подогнана только под R-фактор.  $(R\_free - R) > 10\%$  настораживают в отношении переоптимизации (overfitting).

R-фактор = 0.17122; R-free = 0.18449

R-фактор  $< 25\%$ ,  $(R\_free - R) < 10\%$  – это хорошие значения. Чтобы понять, как это соотносится со всей базой данных, обратимся к рисунку 3 и посмотрим на перцентиль R-free. По сравнению со всей базой данных это отличное значение параметра, среди структур с похожим разрешением это тоже довольно хороший результат.

### ClashScore

ClashScore — число недопустимых наложений атомов на 1000 (т.е. 5 значит 0.5% атомов).

Для данной структуры равен 3. Значит 0.3% недопустимых наложений атомов. Неплохой результат. По перцентильям находится в области хороших значений.

### Карта Рамачандрана

Карты Рамачандрана отражают оценку конформации остова. Ramachandran outliers — полные маргиналы по карте Рамачандрана, лежат вне допустимой области; такие встречаются в 0,05% остатков при высоком разрешении.

Для проверки карты Рамачандрана использовался сервис MolProbity [3]. В нём были добавлены атомы водорода, детектировались и принимались возможные инверсии боковых цепей аспарагина, глутамина и гистидина.

Таблица 1. Индикаторы качества структуры, определенные сервисом MolProbity [3].

All-Atom	Clashscore, all atoms:	2.91	99 <sup>th</sup> percentile* (N=718, 1.60Å ± 0.25Å)
Contacts	Clashscore is the number of serious steric overlaps (> 0.4 Å) per 1000 atoms.		
Protein Geometry	Poor rotamers	0	0.00% Goal: <0.3%
	Favored rotamers	206	98.10% Goal: >98%
	Ramachandran outliers	0	0.00% Goal: <0.05%
	Ramachandran favored	249	99.60% Goal: >98%
	MolProbity score <sup>^</sup>	1.08	99 <sup>th</sup> percentile (N=7200, 1.60Å ± 0.25Å)
	Cβ deviations >0.25Å	0	0.00% Goal: 0
	Bad bonds:	1 / 2109	0.05% Goal: 0%
	Bad angles:	2 / 2858	0.07% Goal: <0.1%
Peptide Omegas	Cis Prolines:	0 / 8	0.00% Expected: ≤1 per chain, or ≤5%
Additional validations	Chiral volume outliers	0/295	
	Waters with clashes	5/170	2.94% See UnDowser table for details

Видим, что ClashScore 2.91 ~ 3, не изменился. Не присутствуют Poor rotamers — остатки с маргинальными по отклонению от ротамеров боковыми цепями. 98.1% ротамеров находятся в предпочитаемых областях (Favored rotamers). Нет Ramachandran outliers — полные маргиналы по карте Рамачандрана, лежащие вне допустимой области, и опять же почти 100% остатков в предпочитаемой области. MolProbity score = 1.08 — интегральная оценка структуры по данным этого сервиса. Cβ deviations > 0.25E — число Cβ атомов с неприемлемым отклонением от ожидаемого положения, таких не нашлось. Bad bonds — число ковалентных связей, существенно отклоняющихся от теории (типично — 0), но 1 связь необычная. Bad angles — число валентных углов, существенно отклоняющихся от теории (типично — 0), таких 2.

Была получена карта Рамачандрана:

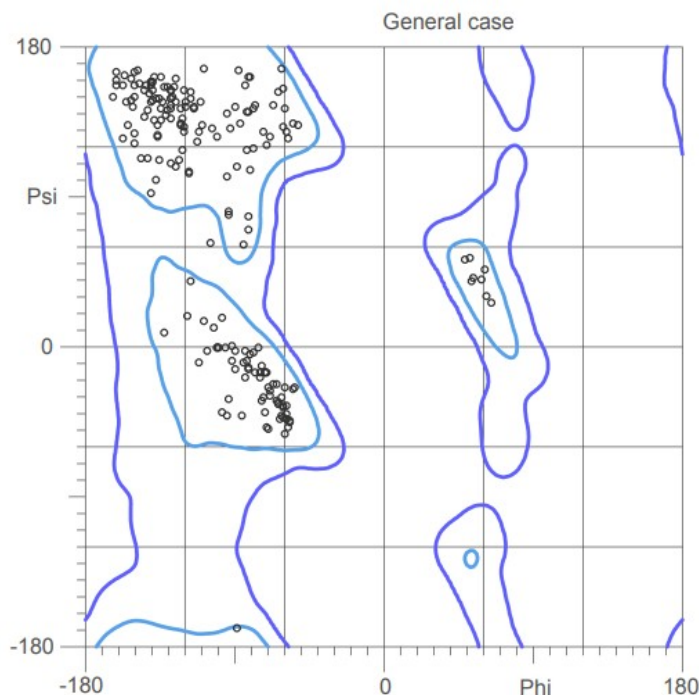


Рис. 4. Карта Рамачандрана. По осям торсионные углы. Области, ограниченные голубым цветом — предпочтительные, синим — допустимые.

Из рисунка 4 можно убедиться, что все остатки находятся в разрешённых областях, поэтому нет полных маргиналов (Ramachandran outliers). 1 остаток не попал в предпочитаемую область, и 1 находится на границе предпочитаемой области.

### Sidechain outliers

Ротамеры — отражают оценку конформации боковых цепей. Маргинальных боковых цепей не было обнаружено.

### RSRZ outliers

Пространственный R-фактор (RSR) характеризует, насколько модель группы атомов (или даже отдельного атома) соответствует “экспериментальной” электронной плотности. RSR можно считать для отдельного остатка, поэтому это один из методов выявления маргинальных остатков. Хорошие



значения: RSR < 10%. Плохие: RSR > 20%. Z-score – относительная оценка RSR. Показывает, насколько RSR для остатка отличается от среднего RSR для такого же типа остатков в структурах PDB со сходным разрешением.

Маргинальным считается остаток, обладающий Z-score > 2. Доля маргинальных остатков по RSR для данной модели составляет 1.2% (3 остатка из 252), это довольно неплохой показатель для PDB.

### 3.3 Таблица со списком нескольких маргинальных остатков

Маргинальных остатков по карте Рамачандрана найдено не было, также у данной структуры нет маргиналов по боковым цепям. В таблице 2 показаны 3 маргинальных остатка по RSRZ.

Таблица 2. Маргинальные остатки по RSRZ [4].

Mol	Chain	Res	Type	RSRZ
1	A	165	THR	3.4
1	A	246	ALA	2.5
1	A	91	THR	2.1

При использовании сервиса Molprobity были найдены неудачные углы (Bad angles) — число ковалентных углов, существенно отклоняющихся от теории (типично — 0), таких 2. В отчёте Full wwPDB X-ray Structure Validation Report [4] их тоже 2, но записи 3 — один из остатков маргинальный вдвойне.

Таблица 3. Маргинальные остатки по углам между атомами [4].

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	242	ARG	NE-CZ-NH2	7.34	123.97	120.30
1	A	242	ARG	NE-CZ-NH1	-5.56	117.52	120.30
1	A	263	TYR	CB-CG-CD2	-5.04	117.98	121.00

Также были найдены атомы, между которыми слишком близкие контакты. Такие были найдены и в белковой цепи, и в лигандах, и между атомами

водорода.

Таблица 4. Атомы с аномально близкими контактами [4].

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
2:A:301:HEM:HBC2	2:A:301:HEM:HMC2	1.65	0.76
1:A:44[A]:GLU:OE2	6:A:401:HOH:O	2.16	0.63
1:A:44[B]:GLU:HG2	6:A:441:HOH:O	2.10	0.51
1:A:207:SER:OG	5:A:308:GOL:C3	2.58	0.51
1:A:207:SER:OG	5:A:308:GOL:H32	2.12	0.49
1:A:195:VAL:HG22	1:A:202[A]:ILE:HG12	1.95	0.49
1:A:45:LYS:HB2	1:A:78:LYS:HE3	1.96	0.47
1:A:218:LYS:NZ	6:A:402:HOH:O	2.33	0.44
1:A:104:ALA:HB1	1:A:106[B]:MET:SD	2.59	0.42
1:A:175:VAL:HG12	1:A:206:LEU:HD21	2.02	0.42
1:A:180:LYS:HE2	1:A:180:LYS:HB3	1.91	0.41
1:A:106[B]:MET:HE1	1:A:109:MET:HE3	2.03	0.40

### 3.4 Анализ 5 маргинальных остатков и гетеромолекул

1. ARG242 (отклоняющиеся от теории значения углов NE-CZ-NH<sub>2</sub>, NE-CZ-NH<sub>1</sub>).

Это углы NE-CZ-N на конце аргинина. Из-за того, что это резонансная структура, то идеальные углы приближены к  $360/3 \sim 120^\circ$ . Но к данной структуре один из углов =  $123.97^\circ$ , а другой  $117.52^\circ$ . Рассмотрим электронную плотность на этих атомах и приближенные к ним атомы.

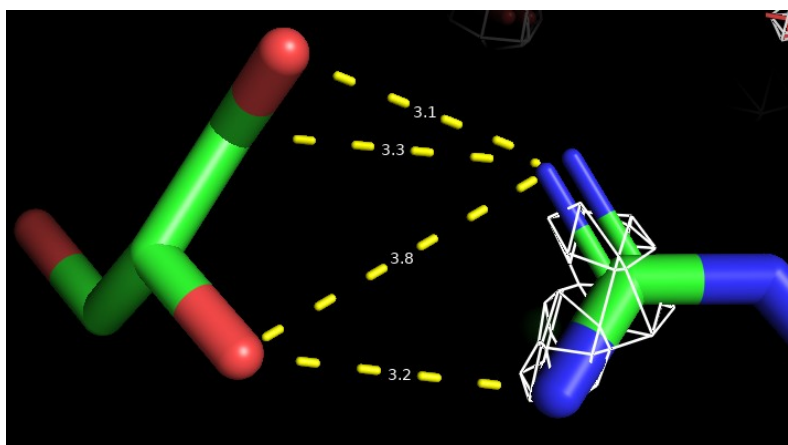


Рис. 5. Взаимодействие остатка аргинина и лиганда глицерина.

На рисунке 5 показана электронная плотность на аргинине с уровнем подрезки 3.5. Был выбран этот уровень подрезки, так как на нём видно, что на атоме NH2 (нижний на рисунке), угол NE-CZ-N с которым равен  $123.97^\circ$ , электронная плотность выше. Нехватка электронной плотности на атоме NH1 (верхний азот) объясняется более сильным влиянием на него лиганда глицерина и его электроотрицательных кислородов. Смещение в углах также соответствует логике: на нижнем атоме (NH2) электронная плотность высокая и он сильнее отталкивается от атома NE, у верхнего атома (NH1) она ниже, и поэтому он чуть ближе к атому NE и угол  $NE-CZ-N1 = 117,52^\circ$ .

2. TYR263 (отклоняющийся от теории угол CB-CG-CD2).

Вместо  $121^\circ$ , этот угол равен  $117.98^\circ$ . Углы отображены на рисунке 6.

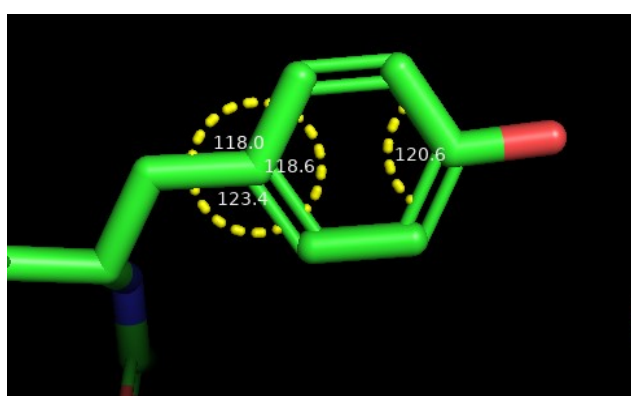


Рис. 6. Углы в маргинальном тирозине.

После визуализации электронной плотности на этом остатке (рис. 7) и при том, что углеродные атомы ароматического кольца не взаимодействуют ни с какими атомами на расстоянии меньше 3.4 ангстрем, кажется, что это

ошибка интерпретации данных.

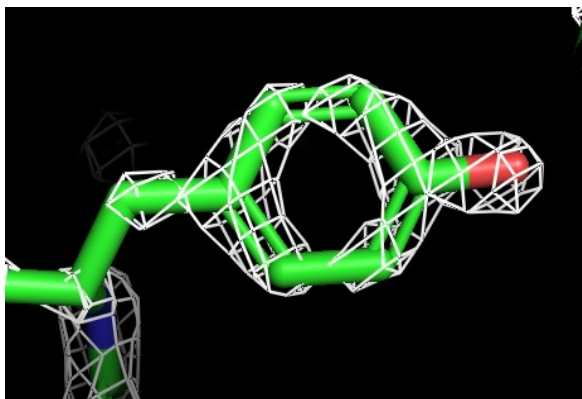


Рис. 7. Электронная плотность на тирозине, уровень подрезки 3.7.

3. THR165 (маргинальный по RSRZ).

Из сравнения ЭП на треонине при разных уровнях подрезки (рис 7, 8, 9) видно, что его реальная электронная плотность была неудачно посчитана экспериментально, возможно, это из-за его положения на поверхности глобулы (рис. 10). Из-за неудачных экспериментальных данных для этого остатка и получилось, что он маргинален по RSRZ.

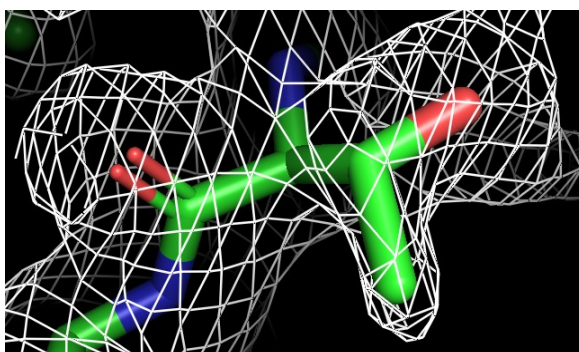


Рис. 8. ЭП на треонине, уровень подрезки 0.5.

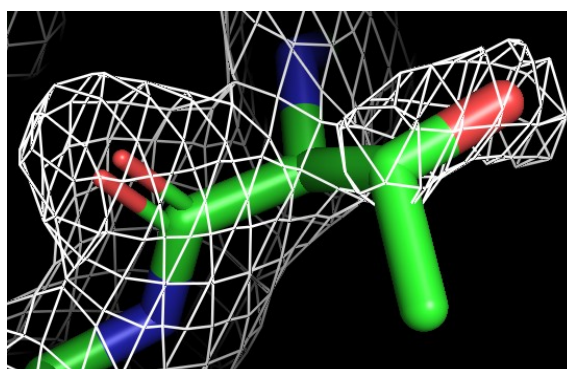


Рис. 9. ЭП на треонине, уровень подрезки 1.

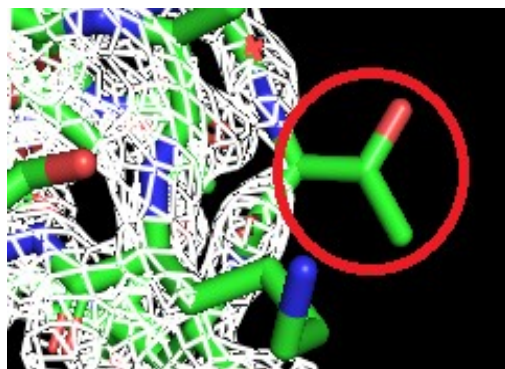


Рис. 10. ЭП на белке, уровень подрезки 2. Поверхностное положение треонина (обведён красным).

4. SER207 (слишком близкое расположение атома 1:A:207:SER:OG с атомами лиганда 5:A:308:GOL:C3 и с 5:A:308:GOL:H32)

В нём 2 взаимодействия 1:A:207:SER:OG с 5:A:308:GOL:C3 1:A:207:SER:OG и 5:A:308:GOL:H32 имеют длину 2.58 Е и 2.12 Е соответственно и имеют перекрытия 0.51 Е и 0.49 Е соответственно. Это взаимодействие лиганда глицерина, в частности одного углерода и водорода на нём (в модели не показан) и кислорода на боковом остатке серина.

На уровне подрезки 1 ЭП глицерина была видна, опять же, это взаимодействие на поверхности белка. В PyMol была показана поверхность белка, а интересующие нас атомы отображены в шаровой модели (рис. 12). Видно, что углерод глицерина проходит через поверхность белка, поэтому можно считать, что это ошибка интерпетации эксперимента и этот углеродный атом глицерина должен находиться дальше.

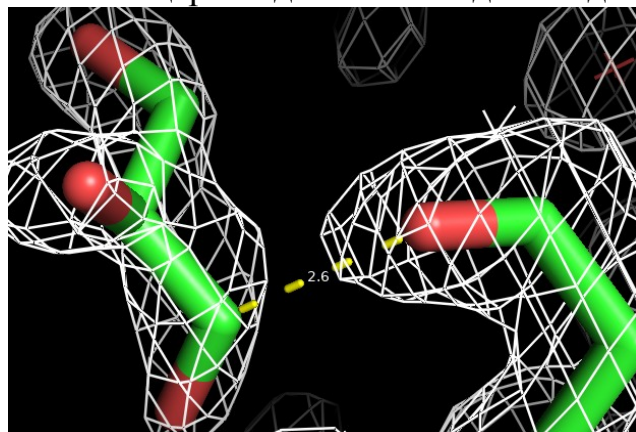


Рис. 11. ЭП, уровень подрезки 1. Взаимодействие серина и глицерина.

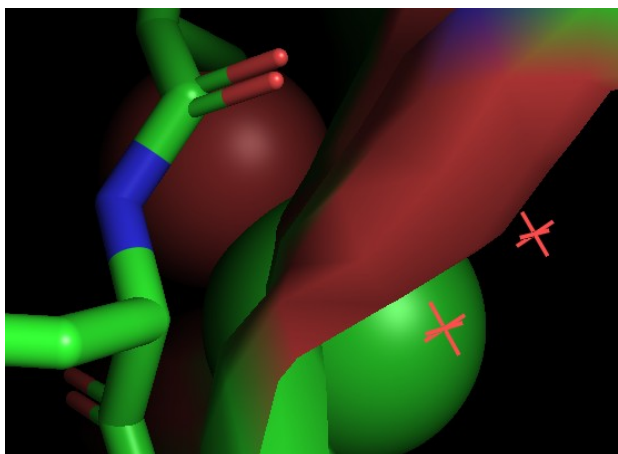


Рис. 12. Поверхность белка, которую «пронзает» углерод глицерина.

5. Лиганд НЕМ301 (внутри лиганда слишком близкое расположение атомов 2:A:301:HEM:HVC2 и 2:A:301:HEM:HMC2)

Произошло перекрытие между водородами гема. Расстояние между ними 1.65 Å, перекрытие 0.76 Å. Вероятно, это ошибка алгоритма восстановления водородов. Сближенные водороды на геме можно увидеть на рис. 13. Модель с водородами была получена с помощью сервиса Molprobity [3].

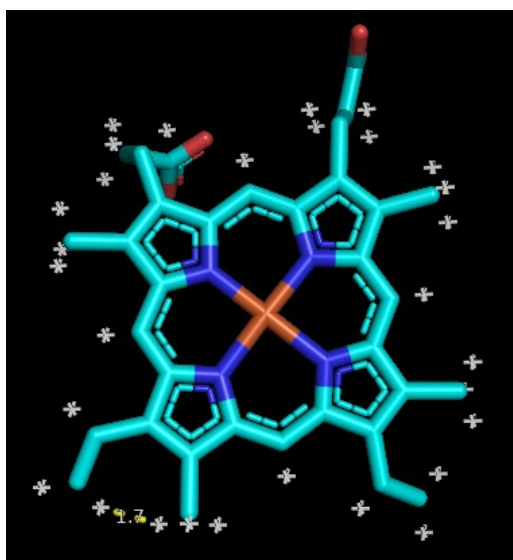


Рис. 13. Гем. Показана короткая связь между водородами.

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

База данных PDB-REDO содержит оптимизированные версии существующих записей PDB с картами электронной плотности, описанием

изменений модели и множеством данных проверки модели. Это хорошая отправная точка для любого проекта структурной биологии.

Все записи обрабатываются с помощью согласованного протокола, который уменьшает влияние различий в давности работы, программного обеспечения и людей, которые работают с данными [5].

В сравнении измененной и начальной версии PDB 6om5 улучшились многие показатели качества модели: R-factor, R-free, отображение на карте Рамачандрана, нормальность ротамеров, оптимизированы некоторые водородные связи и др. Перечисленные улучшения показаны в таблице 5. Количественно они невелики, потому что модель изначально была очень хорошая.

Таблица 5. Сравнение PDB и PDB-redo индикаторов качества модели.

<b>Validation metrics from PDB-REDO</b>		
	<b>PDB</b>	<b>PDB-REDO</b>
<b>Crystallographic refinement</b>		
<i>R</i>	0,1711	0,1275
<i>R-free</i>	0,1843	0,1455
<i>Bond length RMS Z-score</i>	0,412	0,680
<i>Bond angle RMS Z-score</i>	0,838	0,799
<b>Model quality (raw scores   percentiles)</b>		
<i>Ramachandran plot appearance</i>	83	86
<i>Rotamer normality</i>	90	97
<i>Coarse packing</i>	51	52
<i>Fine packing</i>	50	51
<i>Bump severity</i>	82	87
<i>Hydrogen bond satisfaction</i>	44	67

Были сделаны следующие изменения в модели: удалено 4 молекулы воды, улучшен фиттинг 113 а. о., 2 ротамера изменены. Изменения в остове совсем незначительны, их можно увидеть на обновлённой карте Рамачандрана (рис. 14).

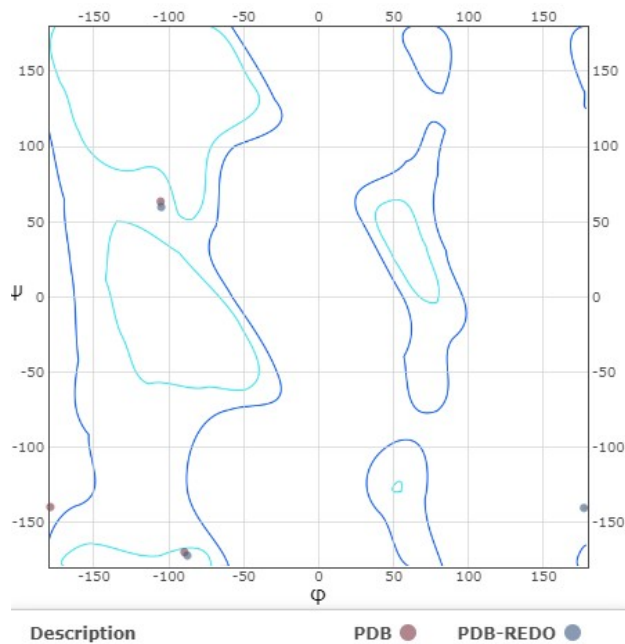


Рис. 14. Отличие карт Рамачандрана для записей PDB и PDB-redo.

#### 4. Заключение

Структура гемофилина бактерии *Haemophilus haemolyticus* сделана качественно: у неё высокое разрешение и модель, не противоречащая теоретическим расчётам. Значения индикаторов качества высокие для структур PDB в общем и для структур с аналогичным разрешением. Некоторые индикаторы отличные — нет остатков в запрещённой области карты Рамачандрана, их боковые цепи также в разрешённых областях. Некоторые маргинальные остатки могут быть оправданы с биологической стороны, другие связаны с положением лигандов или неправильным восстановлением водородов сторонними алгоритмами. Мною не найдено объяснение для маргинальности остатка TYR263 в связи нестандартными углами между атомами в его боковой цепи. Улучшение структуры сервисом PDB\_redo незначительно.



## 5. Список литературы

1. Latham RD, Torrado M, Atto B, Walshe JL, Wilson R, Guss JM, Mackay JP, Tristram S, Gell DA. A heme-binding protein produced by *Haemophilus haemolyticus* inhibits non-typeable *Haemophilus influenzae*. // *Mol Microbiol.* . 2019 Nov 19.
2. <https://www.rcsb.org/pdb/explore/materialsAndMethods.do?structureId=6OM5>
3. <http://molprobit.biochem.duke.edu/>
4. Full wwPDB X-ray Structure Validation Report. [https://files.rcsb.org/pub/pdb/validation\\_reports/om/6om5/6om5\\_full\\_validation.pdf](https://files.rcsb.org/pub/pdb/validation_reports/om/6om5/6om5_full_validation.pdf)
5. PDB-redo <https://pdb-redo.eu/db/6om5>