

## Введение

Мультидоменные белки адгезии широко распространены среди прокариот и эукариот. Обычно один из доменов принимает участие в связывании целевой молекулы, а другой встроен в мембрану клетки и/или органеллы адгезии. Взаимодействие с закоривающим доменом могут модулировать связывание лиганда. Некоторые из них могут работать по принципу “catch bonds”, т.е. усиливать силу связывания при механическом растяжении. Такие связи более устойчивы к водорастворимым ингибиторам (Yakovenko et al. 2008).

Одним из таких белков является комплекс кишечной палочки FimH, который помогает бактериям удерживаться на различных эпителиях, а также неорганических поверхностях. Установление структуры данного комплекса позволило уточнить механизм работы данного комплекса и показало, что он действительно функционирует по механизму “catch bonds”.

PDB ID	3jwn
--------	------

Данная расшифровка сделана для манного-специфичного адгезина *E. coli* FimH, для которого ранее предполагался “catch bonds” механизм. FimH состоит из двух иммуноглобулино-подобных доменов: N-концевой лектиновый домен связывает маннозу, а вот C-концевой домен закоривает белок в фимбриальном участке. В экспериментах взаимодействие этих двух доменов показано не было, т.к. ранее были получены структуры только для отдельных доменов. В данной работе были получены модель комплекса в нативной конформации, что позволило подтвердить первоначальные предположения о механизме его функционирования (Le Trong et al. 2010).

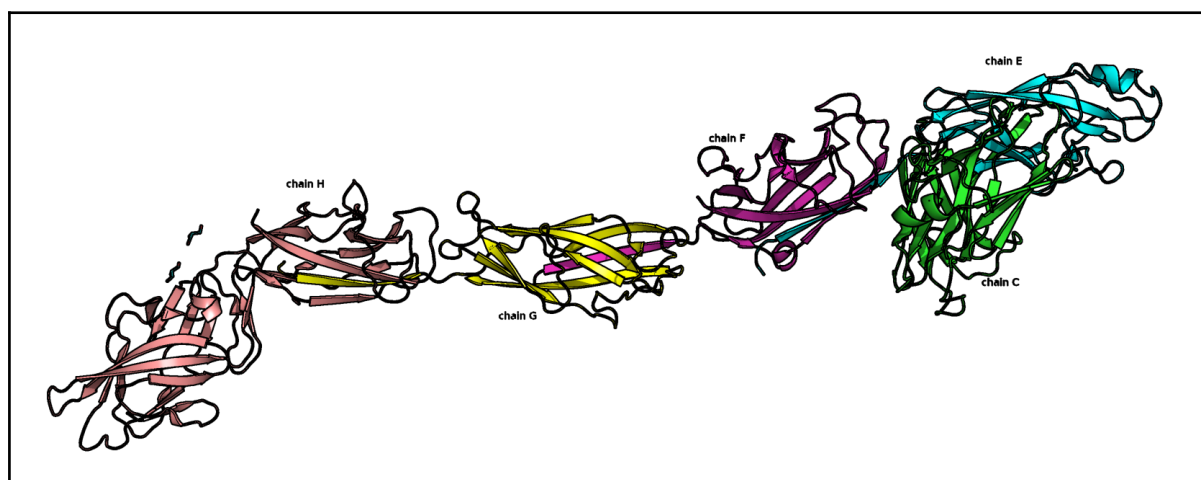


Рисунок 1. Общий вид комплекса.  
Цепь C: белок-шаперон для FimH; цепи E, F: FimF, цепь G: FimG, цепь H: FimH.

Посмотрим на основные характеристики PCA-расшифровки:

Группа симметрии	H 3 2
Параметры ячейки: a, b, c, $\alpha$ , $\beta$ , $\gamma$	216.03Å, 216.03Å, 532.07Å 90.00°, 90.00°, 120.00°
Метод решения фазовой проблемы	Молекулярное замещение
Число рефлексов	119926
Разрешение	48.34Å -- 2.69Å
Полнота данных	90.5% для разрешения 2.69Å

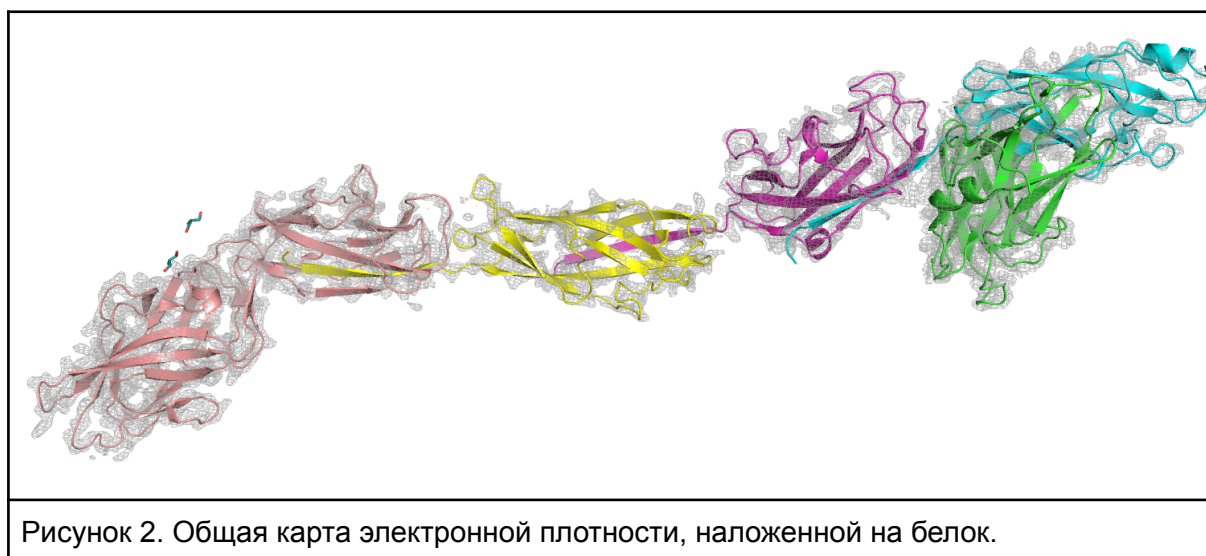
### Метрики качества структуры в целом

В качестве метрик оценки качества структуры в целом воспользуемся величинами R-фактора и  $R_{free}$ .

R-фактор	0.244
$R_{free}$	0.272

Модель имеет удовлетворительное разрешение, близкое к хорошему (2.69Å). Следуя правилу для определения граничного значения для хорошего R-фактора, “хороший” R-фактор должен быть меньше, чем 0.269. Реальное значение R-фактора находится на границе “хорошего” и составляет 0.244. Значение  $R_{free}$  немного выходит за установленные границы, однако разница между ними меньше 10% (в нашем случае -- 2.8%), что позволяет сказать, что структура не переоптимизирована.

Стоит отметить, что, в целом, белок покрыт электронной плотностью и крупные участки, где бы она отсутствовала, не наблюдаются.



Перед тем, как перейти к анализу отдельных маргинальных остатков, посмотрим на общее качество модели подробнее. Рассмотрим метрики качества, которые выдает нам сервис MolProbity до добавления водородов к структуре и инверсии остатков глутамата/аспартата.

All-Atom Contacts	Clashscore, all atoms:	6.59		99 <sup>th</sup> percentile* (N=185, 2.69Å ± 0.25Å)
Protein Geometry	Poor rotamers	67	4.41%	Goal: <0.3%
	Favored rotamers	1358	89.40%	Goal: >98%
	Ramachandran outliers	11	0.60%	Goal: <0.05%
	Ramachandran favored	1755	95.64%	Goal: >98%
	Rama distribution Z-score	-1.50 ± 0.17		Goal: abs(Z score) < 2
	MolProbity score <sup>^</sup>	2.15		96 <sup>th</sup> percentile* (N=5375, 2.69Å ± 0.25Å)
	Cβ deviations >0.25Å	0	0.00%	Goal: 0
	Bad bonds:	1 / 14137	0.01%	Goal: 0%
	Bad angles:	1 / 19358	0.01%	Goal: <0.1%
Peptide Omegas	Cis Prolines:	14 / 98	14.29%	Expected: ≤1 per chain, or ≤5%
	Cis nonProlines:	0 / 1751	0.00%	Goal: <0.05%
	Twisted Peptides:	0 / 1849	0.00%	Goal: 0

При рассмотрении выдачи сервиса мы видим, что некоторые из основных метрик качества достаточно сильно отклоняются от целевых значений: это число маргинальных и “хороших” ротамеров, а также остатков, положение которых сильно отличается от оптимального на карте Рамачандрана. С помощью сервиса добавим водороды и посмотрим на возможные инверсии остатков глутамата/аспартата.

Цепь	Номер остатка	Тип остатка	Оригинальное значение	После изменения	Разница
C	73	ASN	-0.59	-0.016	0.574
F	77	ASN	-5	-4.4	0.6
F	102	GLN	-5	-4.4	0.6
F	135	GLN	-0.55	0.099	0.649
K	94	GLN	-0.24	0.69	0.93
K	102	GLN	-0.24	0.69	0.93

Мы наблюдаем целых шесть остатков, которые, вероятно, были вписаны в ЭП неверно (т.е. с инверсией). Примем предлагаемые изменения.

Мы видим неплохой Clashscore для всех атомов структуры, но более никакие метрики качества не улучшились. Стоит отметить, что чаще всего clash возникают для автоматически добавленных атомов водорода, поэтому значения данного параметра не превышает критическое.

Так как мы наблюдаем достаточно много аутлайеров среди аминокислотных остатков (как для остовных атомов по картам Рамачандрана, так и для атомов боковой цепи по ротамерам), рассмотрим несколько выпадающих остатков и попытаемся выяснить, почему они имеют такое плохое качество.

Маргинальные остатки по карте Рамачандрана		
Цепь, номер и тип	$\varphi$	$\psi$
E, Val-27	52.6	-55.8
E, Val-60	-53.6	106.1
<b>H, Tyr-175</b>	<b>69.9</b>	<b>135.9</b>
H, Asn-228	58.0	-61.0
<b>H, Gly-229</b>	<b>176.3</b>	<b>50.3</b>
K, Arg-12	70.9	134.7
<b>K, Val-27</b>	<b>36.6</b>	<b>-77.1</b>
K, Thr-44	167.6	146.2
N, Thr-169	67.5	122.4
N, Ala-218	-62.7	-172.5

N, Gly-261

-63.5

-103.7

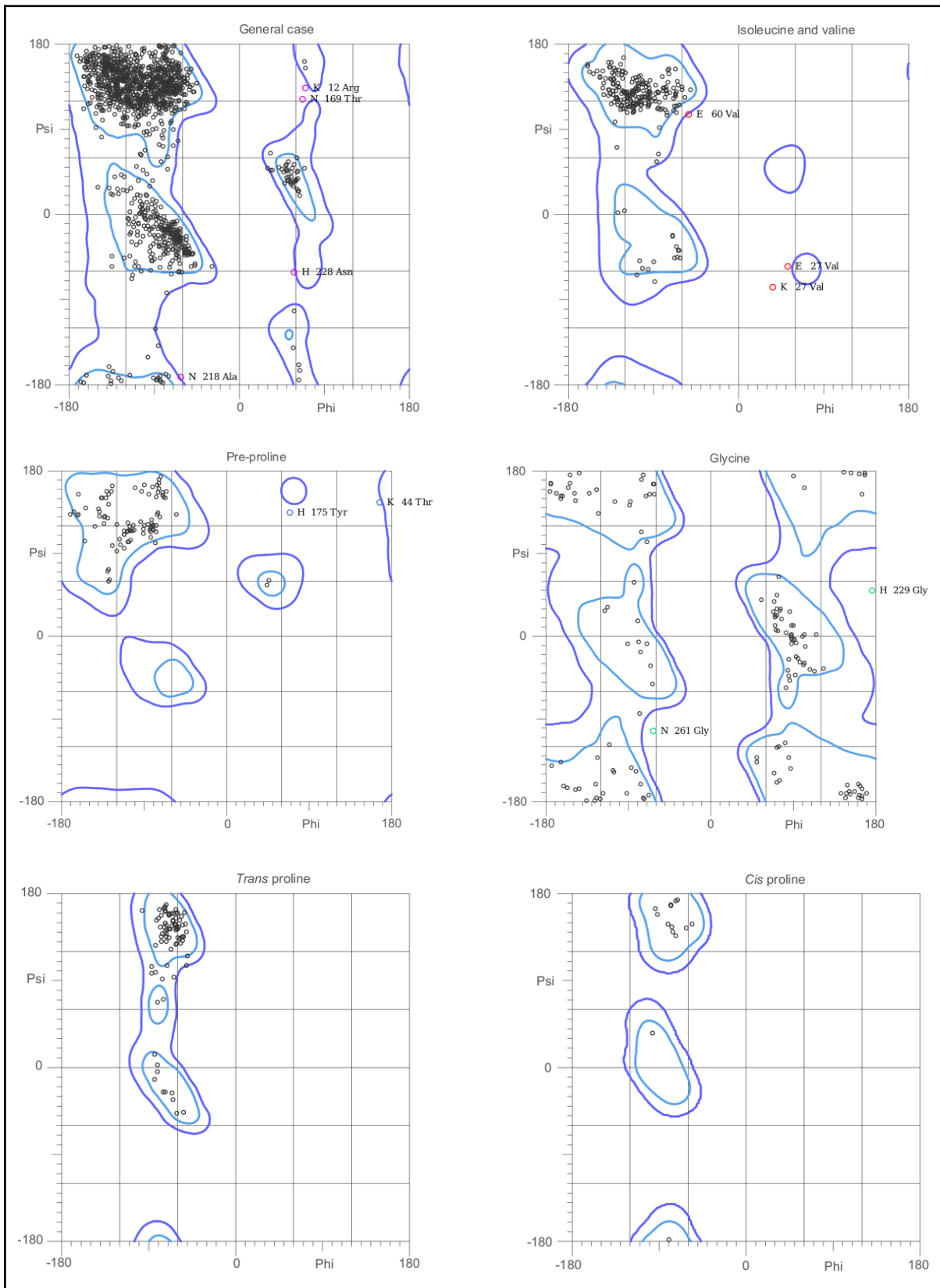
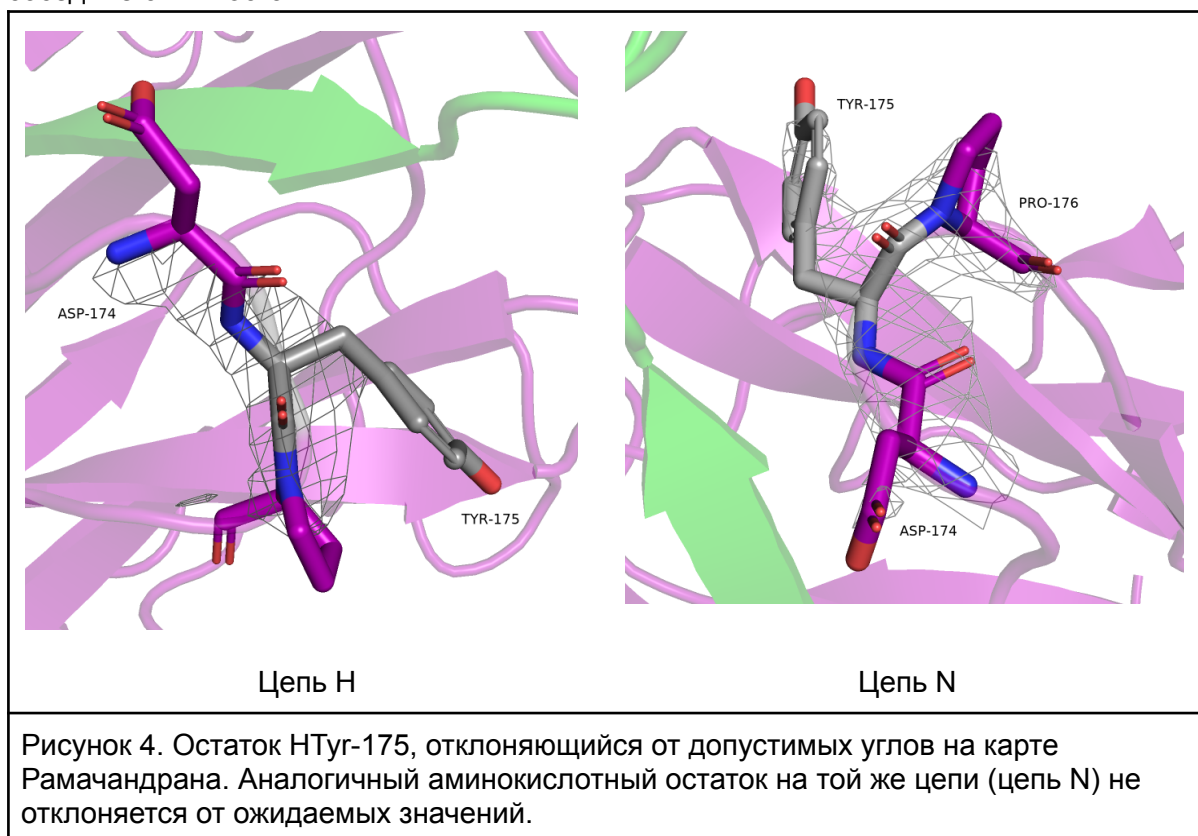


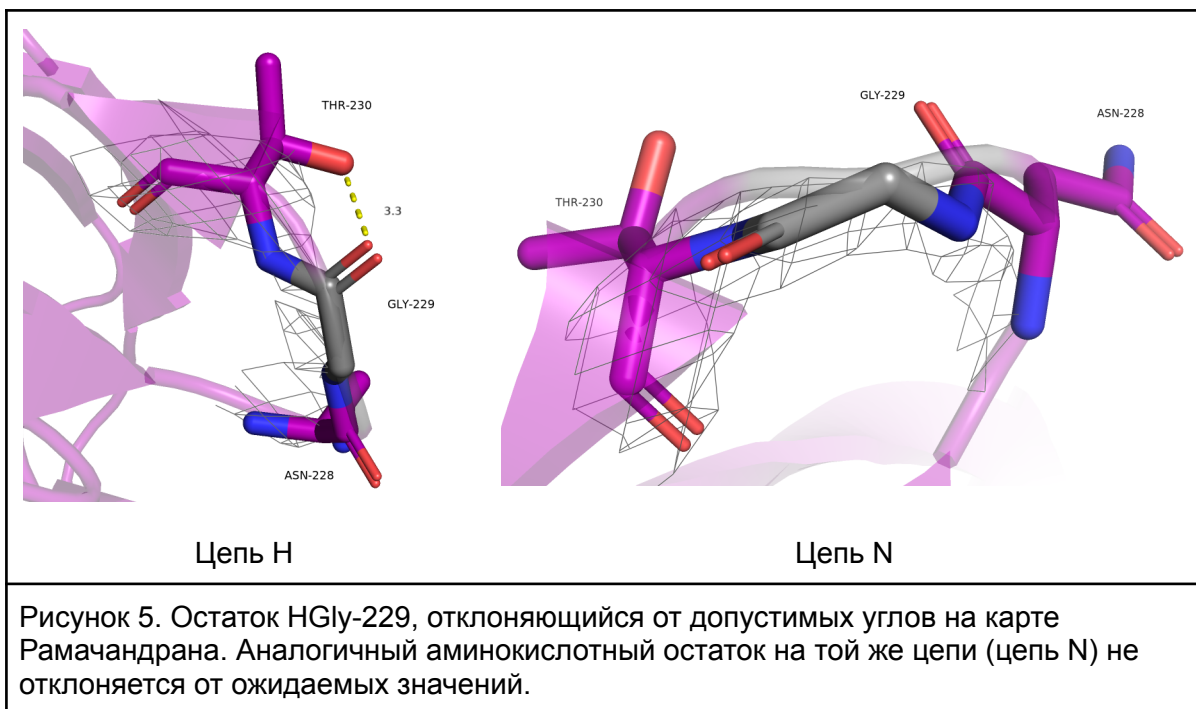
Рисунок 3. Карты Рамачандрана для всех аминокислотных остатков структуры.

Для дальнейшего анализа выберем 4 остатка, наиболее отклоняющихся от “позитивных” областей карты, такие остатки были выделены в таблице красным цветом.

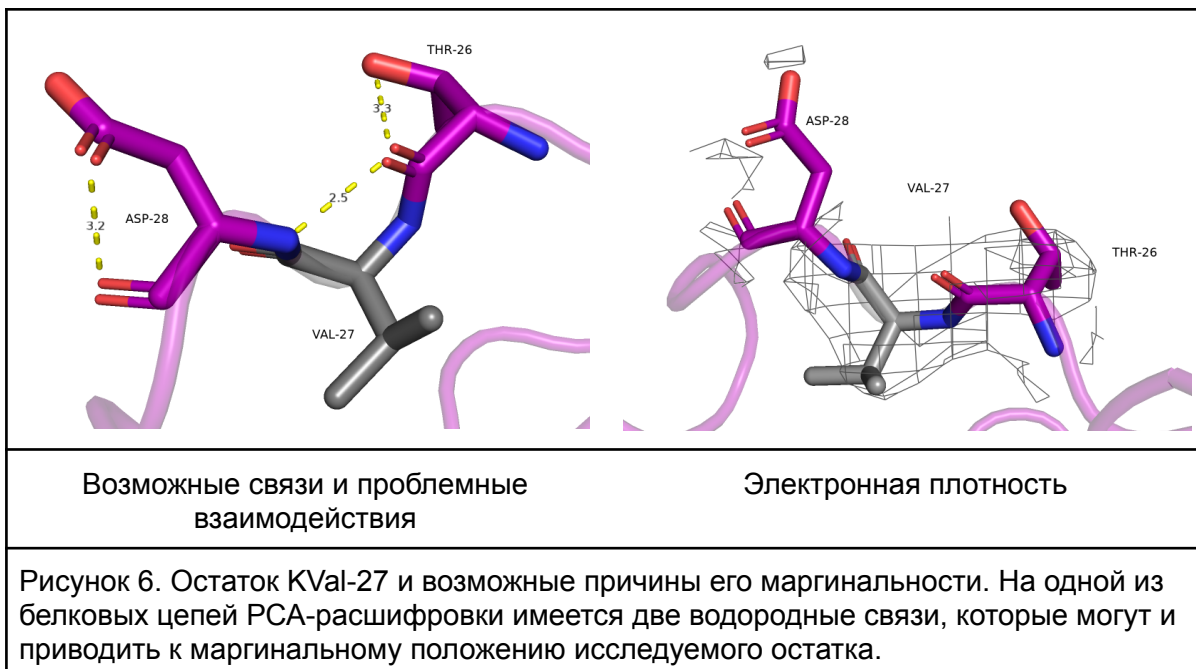
Начнем с остатка NTyr-175, который стоит перед пролином в аминокислотной последовательности. При наложении электронной плотности на этот остаток и на соседние с ним (слева и справа) мы видим, что данный участок довольно плохо разрешен: ЭП покрывает только часть остова соседних с данным остатков, что и может в данном случае объяснить его маргинальность. Стоит отметить, что аналогичный остаток другой цепи (цепь N) маргинальным не является, а ЭП хорошо покрывает соседние с ним остатки.



Рассмотрим остаток NGly-229. В данном случае мы также видим не очень хорошую электронную плотность на данном остатке, однако маргинальность данного остатка может поддерживаться водородной связью между остатками треонина. Если посмотреть на аналогичную цепь, электронная плотность на ней также не очень хорошего качества, однако взаимодействие между остатками отсутствует, что, видимо, и позволяет глицину принять более комфортную для него конформацию.



Рассмотрим остаток KVal-27. Вокруг данного остатка также не имеется хорошей электронной плотности, которая достоверно могла бы подтвердить его положение. Однако данный остаток может быть связан водородной связью с кислородом остова остатка Thr-26, что могло бы объяснить его маргинальность при хорошем наложении ЭП. В данном случае однозначный вывод сделать нельзя.



Рассмотри маргинальные остатки по ротамерным углам боковых групп. Таких в структуре оказалось 67 штук. Рассмотрим один из них.

Некоторые маргинальные остатки по ротамерам боковых групп	
C, Arg8	chi angles: 72.6,254.4,292.5,105.2
C, Leu32	chi angles: 246.4,160.4
C, Lys44	chi angles: 295.2,247.3,292.1,253.3
C, Lys58	chi angles: 297.1,127.8,101.4,304.6
C, Lys61	chi angles: 252.5,69.5,154,293
C, Arg79	chi angles: 93.5,286.8,286.3,185.7
E, Asp13	chi angles: 275.4,71.9
E, Asn14	chi angles: 224.9,80.9
E, Leu29	chi angles: 47.6,158.6
E, Thr84	chi angles: 93.1
F, Asp2	chi angles: 20,294.5
F, Ser86	chi angles: 330.5

Боковая цепь данного аминокислотного остатка вообще не покрыта электронной плотностью, поэтому такие углы в боковой цепи вообще ничем нельзя объяснить и такое положение остатка -- ошибка расшифровки.

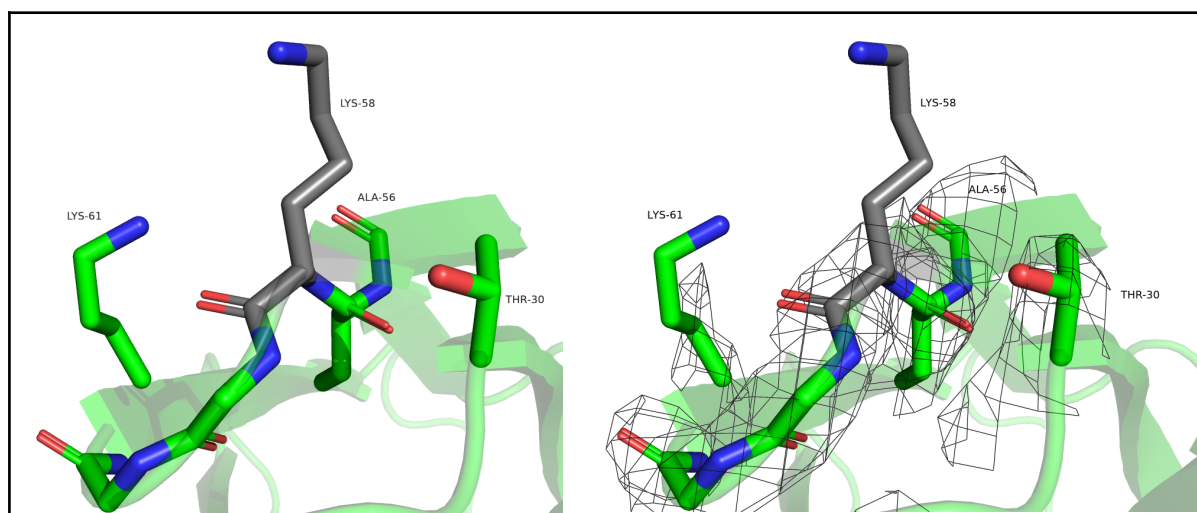


Рисунок 7. Положение остатка Lys58 на цепи C. Ротамерные углы боковой цепи не поддерживаются картой электронной плотности, поэтому данное положение боковой цепи лизина является ошибкой расшифровки.

Напоследок рассмотрим один из остатков, который MolPrody посчитал инвертированным. Действительно, инверсия оправдана: если сделать ее, то между остатком глутамата и  $\pi$ -системой гистидина возможно образование водородной связи.



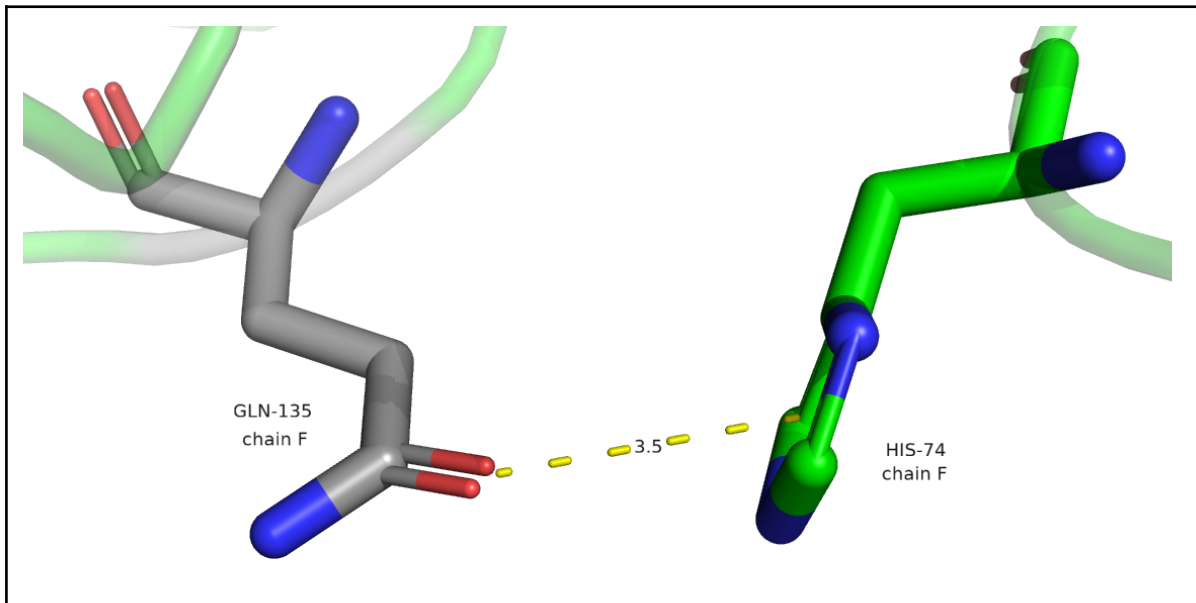


Рисунок 8. Возможное местоположение водородной связи для остатка FQln135, что может указывать на необходимую ему инверсию.

Покрасим все маргинальные остатки на структуре и посмотрим, как они располагаются в ней. Стоит отметить, что большинство маргинальных остатков экспонированы на поверхность белка, а многие вообще располагаются в петлях, которые часто в среднем хуже покрыты электронной плотностью. Скорее всего для большинства остатков маргинальность не подкрепляется электронной плотностью; положение данных остатков не несет функциональной важности и является ошибкой расшифровки.

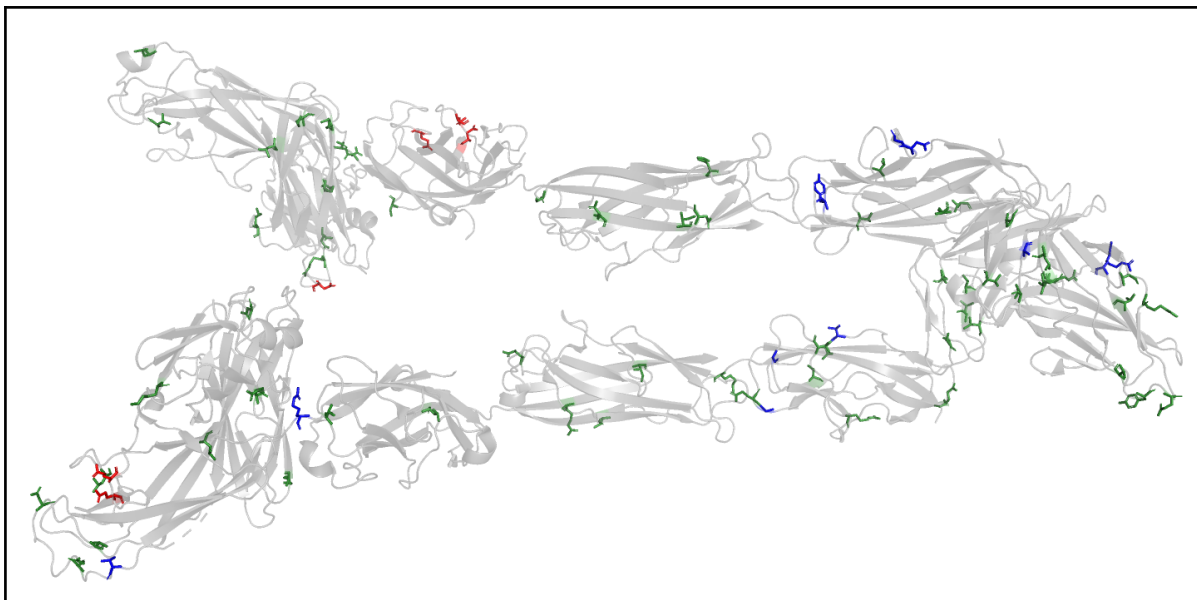
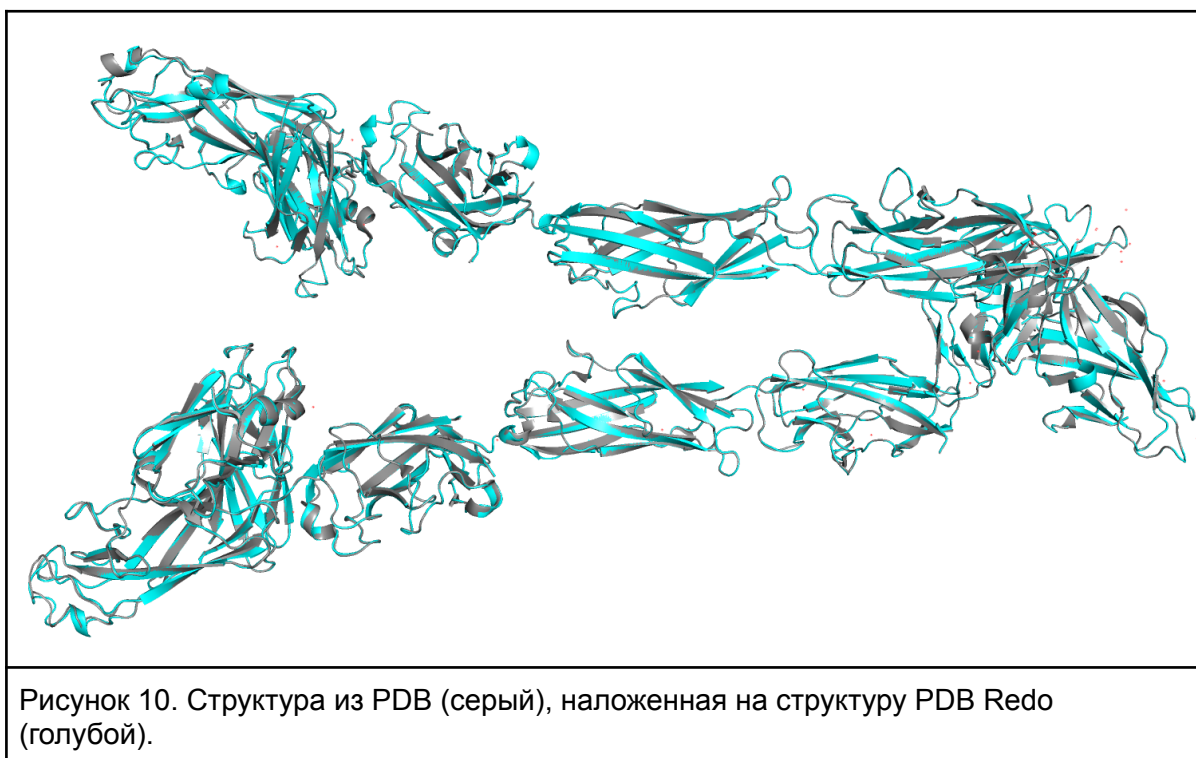


Рисунок 9. Расположение маргинальных аминокислотных остатков в PCA-структуре. Красным показаны остатки с инверсией групп аспарагина/глутамина, синим -- остатки, не вписывающиеся в карту Рамачандрана, зеленым -- остатки, чьи боковые группы плохо вписываются в значения ротационных углов.

В целом, стоит отметить, что основные метрики качества для данной структуры находятся в пределах допустимых значений, однако метрики качества отдельных остатков выбиваются из допустимых значений. Маргинальность исследованных остатков не поддерживается электронной плотностью и можно предположить, что маргинальность большей части остатков, которые не были проанализированы, связана не с их функциональной значимостью, а с ошибками построения модели. При должном анализе функционально значимых участков на предмет маргинальности данной моделью можно пользоваться.

Однако модель, очевидно, можно улучшить, как минимум исправив положения ротамеров боковых групп аминокислот. Проанализируем улучшенную модель с сайта PDB Redo.

На макроуровне структура, как и ожидалось, практически не меняется. Мы видим незначительное изменение положения для некоторых петель



Посмотрим, как изменились метрики качества модели.

<i>R</i>	0,2446	0,2092
<i>R-free</i>	0,2935	0,2270
<i>Bond length RMS Z-score</i>	0,298	0,493
<i>Bond angle RMS Z-score</i>	0,460	0,851
<i>Ramachandran plot appearance</i>	47	74
<i>Rotamer normality</i>	39	26

Мы видим достаточно значительное уменьшение величины R-фактора и  $R_{free}$ , что позволяет сказать, что структура в целом стала лучше вписана в электронную плотность. Также стоит отметить, что мы видим улучшение RMS Z-score, а также увеличение количества остатков, попадающих в благоприятную зону карты Рамачандрана. Но одновременно с этим мы видим ухудшение числа “нормальных” ротамеров.

All-Atom Contacts	Clashscore, all atoms:	5.88		100 <sup>th</sup> percentile* (N=185, 2.69Å ± 0.25Å)
	Clashscore is the number of serious steric overlaps (> 0.4 Å) per 1000 atoms.			
Protein Geometry	Poor rotamers	135	8.85%	Goal: <0.3%
	Favored rotamers	1281	84.00%	Goal: >98%
	Ramachandran outliers	17	0.92%	Goal: <0.05%
	Ramachandran favored	1780	96.58%	Goal: >98%
	Rama distribution Z-score	-0.58 ± 0.18		Goal: abs(Z score) < 2
	MolProbity score <sup>^</sup>	2.26		94 <sup>th</sup> percentile* (N=5375, 2.69Å ± 0.25Å)
	Cβ deviations >0.25Å	10	0.58%	Goal: 0
	Bad bonds:	0 / 14187	0.00%	Goal: 0%
	Bad angles:	40 / 19427	0.21%	Goal: <0.1%
Peptide Omegas	Cis Prolines:	14 / 98	14.29%	Expected: ≤1 per chain, or ≤5%

В целом, я бы отметила, что мы не видим значительного улучшения качества модели в целом: количество маргиналов по картам Рамачандрана уменьшилось, однако, напротив, увеличилось количество маргиналов по углам ротамеров. Появилось много плохих углов, хотя в исходной структуре это не наблюдалось.

Также стоит отметить, что, видимо, в структуре PDB Redo сделано множество ненужных инверсий групп остатков аспарагина и глутамина: число остатков, которые предлагает развернуть Mol Prody сильно увеличивается.

Chain	Res#	Alt	Res ID	Orig	Flip	Flip-Orig	Code	Explanation
C	15		GLN	-0.23	1.3	1.53	FLIP	Some evidence recommending flip.
C	101		ASN	-3.2	-2.1	1.1	CLS-FL	Both orientations clash but flip was preferred.
C	104		GLN	-1.7	-0.11	1.59	FLIP	Some evidence recommending flip.
C	164		ASN	-9.3	-4.5	4.8	FLIP	Clear evidence for flip.
E	14		ASN	-1.5	-0.26	1.24	FLIP	Some evidence recommending flip.
E	74		HIS	-3.3	-2.1	1.2	FLIP	Some evidence recommending flip.
E	77		ASN	-0.86	-0.019	0.841	FLIP	Some evidence recommending flip.
E	100		GLN	-2.4	-1.6	0.8	CLS-FL	Both orientations clash but flip was preferred.
E	102		GLN	-3.4	1.1	4.5	FLIP	Clear evidence for flip.
E	135		GLN	-3.4	-2.1	1.3	CLS-FL	Both orientations clash but flip was preferred.
E	142		HIS	-4.7	-2.1	2.6	FLIP	Clear evidence for flip.
E	154		GLN	-12	-4.5	7.5	FLIP	Clear evidence for flip.
F	24		ASN	-2.9	-1.2	1.7	FLIP	Some evidence recommending flip.
F	74		HIS	-2.6	-1.4	1.2	FLIP	Some evidence recommending flip.
F	101		ASN	-1.2	-0.16	1.04	FLIP	Some evidence recommending flip.
G	105		GLN	-1.1	0.32	1.42	FLIP	Some evidence recommending flip.
G	134		GLN	-2	-0.028	1.972	FLIP	Some evidence recommending flip.
H	59	B	GLN	-11	-4.1	6.9	FLIP	Clear evidence for flip.
H	78		ASN	-2.3	-1.2	1.1	FLIP	Some evidence recommending flip.
H	143		GLN	-7.8	-4.8	3	CLS-FL	Both orientations clash but flip was preferred.
H	147		ASN	-3.7	-1.3	2.4	FLIP	Clear evidence for flip.
H	206		ASN	-3.8	-2.5	1.3	CLS-FL	Both orientations clash but flip was preferred.
H	236		ASN	-0.85	-0.3	0.55	FLIP	Some evidence recommending flip.
H	255		ASN	-1.8	-1.3	0.5	FLIP	Some evidence recommending flip.
H	279		GLN	-0.42	1.1	1.52	FLIP	Some evidence recommending flip.
I	74		GLN	-1.7	-0.051	1.649	FLIP	Some evidence recommending flip.
K	32		ASN	-2.5	-1.1	1.4	FLIP	Some evidence recommending flip.
K	74		HIS	-2.4	-0.91	1.49	FLIP	Some evidence recommending flip.
L	24		ASN	-1.6	-0.07	1.53	FLIP	Some evidence recommending flip.
L	58		ASN	-0.75	0.058	0.808	FLIP	Some evidence recommending flip.
L	74		HIS	-2.3	-1	1.3	FLIP	Some evidence recommending flip.
L	77		ASN	-0.83	-0.051	0.779	FLIP	Some evidence recommending flip.
L	135		GLN	-2.9	-1	1.9	FLIP	Some evidence recommending flip.
M	8		ASN	-0.98	-0.28	0.7	FLIP	Some evidence recommending flip.
M	105		GLN	-0.93	-0.21	0.72	FLIP	Some evidence recommending flip.
M	134		GLN	-3	-0.41	2.59	FLIP	Clear evidence for flip.
N	191		GLN	-2.1	-0.24	1.86	FLIP	Some evidence recommending flip.
N	206		ASN	-1.1	-0.016	1.084	FLIP	Some evidence recommending flip.
N	236		ASN	-0.79	-0.25	0.54	FLIP	Some evidence recommending flip.
N	262		GLN	-0.92	-0.18	0.74	FLIP	Some evidence recommending flip.

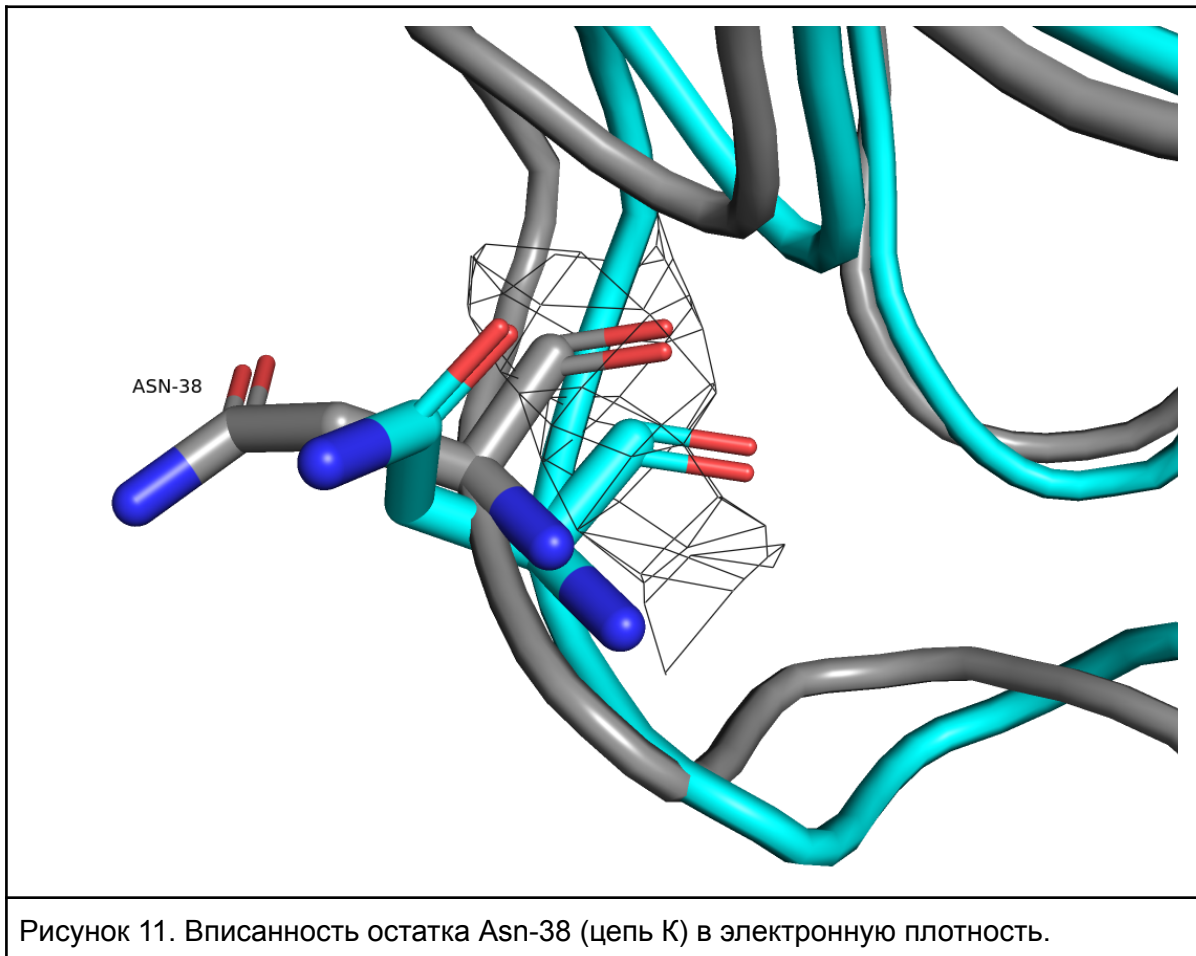
Посмотрим на остатки, для которых PDB Redo максимально увеличил и ухудшил качество. Рассмотрим наиболее улучшенный остаток -- Asn 38 (цепь K). Данный остаток не является маргинальным по анализу MolPrody: да, он располагается на границе нормы карты Рамачандрана и углам ротамеров, но данный остаток не требовал инверсии и вообще в целом не выглядел подозрительно.

K 38	ASN	104.35	-	Favored (3.19%) General / -168.2,-177.2	Allowed (1.8%) <i>t</i> 0 chi angles: 214.6,300.4
------	-----	--------	---	--	--

Загрузим структуру из PDB Redo в MolPrody и посмотрим на него снова.

K 38	ASN	85.51	-	Allowed (0.27%) General / -135.5,-147.6	Favored (19.7%) <i>p</i> 0 chi angles: 51.2,47.6
------	-----	-------	---	--	---

Мы видим для данного остатка ухудшение значений его углов остова, но улучшение ротамеров его боковой группы. Т.к. улучшение ротамеров значительное, а вот ухудшение расположения углов на карте Рамачандрана незначительно, это может объяснять такое улучшение качества. Электронная плотность для данного остатка достаточно плохая, поэтому объяснить такое улучшение лучшей вписанностью нельзя.



Рассмотрим остаток, для которого метрики максимально ухудшились -- Met-30 (цепь E). Данный остаток не является маргинальным по анализу MolPrody: остаток находится в подходящих участках карты, наблюдается лишь незначительный Clash (может быть объяснен неверными водородами).

E 30	MET	104.65	0.45A H with E 27 VAL HG22	Allowed (0.08%) General / -77.8,-83.2	Favored (5.6%) <i>mmm</i> chi angles: 267.8,277.8,285.5
------	-----	--------	-------------------------------	--	--

Посмотрим тот же остаток после PDB Redo.

E 30	MET	94.76	-	Allowed (1.05%) General / -88.7,-65.1	OUTLIER (0%) chi angles: 209.8,225.3,159
------	-----	-------	---	--	---

После переработки остаток стал аутлайером по карте Рамачандрана, незначительно улучшилось положение его боковой цепи. Электронная плотность на данном остатке также очень низкая.

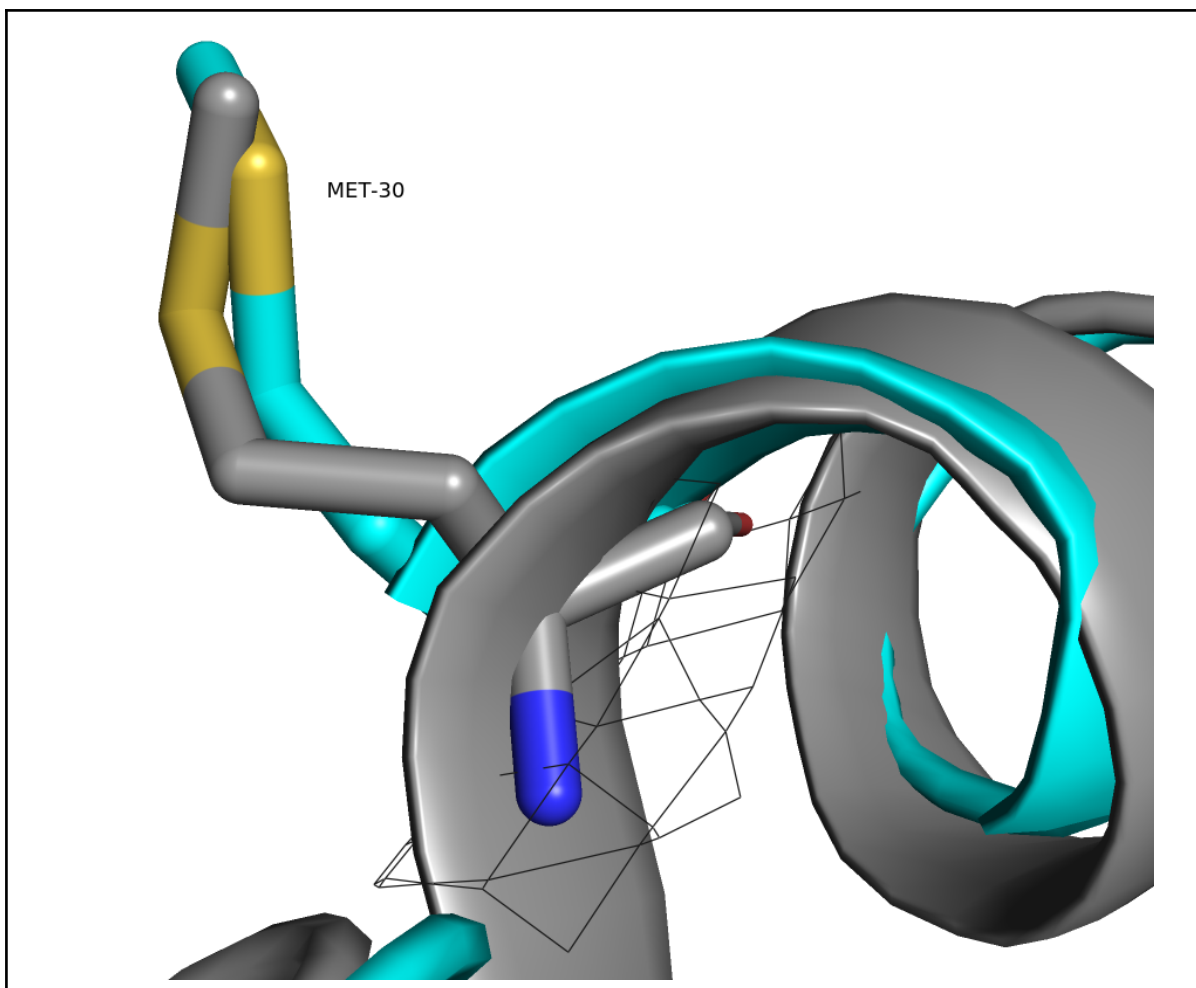


Рисунок 12. Вписанность остатка Met-30 (цепь E) в электронную плотность.

Рассмотрим три ранее маргинальных остатка в структуре PBD Redo и сравним их.

Для остатка KVal-27 практически ничего не изменилось: он все еще остается маргиналом по карте Рамачандрана; немного улучшаются значения ротамеров его боковых цепей, но данный эффект незначителен. PDB Redo считает, что качество для этого остатка изменилось незначительно.

K 27	VAL	69.89	0.41Å O with K 28 ASP HB3	OUTLIER (0%) Ile or Val / 11.4,-66.0	Favored (2.5%) <i>m</i> chi angles: 285.7
------	-----	-------	------------------------------	---	--

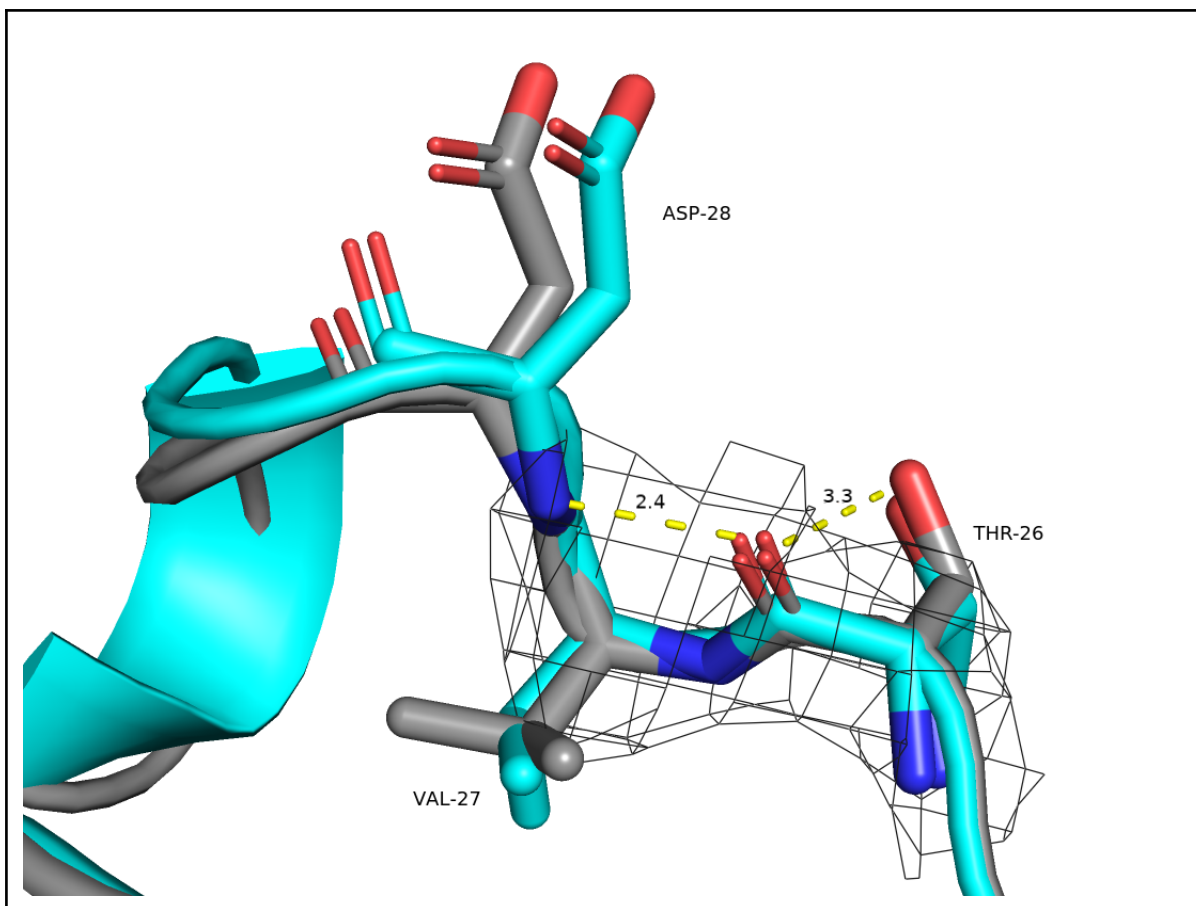


Рисунок 13. Различия между структурами.

Для остатка CLys-58 значительно изменилось положение боковой цепи: Mol Prody больше не выдает его как маргинальный остаток. В данном случае структура, полученная PDB Redo оказалась лучше.

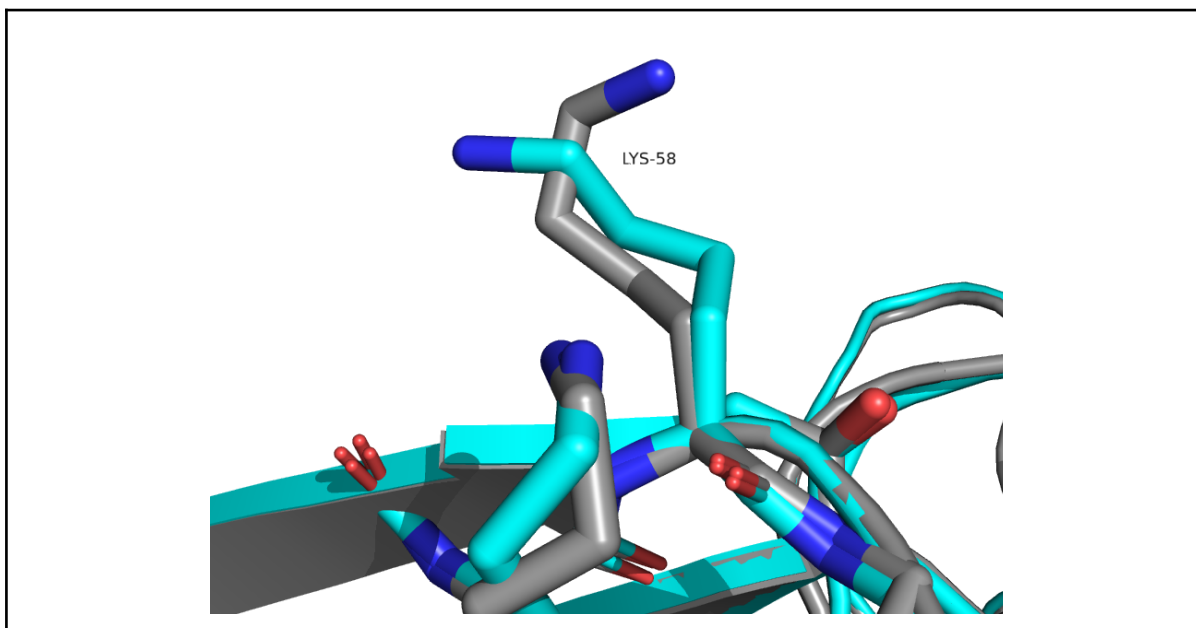
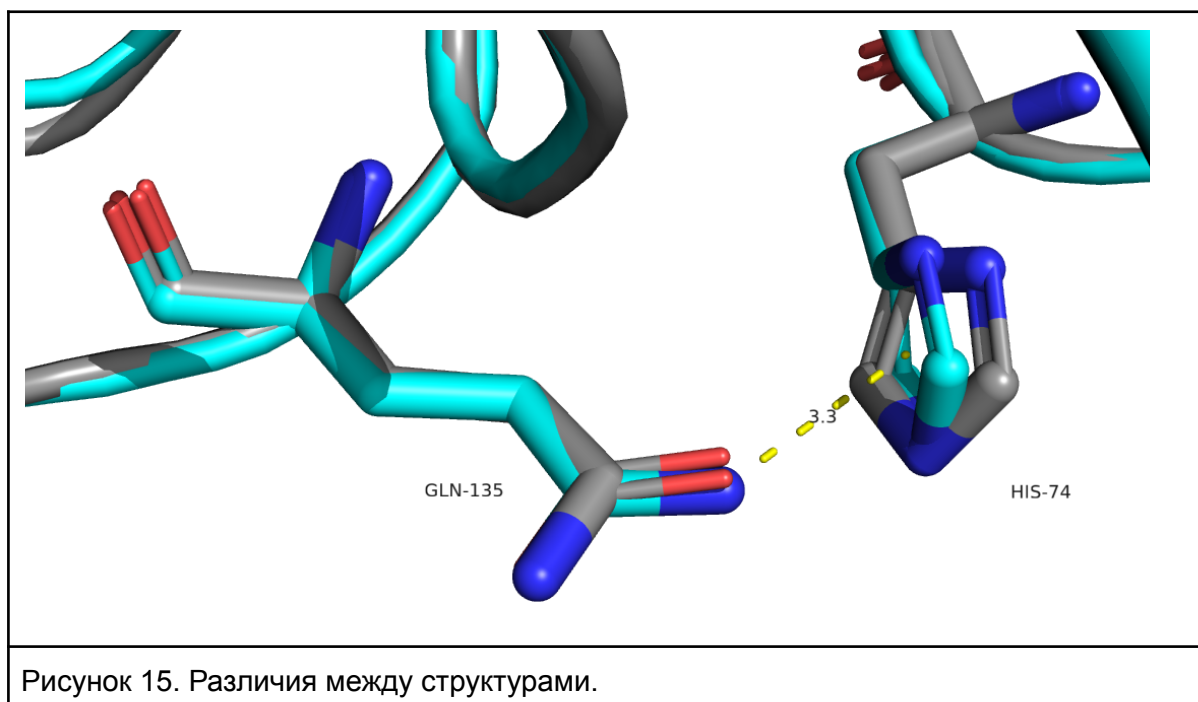


Рисунок 14. Различия между структурами.

Структура в PDB Redo развернула остаток FGln-135, что также улучшило модель.



В целом, стоит отметить, что, несмотря на улучшение некоторых характеристик структуры, модель PDB redo не отличается на порядок лучшим качеством: мы видим достаточно много остатков, для которых качество либо ухудшается, либо остается неизменным. Мы видим значительное ухудшение качества модели по ротамерным остаткам.

В целом, структура PDB Redo также пригодна для использования, но наилучших результатов, на мой взгляд, можно добиться при комбинировании информации о качестве отдельных остатков как исходной структуры, так и переделанной, т.к. довольно большое количество остатков в переделанной модели значительно ухудшаются по качеству.

Le Trong, Isolde, Pavel Aprikian, Brian A. Kidd, Manu Forero-Shelton, Veronika Tchesnokova, Ponni Rajagopal, Victoria Rodriguez, et al. 2010. "Structural Basis for Mechanical Force Regulation of the Adhesin FimH via Finger Trap-like Beta Sheet Twisting." *Cell* 141 (4): 645–55.

Yakovenko, Olga, Shivani Sharma, Manu Forero, Veronika Tchesnokova, Pavel Aprikian, Brian Kidd, Albert Mach, Viola Vogel, Evgeni Sokurenko, and Wendy E. Thomas. 2008. "FimH Forms Catch Bonds That Are Enhanced by Mechanical Force due to Allosteric Regulation." *The Journal of Biological Chemistry* 283 (17): 11596–605.