

Отчет по качеству PCA расшифровки структуры белка TTR (PDB ID: 3KGU)

Бартыш Катя, группа 402

Аннотация

В работе описана основная информация о белке TTR, но, помимо этого, был проведен анализ и оценка качества расшифровки белка, которые были получены методом рентгеноструктурного анализа. Также были проанализированы маргинальные остатки и проведено сравнение разных PDB моделей.

Введение

Транстиретин (TTR) является транспортным белком, богатым тетрамерными бета-листами, непосредственно вовлеченным в амилоидные заболевания человека. Недавно было обнаружено, что изофлавоновый генистеин (GEN) эффективно ингибирует образование амилоидных фибрилл TTR и, следовательно, является перспективным объектом исследования для лечения амилоидоза. Были изучены характеристики связывания генистеина с диким типом (TTRwt) и его наиболее частым амилоидогенным вариантом, мутантом V30M. В зависимости от дозы, генистеин вызывал значительное увеличение стабильности как мутантной, так и TTRwt, что было продемонстрировано с помощью анализов высокого гидростатического давления (ННД) и кислотно-опосредованной диссоциации/денатурации. Эксперименты с кристаллическими комплексами TTR показали, что механизмы связывания генистеина с TTRwt и V30M различны и зависят от конформаций структуры apoTTR. Кроме того, мы могли бы также идентифицировать потенциальные аллостерические движения, вызванные связыванием генистеина с TTR дикого типа, что объясняет, по крайней мере частично, часто наблюдаемый отрицательно кооперативный процесс между двумя сайтами TTRwt при связывании лигандов. Эти результаты показывают, что мутанты TTR могут демонстрировать различное распознавание лигандов и, следовательно, имеют значение в конструкции лигандов для ингибирования амилоидоза TTR [1].



Рис.1 Модель белка TTR

Общая информация о модели

Белок был добавлен в базу данных PDB 29 октября 2009 года, авторами являются Trivella, D.V. и Polikarov, I. Также в этой же статье были описаны такие белки как 3KGT и 3KGS.

TTR содержит 1964 атома и состоит из двух цепей - А и В, каждая из которых содержит лиганд – гистеин. Длина последовательности составляет 230 аминокислотных остатков. Биологическая единица не совпадает с асимметрической и представляет из себя димер. Некристаллографические симметрии отсутствуют. Кристаллографическая группа - P 2₁ 2₁ 2. Параметры кристаллографической ячейки представлены в таблице ниже.

Length (Å)	Angle (°)
a = 43.530	α = 90.00
b = 85.350	β = 90.00
c = 64.970	γ = 90.00

Таблица 1. Параметры кристаллографической ячейки

Информация о методе решения фазовой проблемы отсутствует, возможно, в данном случае она не была решена формально, а лишь с помощью догадок.

Число измеренных рефлексов составляет 20208. Разрешение - 1.85 Å. Диапазон разрешений структурных факторов: 1.85 – 21.4Å. Полнота набора структурных факторов – 94.6% [2].

Значения индикаторов качества модели в целом.

R-фактор равен 0.197, а R-free – 0.234. R-фактор показывает насколько экспериментальные данные структурных факторов соотносятся с рассчитанными факторами. Показателем хорошего качества считается фактор, меньший 0.25 ($0.197 < 0.25$), так что в данном случае можно считать, что качество хорошее. R-free отличается от R-фактора на 15.8%, что довольно большая разница и намного больше критической границы в 10%.

На рисунке 2 приведены результаты работы сервиса MolProbity. При получении этих данных были добавлены водороды для выявления недопустимых наложений атомов. При этом были выявлены возможные инверсии боковых цепей Asn, Gln, His. Всего было выявлено 3 инверсии:

- A 90 HIS
- B 98 ASN A
- B 124 ASN

All-Atom Contacts	Clashscore, all atoms:	2.16	99 th percentile* (N=865, 1.85Å ± 0.25Å)
	Clashscore is the number of serious steric overlaps (> 0.4 Å) per 1000 atoms.		
Protein Geometry	Poor rotamers	4	1.94% Goal: <0.3%
	Favored rotamers	195	94.66% Goal: >98%
	Ramachandran outliers	0	0.00% Goal: <0.05%
	Ramachandran favored	223	98.67% Goal: >98%
	MolProbity score [^]	1.21	99 th percentile* (N=12654, 1.85Å ± 0.25Å)
	Cβ deviations >0.25Å	0	0.00% Goal: 0
	Bad bonds:	0 / 1903	0.00% Goal: 0%
Bad angles:	1 / 2618	0.04% Goal: <0.1%	
Peptide Omegas	Cis Prolines:	0 / 12	0.00% Expected: ≤1 per chain, or ≤5%
Additional validations	Chiral volume outliers	0/302	
	Waters with clashes	4/146	2.74% See UnDowser table for details

Рис. 2 Результат работы сервиса MolProbity для 3KGU

ClashScore – число недопустимых наложений атомов на 1000. То есть в данном случае это 0.216% атомов. Также этот параметр показывает, что 99% структур ClashScore хуже, чем у данной, что является очень хорошим результатом.

Poor rotamers – число остатков с маргинальными по отклонению от ротамеров боковыми цепями для данного белка составляет 4 (1.94%), что значительно превышает критическую границу.

Ramachandran outliers – полные маргиналы по карте Рамачандрана, лежащие вне допустимой области. Для данной структуры значение 0%, что говорит о высоком разрешении.

Ramachandran favores – процент остатков в предпочитаемой области – составляет 98.67% соответственно.

MolProbity score – это интегральная оценка структуры по данным этого сервиса. В данном случае она очень хорошая.

C β deviations >0.25Å - число C β -атомов с неприемлемым отклонением от ожидаемого положения составляет 0%.

Значения Bad backbone bonds и Bad backbone angles, показывающие число ковалентных связей и валентных углов соответственно, существенно отклоняющихся от теории близки к 0.

На рисунке 3 представлены данные тех же параметров, что и на рисунке 2, представленные в базе данных PDB. И, как можно заметить, такие значения как Rfree, Clashscore и Ramachandran outliers совпадают с теми же показателями, полученными в результате работы сервиса MolProbity.

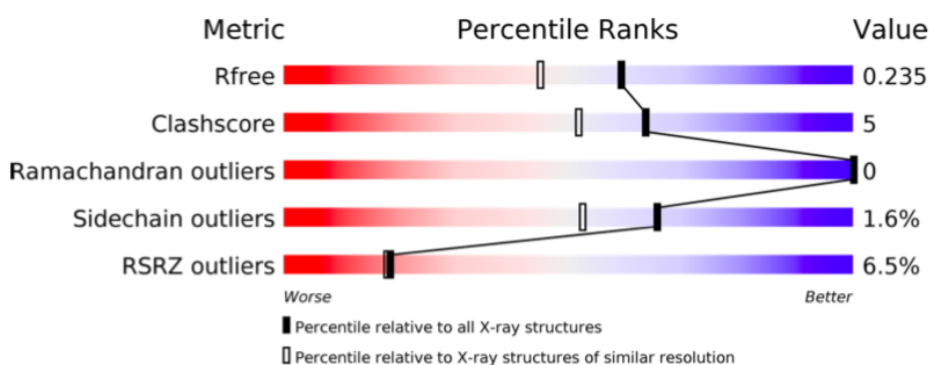


Рис. 3 Данные валидации для 3KGU из базы данных PDB.

В целом, можно сказать, что качество структуры довольно хорошее, хотя и содержит превышающее критический порог число остатков с маргинальным по отклонению от ротамеров боковыми цепями.

Помимо данных, представленных выше, в базе данных PDB и сервисом MolProbity также представляется карта Рамачандрана (Рис. 4), которая иллюстрирует допустимые углы вращения между атомами остова. В предпочтительной зоне находится 98.8% всех аминокислотных остатков, в допустимой – 1.2%.

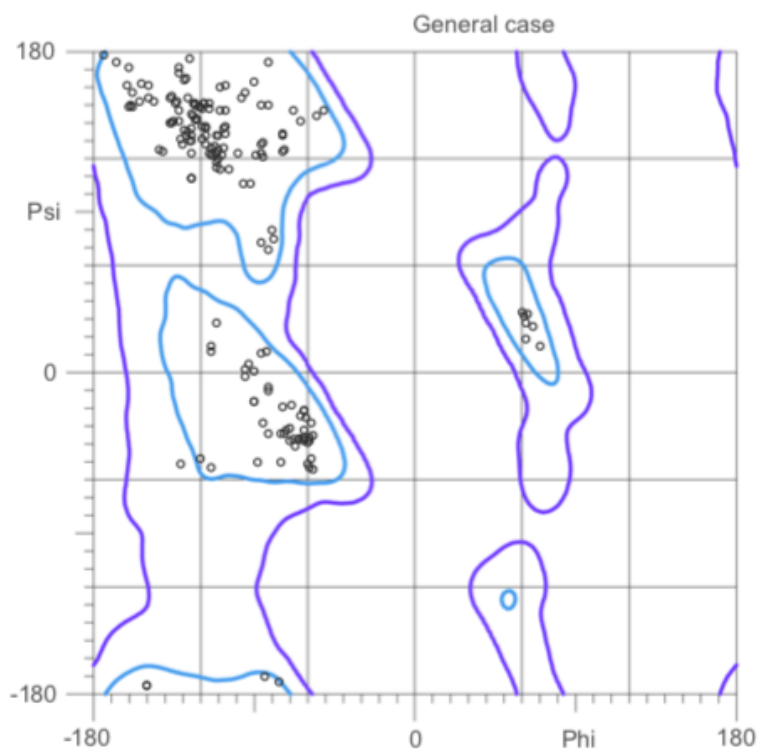


Рис.4 Карта Рамачандрана для 3KGU. Черными точками представлены аминокислотные остатки, голубые линии отделяют предпочтительные зоны, фиолетовые – допустимые.

Маргинальные остатки и их анализ.

Как видно из Рис. 3, значение RSRZ для данного белка хуже, чем для всех структур, находящихся в базе данных. Параметр RSR характеризует то, насколько аминокислотный остаток соответствует экспериментальным данным. RSR оценивается числом Z-score, которое показывает насколько RSR для остатка данной модели отличается от среднего значения RSR структур такого же разрешения. Остаток считается маргинальным, если его $Z\text{-score} > 2$.

Для данного белка число маргиналов RSRZ составляет 6% (15/230). Однако, стоит отметить, что проанализированы были не все остатки (230/254), а только 90% от общего количества, то есть общее количество маргинальных остатков может на самом деле варьироваться. На рисунке 5 представлены все 15 маргинальных остатков по параметру RSRZ [3].

Также с помощью сервиса EDS было получено наглядное изображение RSR для остатков, оно представлено на рисунке 6 [4].

Mol	Chain	Res	Type	RSRZ
1	B	37	ALA	5.6
1	B	102	PRO	5.4
1	B	10	CYS	4.4
1	B	124	ASN	3.8
1	B	39	ASP	3.7
1	B	103	ARG	3.6
1	B	40	THR	3.3
1	B	36	ALA	3.2
1	A	39	ASP	3.1
1	A	10	CYS	2.9
1	B	41	TRP	2.6
1	B	100	SER	2.6
1	A	100	SER	2.5
1	B	38	ASP	2.3
1	A	124	ASN	2.2

Рис. 5 Список аутлайнеров RSRZ

Chain	Nres	<Z>	Sigma	%<-1	%<0	%>1	%>2
A	115	-0.15	0.80	3.48	73.04	6.96	4.35
B	115	0.21	1.26	0.87	65.22	13.91	7.83
All	230	0.03	1.05	2.17	69.13	10.43	6.09

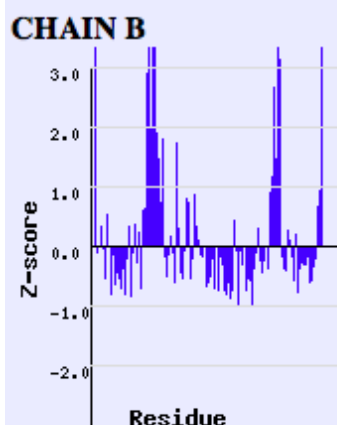
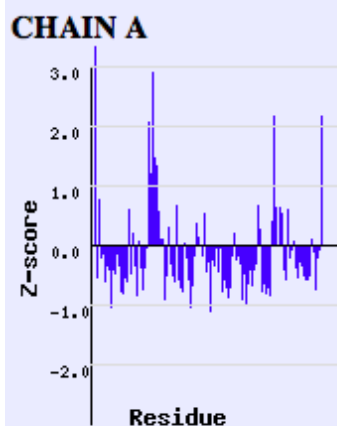
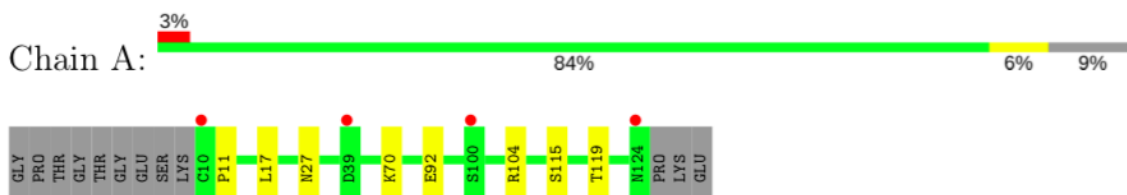


Рис. 6 Изображение RSR, полученное с помощью сервиса EDS

На рисунке 7 отображена визуализация последовательности аминокислотных остатков, аннотированная относительно качества геометрии электронной плотности. Остатки окрашены в разные цвета согласно тому, какое количество критериев определяет данную аминокислоту в качестве выброса (если он есть). Красный цвет – более 3 критериев, оранжевый – 2, желтый – 1, зеленый – 0. Красными точками помечены аутлайнеры RSRZ.

- Molecule 1: Transthyretin



- Molecule 1: Transthyretin

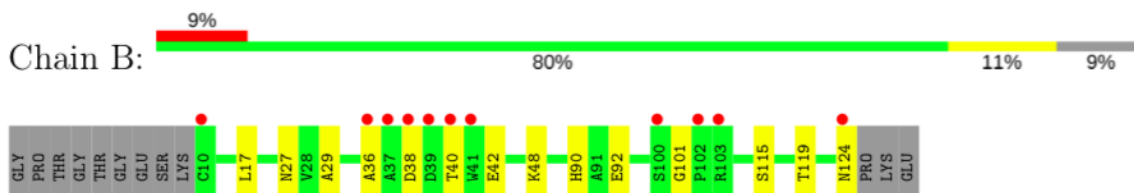


Рис. 7 Визуализация последовательности аминокислот 3KGU (качество геометрии электронной плотности)

В таблице 2 представлен список маргинальных остатков, характеристики для которых были получены с помощью сервера MolProbity.

Остаток	Причина маргинальности
A 92 GLU	Перекрытие OE2 с A 70 LYS HD2
B 90 HIS	Перекрытие CD2 с B 92 GLU HG3
B 101 GLY	Недопустимый угол (N-CA-C)
A 11 PRO	Перекрытие HA с A 104 ARG HD3
B 119 THR A	Перекрытие HG23 с A 115 SER HB2

Таблица 2. Список маргинальных аминокислотных остатков для белка 3KGU.

A 92 GLU

На рисунке 8 представлено изображение двух аминокислотных остатков – глутаминовой кислоты (справа) и лизина (слева). Видно, что их электронные плотности очень сильно перекрываются (перекрытие составляет 0.91 ангстрема). Я думаю, что это не ошибка расшифровки, а особенности структуры.

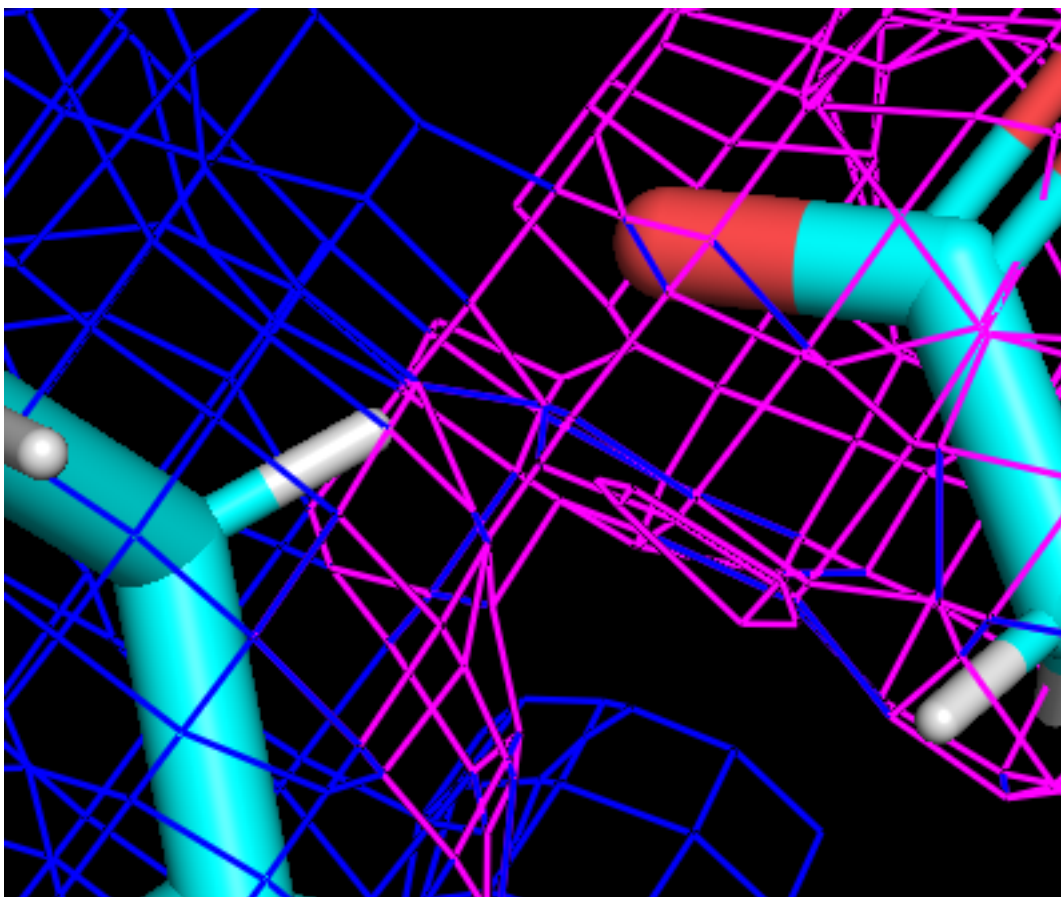


Рис. 8 Изображение A 92 GLU и A 70 LYS с построенной электронной плотностью. Серым обозначен атом HD2, красным - OE2. Уровень подрезки = 0.4

B 90 HIS

На рисунке 9 представлено изображение двух аминокислотных остатков – гистидина (слева) и глутаминовой кислоты (справа). Видно, что их электронные плотности вообще не пересекается (перекрытие 0.42 ангстрема). Видимо, такое маленькое перекрытие нельзя отследить в PyMol, либо это ошибка расшифровки.

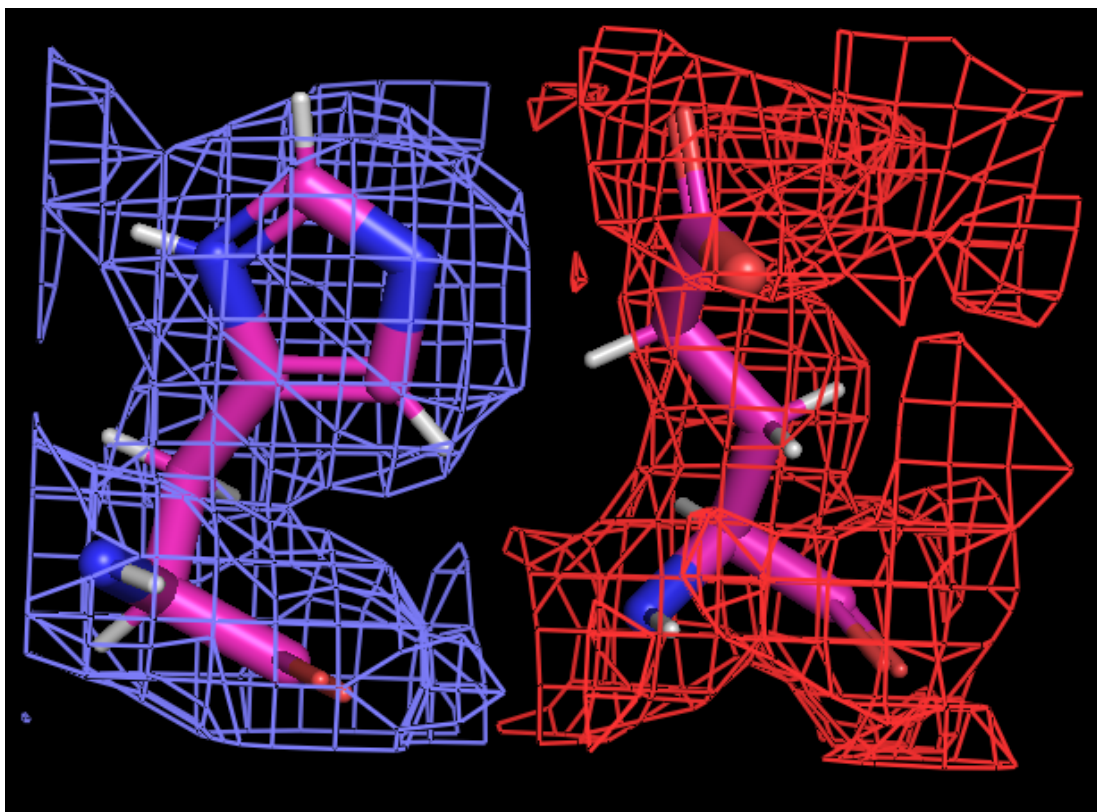


Рис. 9 Изображение B 90 HIS и B 92 GLU с построенной электронной плотностью. Уровень подрезки = 0.5

B 101 GLY

На рисунке 10 представлено изображение глицина с измеренным углом между N-CA-C. Получившееся значение составляет – 96.526 градусов, что является особенностью структуры.

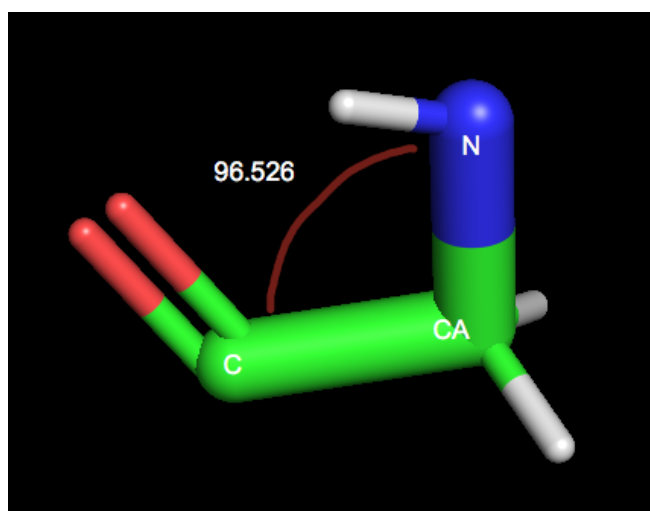


Рис. 10 Изображение B 101 GLY с измеренным углом

A 11 PRO

На рисунке 11 представлено изображение двух аминокислотных остатков – пролина (снизу) и аргинина (сверху).

Видно, что их электронные плотности вообще не пересекается (перекрытие 0.51 ангстрема). Видимо, это ошибка расшифровки, потому что я думаю, что при таком перекрытии это было бы видно в PyMol.

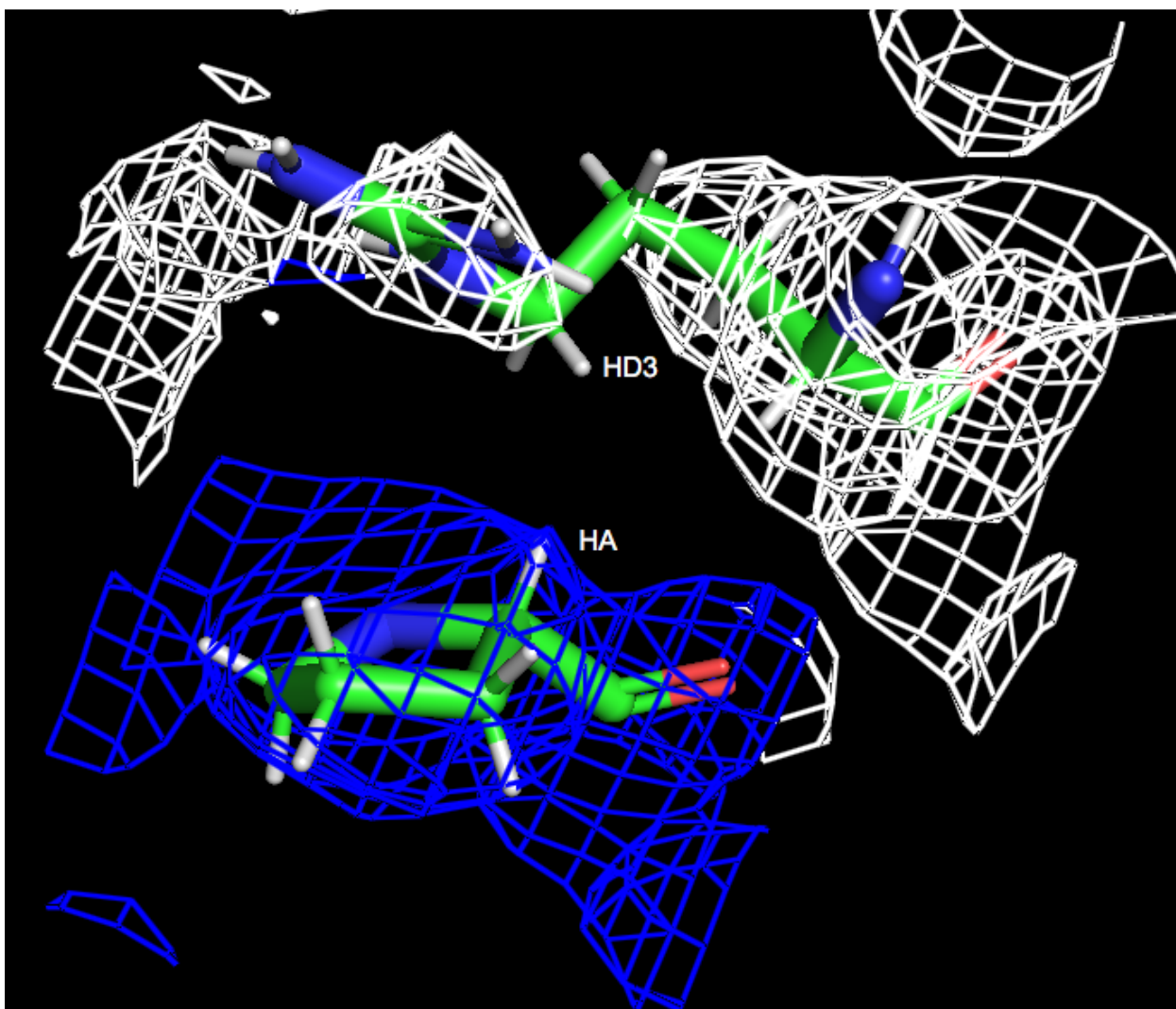


Рис. 11 Изображение A 11 PRO и A 104 ARG с построенной электронной плотностью. Уровень подрезки = 0.5

B 119 THR

На рисунке 12 представлено изображение двух аминокислотных остатков – триптофан (справа) и серин (слева). Их перекрытие составляет 0.51 ангстрем. Электронная плотность серина очень обширная, охватывает большую площадь и перекрывается с электронной плотностью триптофана. Это можно считать особенностью структуры.

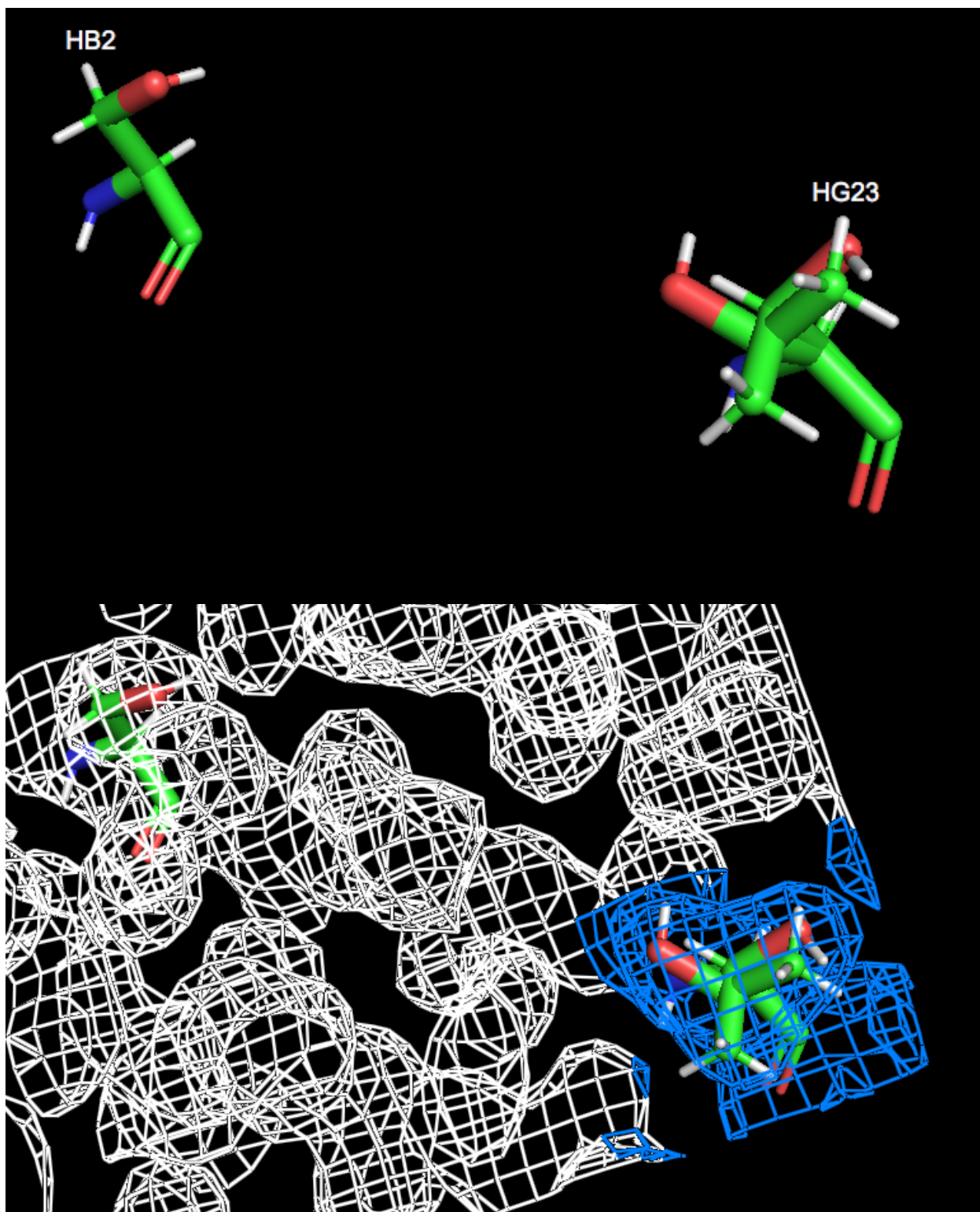


Рис. 12 Изображение B 119 THR и A 115 SER с построенной электронной плотностью. Уровень подрезки = 0.5. (На верхнем изображении показаны интересующие нас атомы)

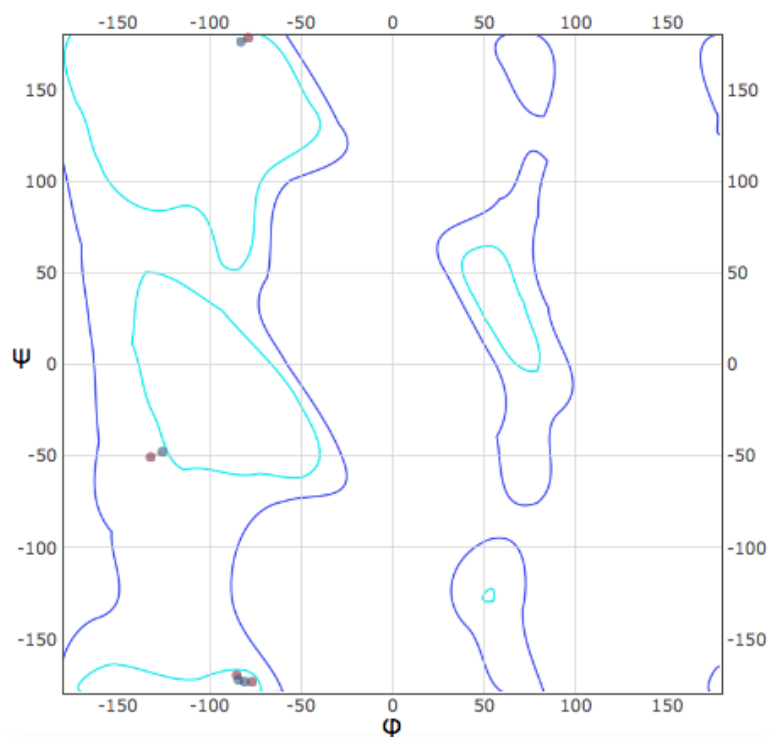
Сравнение модели из PDB с моделью из PDB_redo

Модель в PDB_redo несколько отличается от описанной модели PDB. Во-первых, значения R-фактора и R-free в ней улучшены, что на

самом деле улучшает качество расшифровки. Что касается карты Рамачандрана, то такая из PDB-redo несильно отличается от карты из PDB. В связи с этим изменяется и количество маргинальных аминокислотных остатков, что, по моему мнению, несильно, но влияет на качество расшифровки [5].

Validation metrics from PDB-REDO		
	PDB	PDB-REDO
Crystallographic refinement		
<i>R</i>	0,1947	0,1773
<i>R-free</i>	0,2267	0,2002
<i>Bond length RMS Z-score</i>	0,475	0,685
<i>Bond angle RMS Z-score</i>	0,629	0,833
Model quality (raw scores percentiles)		
<i>Ramachandran plot appearance</i>	99	99
<i>Rotamer normality</i>	77	86
<i>Coarse packing</i>	N/A	N/A
<i>Fine packing</i>	99	99
<i>Bump severity</i>	61	99
<i>Hydrogen bond satisfaction</i>	61	73
<i>WHAT_CHECK</i>	Report	Report

PDB-REDO Kleywegt-like plot for 3kgu



Description	PDB ●	PDB-REDO ●
Ramachandran Z-score	2.352	2.364
<input type="checkbox"/> Preferred regions	222	224
<input checked="" type="checkbox"/> Allowed regions	4	2
<input checked="" type="checkbox"/> Outliers	0	0

Significant model changes

Description	Count
<i>Rotamers changed</i>	6
<i>Side chains flipped</i>	4
<i>Waters removed</i>	10
<i>Peptides flipped</i>	0
<i>Chiralities fixed</i>	0
<i>Residues fitting density better</i>	12
<i>Residues fitting density worse</i>	0

Рис. 13 Сравнение качества модели PDB и PDB-redo

Выводы

Я думаю, что в данном случае судить как-то однозначно о качестве расшифровки нельзя. Потому что есть показатели, по которым расшифровку можно считать хорошей, а есть параметры, по которым она плоха относительно других структур с аналогичным разрешением в PDB. Также некоторые аминокислотные остатки оказались неправильно принятыми за маргинальные, хотя по протоколу получалось, что их можно отнести к этой категории. Поэтому получается, что данную расшифровку можно отнести к средней по качеству, и, я думаю, что данные из PDB_redo показывают, что она может быть оптимизирована и улучшена.

Источники

1. Conformational differences between the wild type and V30M mutant transthyretin modulate its binding to genistein: implications to tetramer stability and ligand-binding. Trivella, D.B., Bleicher, L., Palmieri, L.C., Wiggers, H.J., Montanari, C.A., Kelly, J.W., Lima, L.M., Foguel, D., Polikarpov, I. (2010) J.Struct.Biol. 170: 522-531 DOI: 10.1016/j.jsb.2010.03.002
2. <https://www.rcsb.org/pdb/explore/materialsAndMethods.do?structureId=3KGU>
3. https://files.rcsb.org/pub/pdb/validation_reports/kg/3kgu/3kgu_full_validation.pdf
4. <http://eds.bmc.uu.se/cgi-bin/eds/zscore?3kgu>
5. <https://pdb-redo.eu/db/3kgu>