

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

ФАКУЛЬТЕТ БИОИНЖЕНЕРИИ И БИОИНФОРМАТИКИ

**Практикум по дисциплине “Биоинформатика”
Отчет по качеству расшифровки РСА структуры
трансляционного фактора eIF2A.**

*Отчет студентки 4-го курса
Григорьевой Марии Владимировны*

Оглавление

1. Аннотация
2. Введение
3. Результаты и обсуждение
 - 1) общая информация о модели
 - 2) значение идентификаторов качества модели в целом
 - 3) анализ маргинальных остатков
 - 4) сравнение с PDB-REDO
4. Заключение
5. Список литературы

Аннотация

В данной работе был проведен анализ качества расшифровки структуры эукариотического трансляционного фактора eIF2A (PDB ID: 3WJ9), полученной методом рентгеноструктурного анализа. Приведены различные параметры модели, в том числе общие параметры и информация о маргинальных остатках.

В результате проведенной работы появилась возможность сделать вывод о качестве расшифровки этой структуры и сравнить рассматриваемую модель с подобной моделью с PDB-REDO.

Введение

EIF2A - это фактор инициации трансляции у эукариот (не путать с eIF2-alpha). Участвует в регуляции трансляции и сборки рибосом.

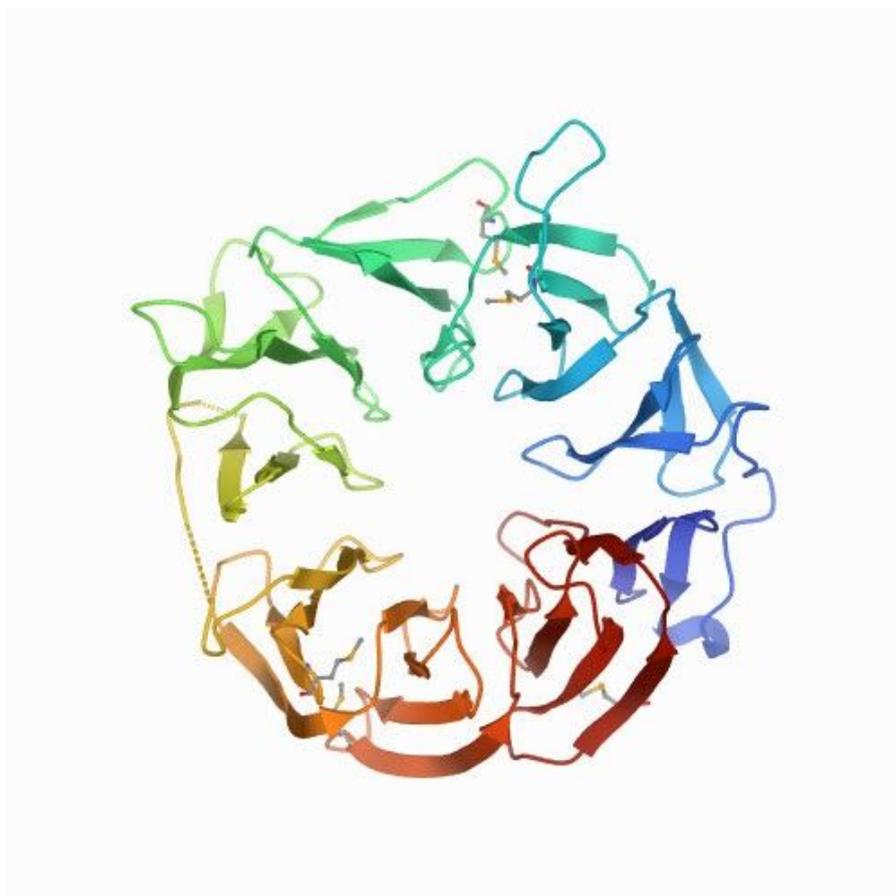


Рис 1. Структура eIF2A из PDB (3WJ9)

Функции и механизмы работы этого фактора все еще не до конца изучены. Работа над этим на данный момент активно проводится группой моих коллег.

Получение структуры может значительно упростить работу и, возможно, позволит предположить механизмы взаимодействий eIF2A .

Общая информация о модели

Информация, приведенная далее, взята с ресурса RCSB PDB по ID 3WJ9.

Состав комплекса: Crystal structure of the eukaryotic translation initiation factor 2A from *Schizosaccharomyces pombe*

Год: 2014

Авторы: Kashiwagi, K., Ito, T., Yokoyama, S.

Статья (DOI): 10.1007/s10969-014-9177-y

Метод решения фазовой проблемы: молекулярное замещение

Число измеренных рефлексов: 47788

Разрешение: 2.51 Å

Полнота набора структурных факторов: 99.9%

Параметры кристаллографической ячейки:

$a = 93.443 \text{ \AA}$, $b = 93.443 \text{ \AA}$, $c = 269.387 \text{ \AA}$

$\alpha = 90.00^\circ$, $\beta = 90.00^\circ$, $\gamma = 120.00^\circ$

Кристаллографическая группа: $P 3_1 2 1$

Значение идентификаторов качества модели в целом

Параметры для оценки качества полученной модели:

R-Value Free (тестовая выборка): 0.258

R-Value Work (тренировочная выборка): 0.238

R-Value Observed (весь набор): 0.240

В выбранной модели R-фактор, характеризующий соответствие модели ее экспериментальным значениям, равен 0.238, то есть < 0.25 , поэтому можем сделать вывод, что соответствие достаточно.

При этом R-free фактор все же больше значения 0.2, пусть и вписывается в допустимые значения.

По приведенным данным можно предварительно судить об отсутствии переоптимизации модели и ее соответствии экспериментальным данным.

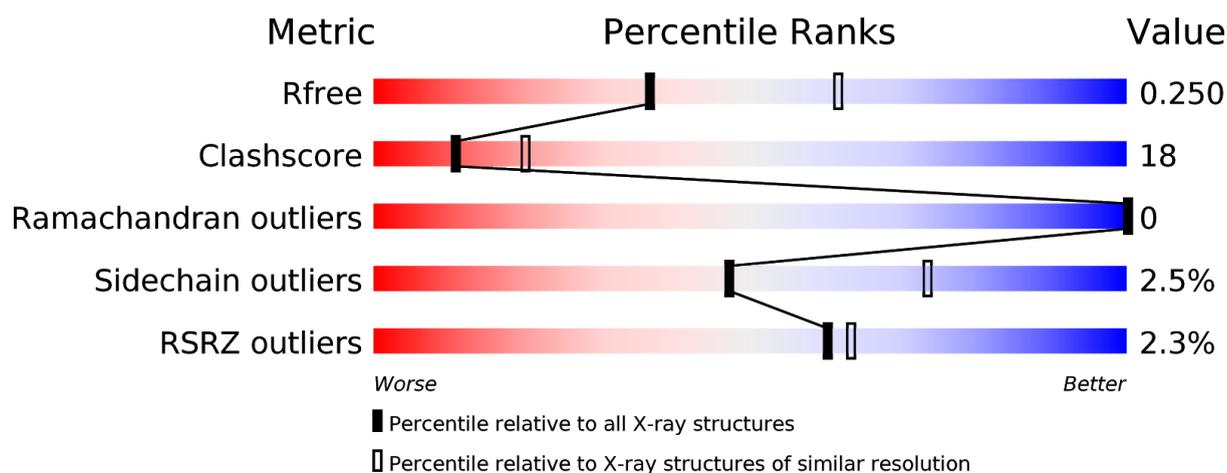


Рис 2. Показатели качества eIF2A из PDB (3WJ9)

Также структура была оценена с помощью сервера MolProbity.

Таблица 1. Оценка качества с помощью сервера MolProbity.

All-Atom Contacts	Clashscore, all atoms:	17.6	79 th percentile* (N=262, 2.51Å ± 0.25Å)
	Clashscore is the number of serious steric overlaps (> 0.4 Å) per 1000 atoms.		
Protein Geometry	Poor rotamers	26	3.83% Goal: <0.3%
	Favored rotamers	582	85.84% Goal: >98%
	Ramachandran outliers	0	0.00% Goal: <0.05%
	Ramachandran favored	758	96.68% Goal: >98%
	Rama distribution Z-score	-2.92 ± 0.26	Goal: abs(Z score) < 2
	MolProbity score [^]	2.40	82 nd percentile* (N=6730, 2.51Å ± 0.25Å)
	Cβ deviations >0.25Å	5	0.67% Goal: 0
	Bad bonds:	0 / 6438	0.00% Goal: 0%
Bad angles:	10 / 8749	0.11% Goal: <0.1%	
Peptide Omegas	Cis Prolines:	4 / 50	8.00% Expected: ≤1 per chain, or ≤5%
	Cis nonProlines:	1 / 741	0.13% Goal: <0.05%
Low-resolution Criteria	CaBLAM outliers	11	1.4% Goal: <1.0%
	CA Geometry outliers	9	1.17% Goal: <0.5%
Additional validations	Chiral volume outliers	0/945	
	Waters with clashes	2/61	3.28% See UnDowser table for details

Для лучшей оценки были добавлены водороды и произведена оптимизация связей.

ClashScore - число недопустимых наложений на 1000 атомов; по результатам модели у 79% структур с таким же разрешением ClashScore хуже. Это значит, что качество структуры в целом хорошее.

Далее будет приведена расшифровка некоторых параметров таблицы - тех, в которых у текущей модели есть "погрешности" (желтый и красный цвет).

Poor rotamers - количество маргинальных остатков по отклонению боковой цепи; данное значение не должно превышать 0.3, однако здесь значение составляет 3.86% (26 остатков).

Ramachandran outliers - маргиналы по карте Рамачандрана, которые лежат вне допустимой области; в данной модели таких остатков нет.

Ramachandran favored - остатки, лежащие в предпочитаемой области; в данной модели значение в пределах допустимого (96.68% - это меньше, чем идеальные 98, но очень близко).

C β deviations > 0.25Å - число C β -атомов с неприемлемым отклонением от ожидаемого положения; таких в структуре 0,67% (почти 0, поэтому значение считается приемлемым).

Bad angles - число углов, существенно отклоняющихся от теории, в данном случае их 0,11%. Ресурс считает, что это плохой результат, несмотря на то, что он лишь слегка превышает ограничение (<0,1%).

По карте Рамачандрана аутлаеров выявлено не было, 96.7% остатков попадают в предпочтительную область и 100% всех остатков находятся в разрешенных областях (рисунок 3).

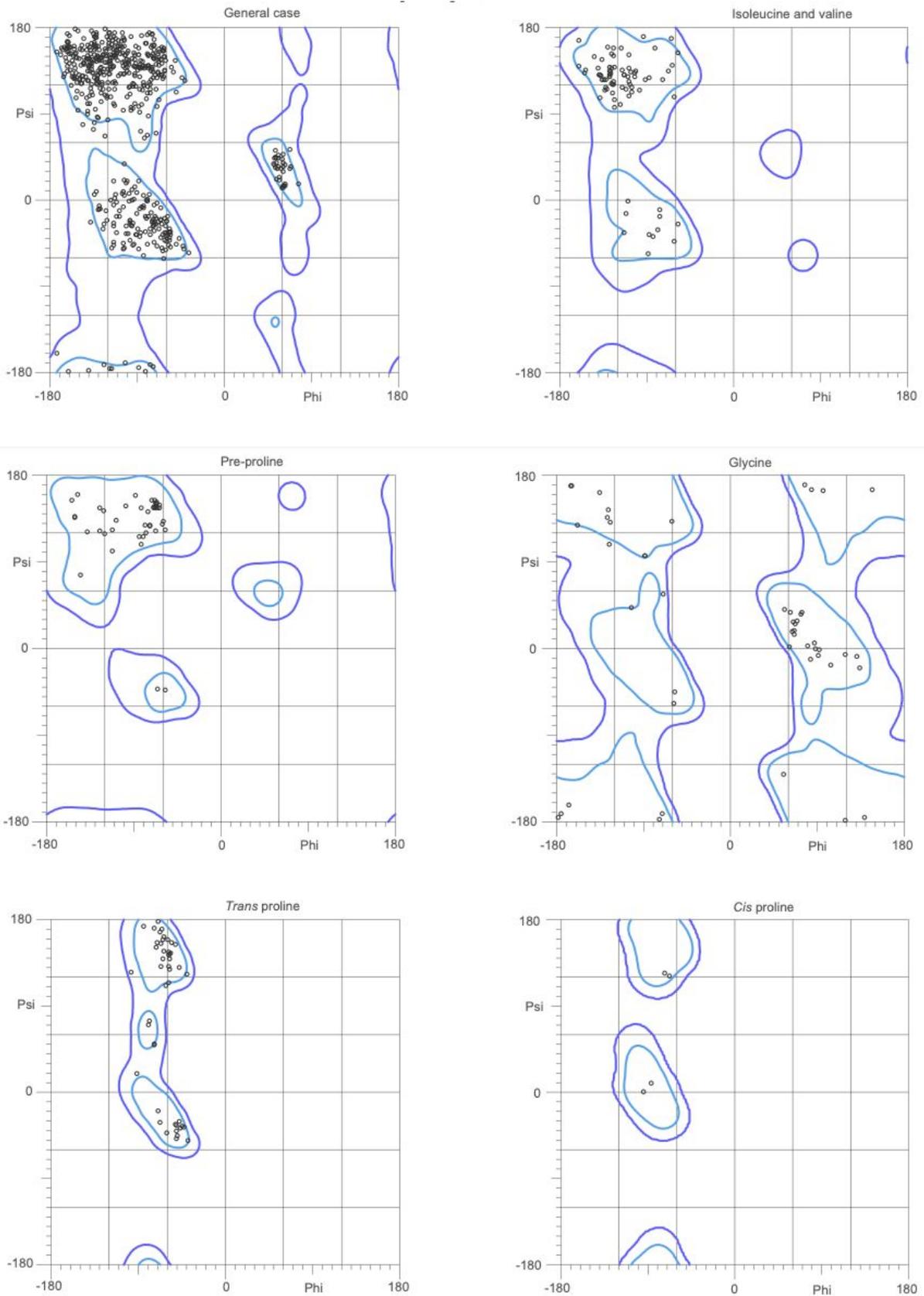


Рис 3. Карты Рамачандрана для структуры 3WJ9. 96.7% (758/784) of all residues were in favored (98%) regions. 100.0% (784/784) of all residues were in allowed (>99.8%) regions. There were no outliers.

Анализ маргинальных остатков

Для составления списка маргинальных остатков использовалась совокупность данных, полученных из указанных ранее источников.

Таблица 2. Список маргинальных остатков

Остаток	Показатель маргинальности
B:LEU-411	- отклонение положения атомов 0.27Å - отклоняющиеся углы OUTLIER(S) worst is C-CA-CB: 4.7 σ
B:ILE-388	- перекрывание HD11 with B 366 GLN HA - недопустимый ротамер OUTLIER (0.2%) chi angles: 324.2,97.2
B:MSE-345	- перекрывание 1.33Å HE1 with B 324 ILE CD1
B:TYR-244	- перекрывание 1.34Å CB with B 261 CYS HA - RSRZ 2,9 - недопустимый ротамер OUTLIER (0.3%) chi angles: 221.7,48.2 - CaBLAM CA Geom Outlier (0.057%)
A:GLN-234	- RSRZ 5,9
A:PHE-367	- перекрывание 1.01Å CD1 with A 393 MSE HE1
B:GLY-288	- перекрывание 0.52Å HA3 with B 267 ARG HB3
B:VAL-263	- перекрывание 1.08Å HG23 with B 244 TYR CE2
B:ASP-216	- перекрывание 0.47Å N with B 234 GLN O - CaBLAM Outlier (0.642%)
A:LYS-13	- недопустимый ротамер OUTLIER (0.3%) chi angles: 135.4,159.7,175.1,198.3
A:ASN-23	- CaBLAM CA Geom Outlier (0.383%)
B:SER-374	- CaBLAM CA Geom Outlier (0.083%)

B:LEU-411

- отклонение положения атомов 0.27\AA
- отклоняющиеся углы OUTLIER(S) worst is C-CA-CB: 4.7σ

Предположительно отклонение связано с тем, что остаток находится почти на самом конце цепи (предпоследний остаток), и этот фрагмент может быть довольно подвижным.

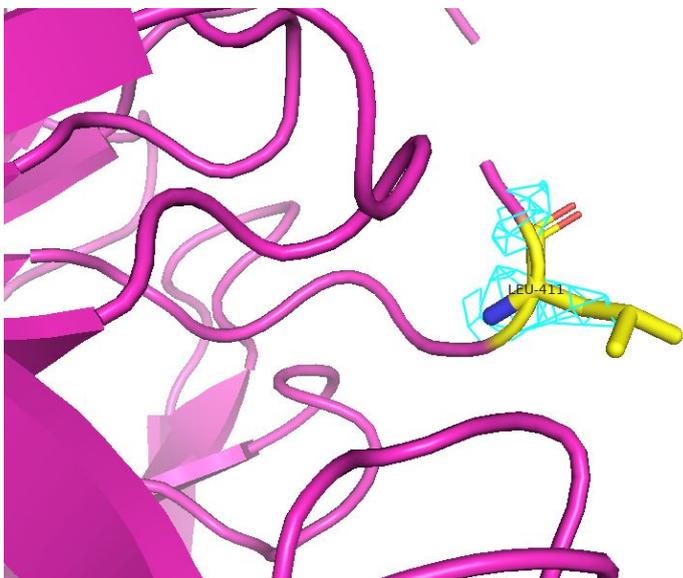


Рис 4. визуализация остатков B:LEU-411 и его электронная плотность при $carve=1.5$.

B:ILE-388

- перекрытие HD11 with B 366 GLN HA
- недопустимый ротамер OUTLIER (0.2%) chi angles: $324.2, 97.2$

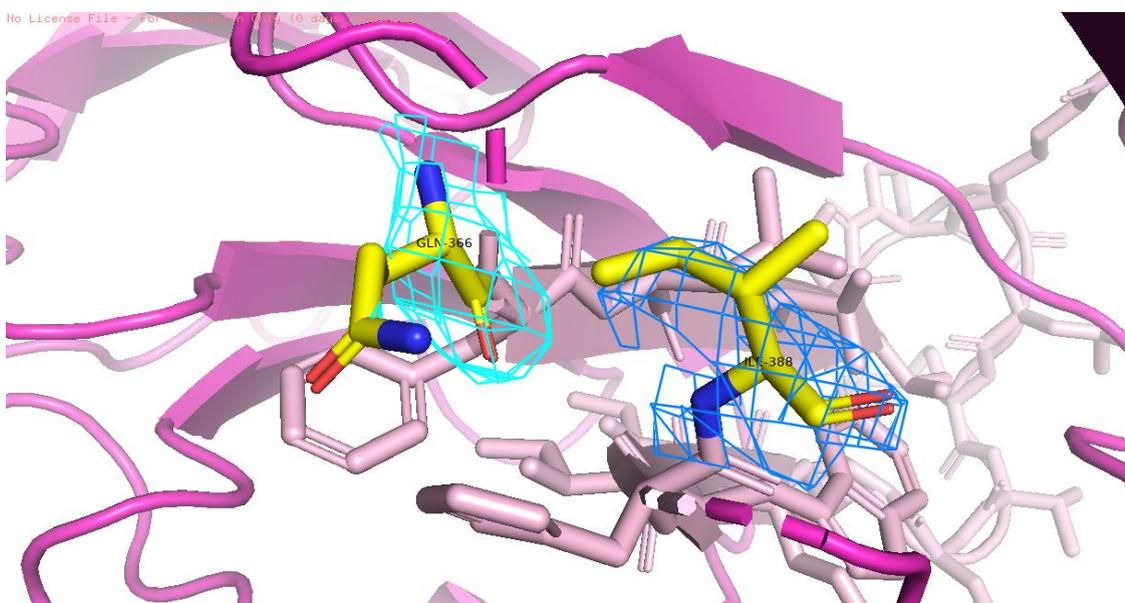


Рис 5. визуализация остатков B:ILE-388 и B:GLN-366 и их электронных плотностей при $carve=1.5$.

Перекрытие:

1:B:366:GLN:HA 1:B:388:ILE:HD11

Interatomic distance 1.54Å

Clash overlap 0.89Å

Что касается ротамера, остаток здесь находится в петле, смотрящей «наружу» от центра белка, при этом он крайний остаток близ листа. Можно заметить, что единственное взаимодействие, обуславливающее такой ротамер - взаимодействие с B:THR-349, который включен в структуру листа. Поэтому я предполагаю, что для данного примера нужно еще раз рассмотреть параметры разметки.

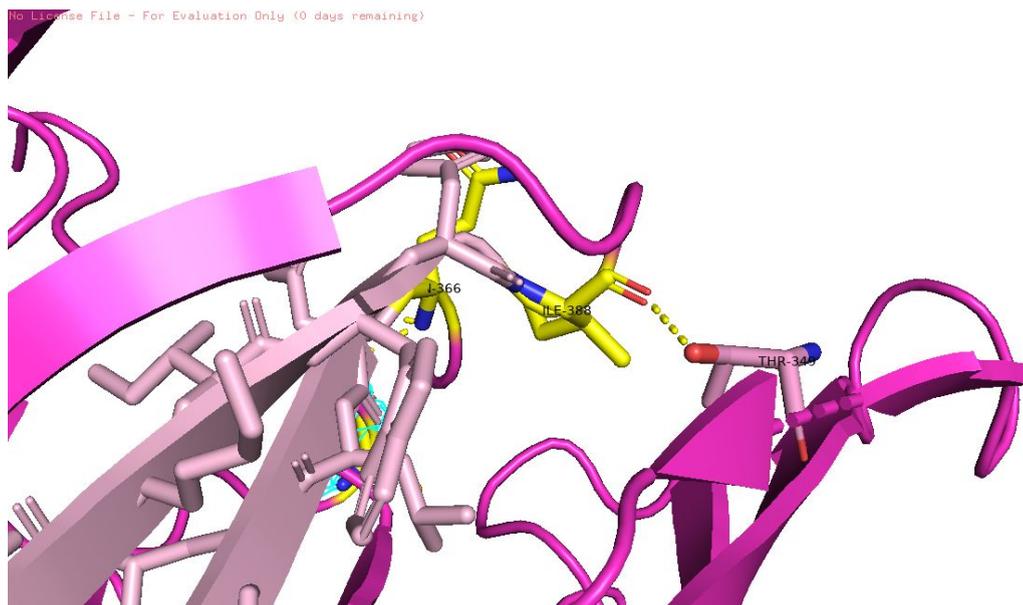


Рис 6. визуализация остатков B:ILE-388 и B:THR-349

B:MSE-345

- перекрытие 1.33Å HE1 with B 324 ILE CD1

Перекрытие:

1:B:324:ILE:CD1 1:B:345:MSE:HE1

Interatomic distance 1.57Å

Clash overlap 1.33Å

1:B:324:ILE:HD12 1:B:345:MSE:CE

Interatomic distance 1.61Å

Clash overlap 1.31Å

1:B:344:ASN:O 1:B:345:MSE:CB

Interatomic distance 2.64Å

Clash overlap 0.45Å



Рис 7. визуализация остатков B:MSE-345 и B:ILE-324 и их электронных плотностей при $carve=1.5$.

Можно заметить, что перекрытие сильно выше, чем в предыдущем случае. Перекрытие идет по двум атомам с заявленным B:ILE-324 и еще одно перекрытие с соседним B:ASN-344, которое, на мой взгляд, обуславливается неудобным положением, которое принимается из-за этого перекрытия.

B:TYR-244

- перекрывание 1.34Å CB with B 261 CYS HA
- RSRZ 2,9
- недопустимый ротамер OUTLIER (0.3%) chi angles: 221.7,48.2
- CaBLAM CA Geom Outlier (0.057%)

Перекрытие:

1:B:244:TYR:CB 1:B:261:CYS:HA
Interatomic distance 1.57 Clash overlap 1.34
1:B:244:TYR:CG 1:B:261:CYS:HA
Interatomic distance 1.84 Clash overlap 1.11
1:B:244:TYR:HB3 1:B:261:CYS:HA
Interatomic distance 1.26 Clash overlap 1.09
1:B:244:TYR:CE2 1:B:263:VAL:HG23
Interatomic distance 1.87 Clash overlap 1.08
1:B:244:TYR:CB 1:B:261:CYS:CA
Interatomic distance 2.36 Clash overlap 1.03
1:B:244:TYR:CD2 1:B:261:CYS:HA
Interatomic distance 1.94 Clash overlap 1.01

1:B:244:TYR:CZ 1:B:263:VAL:HG23
Interatomic distance 2.03 Clash overlap 0.93
1:B:244:TYR:HB3 1:B:261:CYS:CA
Interatomic distance 2.01 Clash overlap 0.88

...

Всего было найдено 27 перекрытий. Все имели разные параметры. Выше привожу 8 перекрытий с максимальным значением перекрытия.

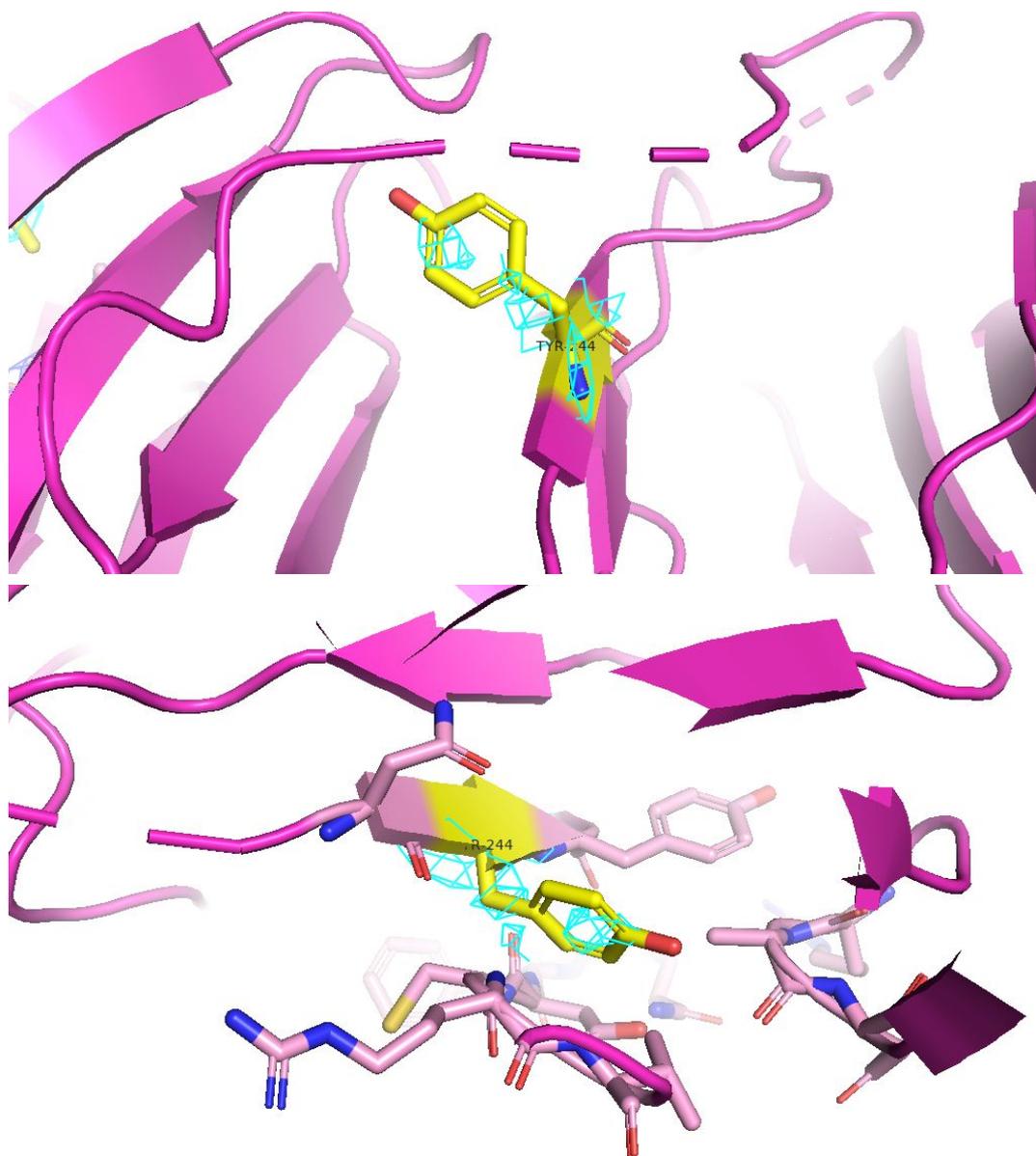


Рис 8. визуализация остатков B:TYR-244 и его электронная плотность при $carve=1.5$.

В целом эта структура вгоняет в уныние, так как совсем непонятно, как расположить в этом “кармане” из двух листов и петель.

Скорее всего, недопустимый ротамер обуславливается тем же.

A:GLN-234
- RSRZ 5,9

Маргинальность по RSRZ говорит о том, что этот остаток плохо вписывается в электронную плотность, что можно пронаблюдать на рисунке ниже.

Этот остаток находится на выходе из листа, поэтому, возможно, он в большей степени участвует в формировании этого листа, чем предполагает исследуемая модель.

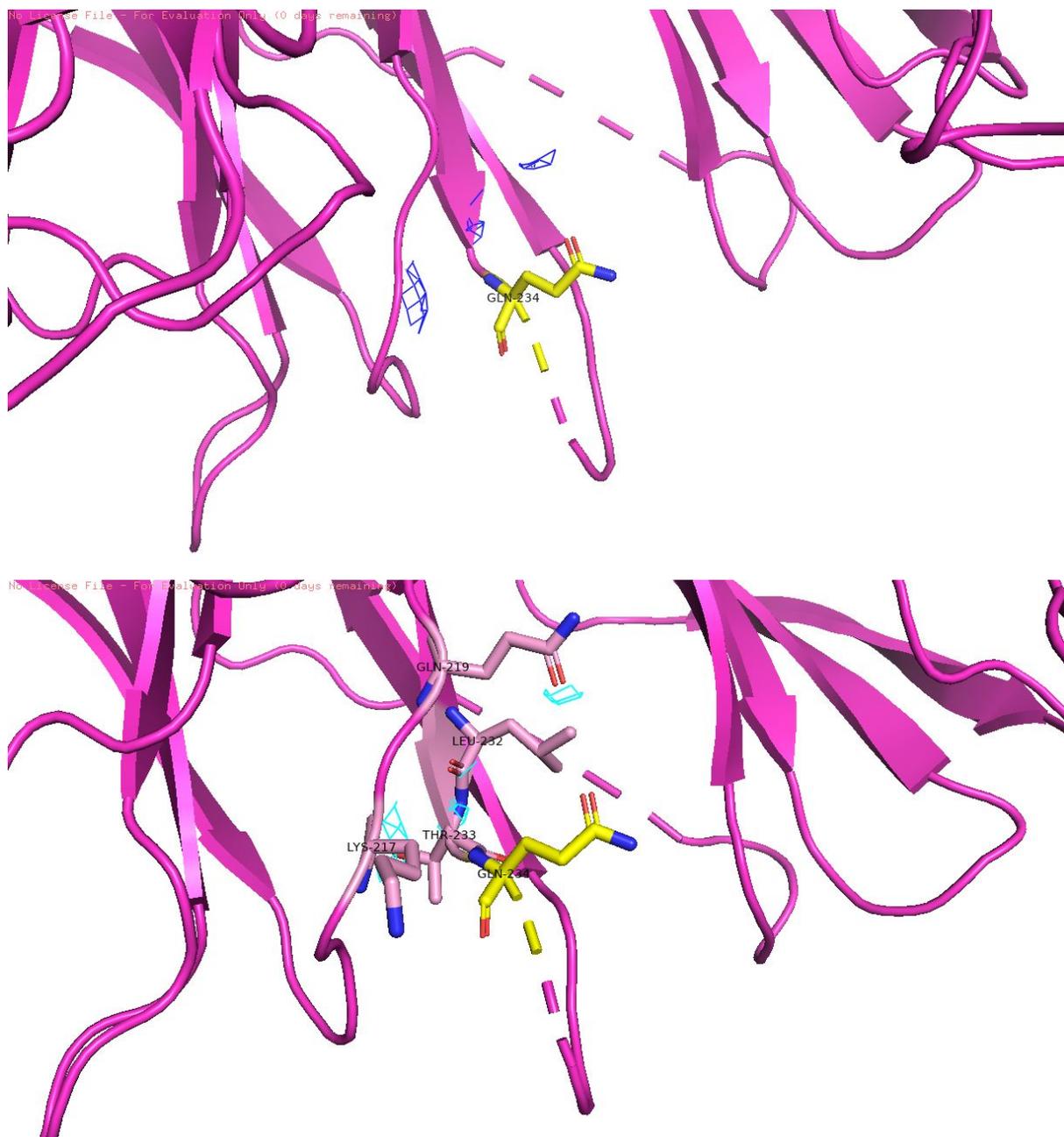


Рис 9. визуализация остатка A:GLN-234 и его электронная плотность при $carve=4$ (сверху ЭП синего цвета), окружение остатка относительно электронной плотности (снизу ЭП бирюзового цвета).

Сравнение с PDB-REDO

PDB-REDO – это база данных, где содержатся оптимизированные структурные модели (относительно моделей из банка PDB).

Модель из PDB-redo имеет R-фактор = 0.203 (против 0.238 в PDB), и R-free = 0.2539 (против 0.258 в PDB). Можно заметить, что оба параметра немного уменьшились, однако сложно сказать, является ли такое изменение весомым.

Интересно, что PDB-REDO счел R-free в PDB “предвзятым”, поэтому использовал новый (“the estimated unbiased”) R-free = 0.3095 (R-фактор тоже поменялся, но, на мой взгляд, незначительно - до 0.2439).

Качество параметра “ротамеров” упало.

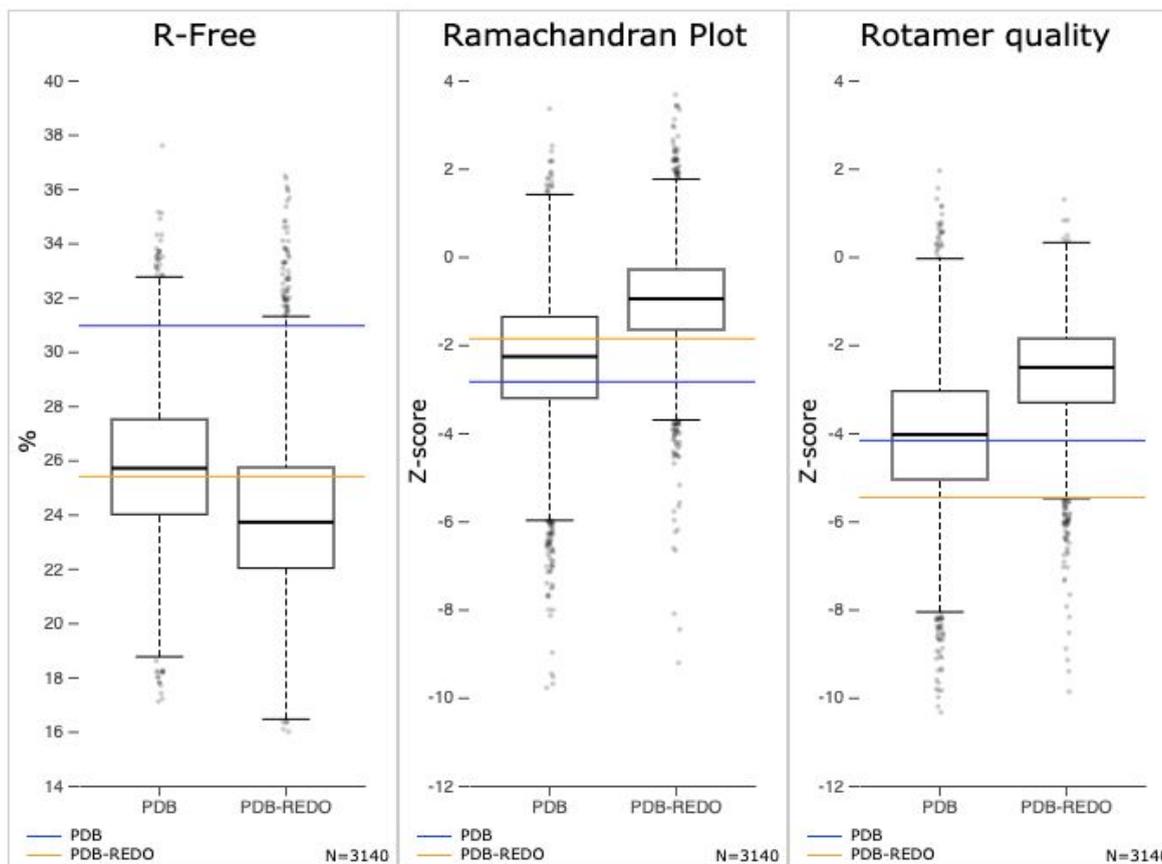


Рис 10. Сравнение показателей качества моделей в PDB и PDB-REDO

Заключение

В данной работе была рассмотрена структура eIF2A из базы данных PDB (3WJ9) и проведен анализ ее качества.

Рассматриваемая структура имеет достаточное разрешение и полноту данных, модель достаточно хорошо оптимизирована. Разница между моделями из PDB и PDB-REDO, на мой взгляд, неоднозначна.

При этом показатели модели довольно хороши, нет лишней воды, но если большое количество перекрываний. С учетом возможности отладки структуры вручную, модель вполне рабочая.

Список литературы

1. Crystal structure of the eukaryotic translation initiation factor 2A from *Schizosaccharomyces pombe* // Kashiwagi, K., Ito, T., Yokoyama, S. (2014) J Struct Funct Genomics DOI: 10.1007/s10969-014-9177-y
2. Запись PDB: <https://www.rcsb.org/structure/3WJ9>
3. Сервис MolProbity: <http://molprobity.biochem.duke.edu/>
4. PDB-redo: <https://pdb-redo.eu/db/3wj9>
5. UniProtKB P53235 (eIF2A_yeast) <https://www.uniprot.org/uniprot/P53235>
6. Komar AA, et al. (2005) Novel characteristics of the biological properties of the yeast *Saccharomyces cerevisiae* eukaryotic initiation factor 2A. J Biol Chem 280(16):15601-11