

МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
имени М.В.ЛОМОНОСОВА

ФАКУЛЬТЕТ БИОИНЖЕНЕРИИ И БИОИНФОРМАТИКИ

Оценка качества расшифровки структуры РНК-распознающего
мотива 65 кДа субъединицы фактора сплайсинга U2AF из
Homo sapiens (PDB id: 5W0H)

Отчёт выполнен студенткой 4-го курса
Буяновой Мишель Фёдоровной

Москва

2019

ОГЛАВЛЕНИЕ

ОГЛАВЛЕНИЕ	1
СПИСОК СОКРАЩЕНИЙ	2
АННОТАЦИЯ	3
ВВЕДЕНИЕ	4
РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ	6
Общая информация о модели	6
Анализ маргинальных остатков	12
Сравнение структур из PDB и PDB-REDO	18
ЗАКЛЮЧЕНИЕ	23
СПИСОК ЛИТЕРАТУРЫ	25

СПИСОК СОКРАЩЕНИЙ

PDB — Protein Data Bank

U2 snRNP — малый ядерный рибонуклеопротеиновый комплекс U2 (от *англ.* U2 small nuclear RNA)

U2AF — вспомогательный фактор U2 (от *англ.* U2 small nuclear RNA auxiliary factor)

RRM2 — РНК-распознающий мотив (от *англ.* RNA-recognition motif)

РСА — рентгеноструктурный анализ

EDS — Electron Density Server at Uppsala University

ЭП — электронная плотность

RSR — от *англ.* real-space R-value

I. АННОТАЦИЯ

Данный отчет представляет собой анализ качества расшифровки белка, а именно — РНК-распознающего мотива 65 кДа субъединицы фактора сплайсинга U2AF из *Homo sapiens*. Приведена общая информация о рентгеноструктурном эксперименте, указаны значения индикаторов качества модели в целом. Проведено исследование на маргинальность по многим параметрам, связанным как с остовом белка, так и с отдельными боковыми цепями аминокислотных остатков. Также была осуществлена процедура оптимизации средствами PDB-REDO и сделан анализ, систематизация и интерпретация результатов.

II. ВВЕДЕНИЕ

Для созревания большинства мРНК необходимо протекание такого процесса как сплайсинг, суть которого состоит в вырезании некодирующих частей, — интронов, — из прекурсорной мРНК, что является важным этапом в посттранскрипционной регуляции генной экспрессии. Сплайсинг как многоэтапный динамический процесс осуществляется с привлечением множества белок-белковых комплексов и рибонуклеопротеинов. На **Рис. 1** представлен сплайсосомальный E-комплекс (от *англ.* early), который ответственен за узнавание интрона, и соответствующие консенсусные последовательности РНК, с которыми связываются элементы сплайсосомы. В их число входит U2AF, — так называемый вспомогательный U2 фактор (*англ.* U2 auxiliary factor), состоящий из большой и малой субъединиц, называемых U2AF2 (65 кДа) и U2AF1 (35 кДа), соответственно.

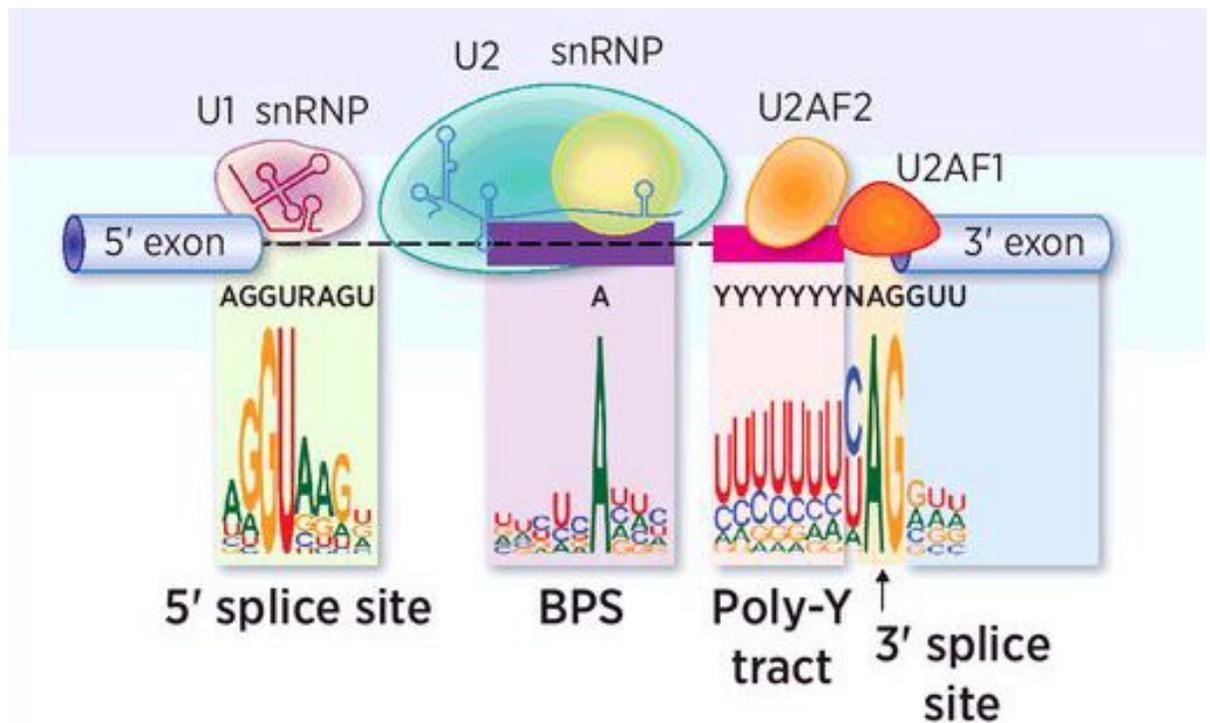


Рис.1. Устройство раннего E-комплекса сплайсосомы.

Взято с изменениями из [1]

U2AF1 связывает динуклеотид AG, располагающийся в месте 3'-сайта сплайсинга, а U2AF2 — предшествующий ему полипиримидиновый тракт. Таким образом, U2AF-гетеродимер способствует узнаванию 3'-сайта разрезания U2-зависимой сплайсосомой, которая отвечает за процессинг пре-мРНК в 95% случаев [1].

Исследования геномов людей с онкологическими заболеваниями выявили наличие рекуррентных (то есть одинаковых и возникших несколько раз независимо) мутаций в генах, кодирующих факторы сплайсинга [2]. Такие мутации, очевидно, могут приводить к неправильному процессингу пре-мРНК за счет неточного считывания экзон-интронной структуры гена, в результате чего нарушится правильный синтез его продукта. В случае, например, если пре-мРНК кодирует опухолевый супрессор, такие мутации могут привести к развитию рака [3].

Рассматриваемый в данной работе белок представляет собой RRM2 (от *англ.* RNA-recognition motif — РНК-распознающий мотив) фактора U2AF2. Его структура в высоком разрешении была получена методом РСА (рентгеноструктурного анализа) с целью сопоставить ассоциированные с раком мутации и пространственное расположение затронутых ими аминокислотных остатков и выяснить, таким образом, функциональную роль U2AF2 в развитии опухолей [4].

II. РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

2.1. Общая информация о модели

Структура RRM2 из U2AF2 (PDB id: 5W0H) представляет собой единственный белок и была получена методом РСА в 2017 году группой исследователей из Медицинского центра Рочестерского университета [4]. Заявленное разрешение структуры составляет 1.11\AA . Разброс разрешения — от 20.69 до 1.11\AA по данным авторов. По данным же EDS (Electron Density Server at Uppsala University) он составляет от 36.36\AA до 1.11\AA . При этом полнота данных о структурных факторах обладает величиной в 96.7% .

Для решения фазовой проблемы использовался метод молекулярного замещения с привлечением ранее разрешенной структуры U2AF2 (PDB id: 2G4B, заявленное разрешение 2.5\AA). Данная процедура проводилась с помощью программного обеспечения Phaser [5]. На **Рис. 2** представлено структурное выравнивание, демонстрирующее степень сходства структур и подчеркивающее удачный выбор 2G4B в качестве первого приближения для построения модели. RMSD выравнивания составило 0.56\AA .

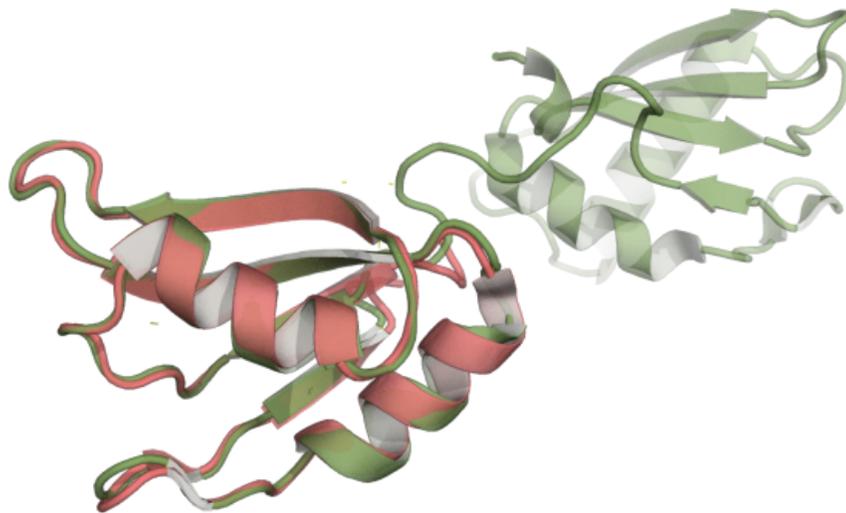


Рис.2. Выравнивание 5W0H со структурой 2G4B.
Розовым цветом показана 5W0H, зеленым — 2G4B.

Кристаллографическая группа ячейки в нотации Германа-Могена — “ $P2_12_12_1$ ”. Параметры кристаллографической ячейки, представляющей собой прямой параллелепипед, приведены в **Таблице 1**.

Таблица 1. Параметры кристаллографической ячейки

a	b	c	α	β	γ
33.43 Å	38.20 Å	72.72 Å	90°	90°	90°

Некристаллографические симметрии отсутствуют. Разрешенные операции симметрии включают в себя:

- $-X+1/2, -Y, Z+1/2$
- $X, Y+1/2, -Z+1/2$
- $X+1/2, -Y+1/2, -Z$

Восстановленный с помощью инструмента The PyMOL Molecular Graphics System [6] кристалл изображен на **Рис. 3**. В одной ячейке содержится 4 асимметрические единицы.

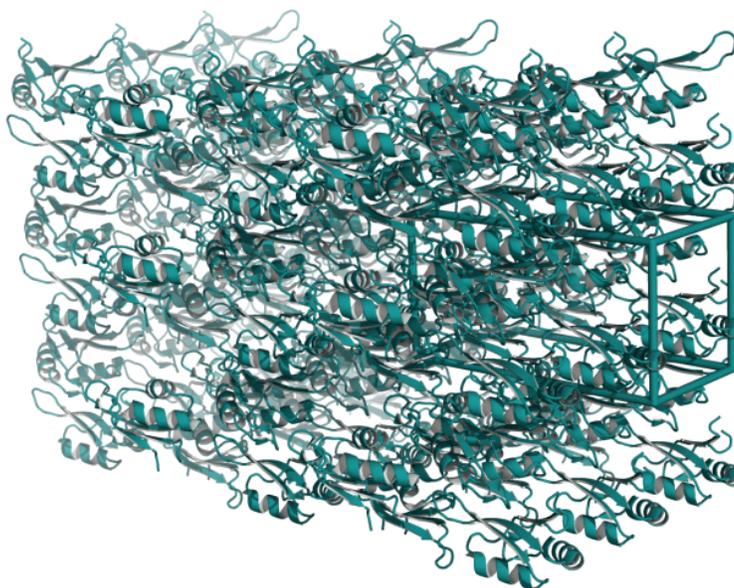


Рис.3. Структура кристалла 5W0H.
Обведены границы одной элементарной ячейки кристалла.

Асимметрическая единица, содержащая одну полимерную цепь, совпадает с биологической сборкой как сообщается в PDB [7]. При этом экспериментальной проверки данного факта не проводилось: это отмечено в поле "Biological Assembly Evidence". На **Рис. 4** приведена одна единица в окружении соседей по кристаллу.

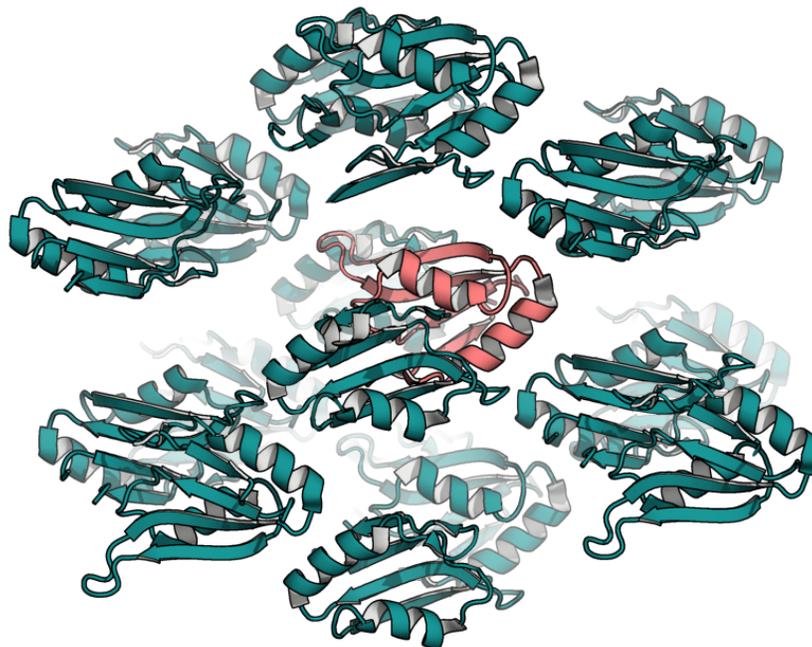


Рис.4. Асимметрическая единица в окружении ближайших соседей.
Розовым цветом выделена асимметрическая единица.

Число измеренных рефлексов в рентгеноструктурном эксперименте составило 36104. R-фактор, метрика, характеризующая соответствие модели экспериментальным данным, составила 0.153, что является примером хорошего значения. 1938 (5.37%) рефлексов были предварительно отложены для валидации оптимизационного процесса. Величина R_{free} , вычисленная по этим контрольным рефлексам, составила 0.168, что лишь немногим выше R-фактора. Данное наблюдение говорит в пользу отсутствия серьезной переоптимизации модели.

В ходе дальнейшей работы использовался сервис MolProbity [8], предназначенный для решения широкого спектра задач при валидации структур для оценки качества модели как на глобальном, так и на локальном уровнях. На первом этапе при обработке сервисом структуры была

произведена дотройка атомов водорода инструментом Reduce, в котором при разрешена операция инверсии таких остатков как аспарагин, глутамин и гистидин. Но в ходе работы Reduce не совершал таких перестроек, что может являться показателем изначально правильной ориентации этих остатков в структуре.

Построенные сервисом MolProbity карты Рамачандрана приведены на **Рис. 5.**

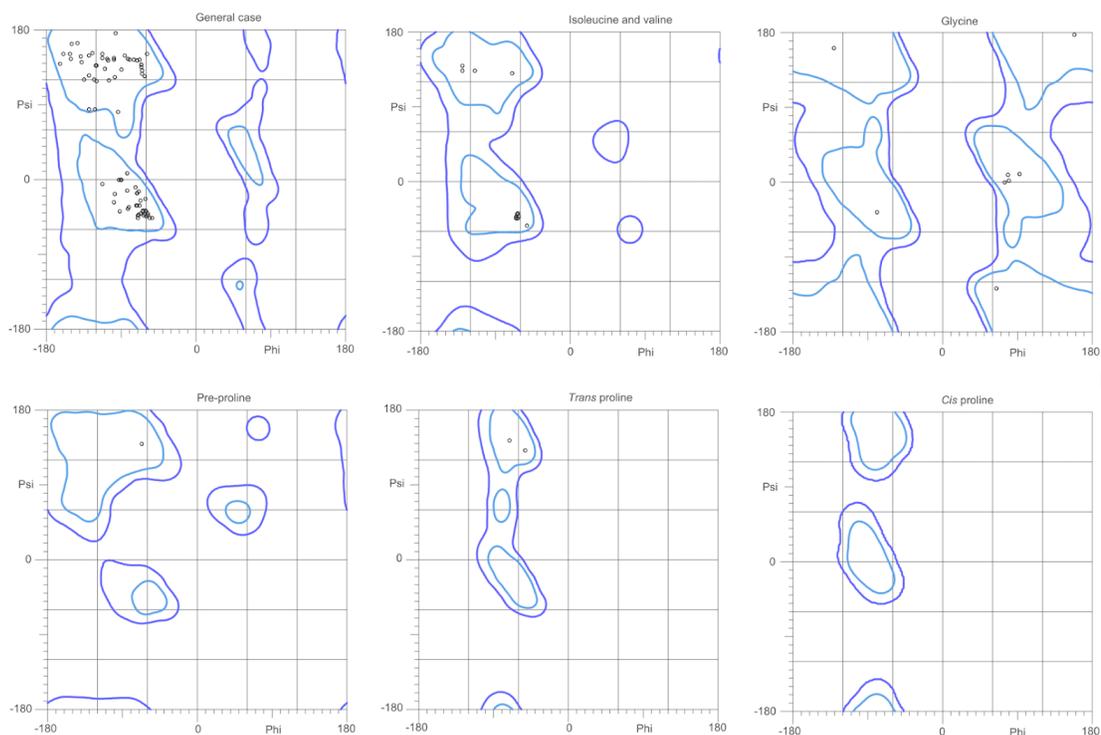


Рис.5. Карты Рамачандрана из сервиса MolProbity.

Синим: разрешенные (*allowed*) области, голубым: благоприятные (*favored*).

Верхний ряд: общая карта, *Ile+Val, Gly*;

Нижний ряд: *pre-Pro, cis-Pro, trans-Pro*.

Как можно видеть, абсолютно все остатки попадают в благоприятные регионы карт. Выбросов по областям Рамачандрана (от *англ.* outliers) не обнаружено (0.0%).

Clashscore, — число значительных стерических наложений ($> 0.4 \text{ \AA}$) на 1000 атомов, — составило 5.46, и данное значение представляет собой 73-ий перцентиль в распределении Clash-скоров, посчитанных для структур с

подобным разрешением ($1.11\text{\AA} \pm 0.25\text{\AA}$). Получается, это достаточно хороший результат: стерических пересечений меньше, чем в 73% подобных структур.

На **Рис. 6** приведены примеры всех таких Clash-перекрытий, выявленных PDB.

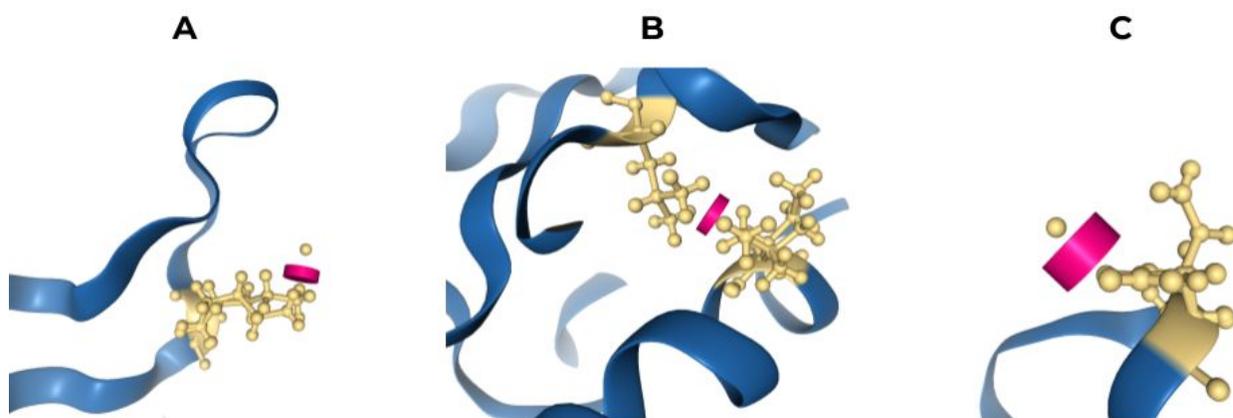


Рис.6. Clash-наложения, выявленные PDB.

Желтым: проблемные места структуры; розовым: репрезентация наложения; синим: остальная структура.

A: Clash между Lys300 и HOH406; B: Clash между Leu261 и Ile317;

C: Clash между Asn311 и HOH401.

Визуализировано при помощи [9]

Интересно отметить, что, в то время как PDB находит 3 clash-наложения, в расширенном отчете, сгенерированном в MolProbity, их содержится больше, и их полный список приведен в **Таблице 2**. В то же время там не описан clash, показанный на **Рис. 6, C**. Префикс перед трехбуквенным обозначением аминокислоты указывает на определенное альтернативное положение (от *англ.* altloc, alternative location).

Таблица 2. Clashlist из сервиса MolProbity

Остаток1	Атом1	Остаток2	Атом2	Расстояние (Å)	Показан в PDB
LEU 261	HD21	A-ILE 317	CG1	:0.496	
LEU 261	HD22	A-ILE 317	HG12	:0.472	
B-LYS 286	HE2	HOH 412	O	:0.455	
LEU 261	CD2	A-ILE 317	CG1	:0.442	+

A-ILE 317	HD11	A-ARG 334	HG2	:0.429	
LEU 261	HD21	A-ILE 317	HD11	:0.428	
B-LYS 300	NZ	HOH 406	O	:0.417	+
LEU 261	HD21	A-ILE 317	HG13	:0.407	

Интересно отметить, что остаток Ile317, участвующий в наложении, указан тут в альтернативном положении “А”. Следуя логике, altloc присваивается в порядке понижения occupancy, но в данном случае это не так: у A-Ile occupancy=0.48, а у B-Ile — 0.52.

Оказывается, в Full Validation Report от PDB (доступен по ссылке https://files.rcsb.org/pub/pdb/validation_reports/w0/5w0h/5w0h_full_validation.pdf) содержатся все записи, указанные в таблице выше, но инструмент 3D-визуализации на сайте PDB [10] настроен исключительно на отображение clash-наложений, не учитывающих атомы водорода (отображенные в PDB отмечены “+” в последней колонке **Таблицы 2**). Выяснилось также, что “утерянный” в отчете MolProbity clash между Asn311 и HOH401, тоже связан с альтернативными положениями. Наложение отмечено именно у B-Asn311, у которого occupancy составляет 0.43, и, следовательно, преимущество над ним имеет A-Asn311, который, по всей видимости, не создает clash-проблемы и потому не указан в MolProbity.

Ротамеры — типичные конформации боковых цепей остатков. По базе данных PDB было составлено представление о распределениях значений χ_i торсионных углов для разных остатков. Таким образом, каждый остаток структуры может быть оценен по соответствию конформации боковой цепи. В категорию предпочитаемых (от *англ.* favored) ротамеров попал 71 остаток, тогда как плохих (от *англ.* poor) — 3. Оба эти числа не проходят порог качественной структуры, если опираться на пороги, данные в MolProbity. Предпочитаемых ротамеров должно быть более 98%, в то время как в исследуемой структуре выходит лишь 91%. Порог содержания для плохих ротамеров в MolProbity составляет менее 0.3%, у нас же выходит целых 3.85%.

Объединив показатели по Рамачандрану, Clash-скорам и ротамерам, в MolProbity ввели собственную, интегральную, оценку качества структуры. MolProbity score рассматриваемой структуры составил 1.74, и это значение находится в 43-ем перцентиле распределения данной оценки для схожих по разрешению структур. Полученный результат может указывать на не вполне хорошее качество структуры, основанное на плохих оценках для ротамеров боковых цепей.

Однако также по критериям MolProbity не обнаружено ни некорректных связей, ни некорректных углов, ни значительных ($>0.25\text{\AA}$) отклонений в положениях $C\beta$ атомов. Это говорит в пользу хорошего качества структуры.

2.2. Анализ маргинальных остатков

Обратим внимание на плохие ротамеры, которые были причиной занижения интегрального MolProbity-скора.

1. Ser257

Наблюдаемый ротамер не соответствует представлениям о том, как должна быть расположена боковая цепь серина в белке, поскольку этот остаток располагается в начале цепи (см. **Рис. 7**).

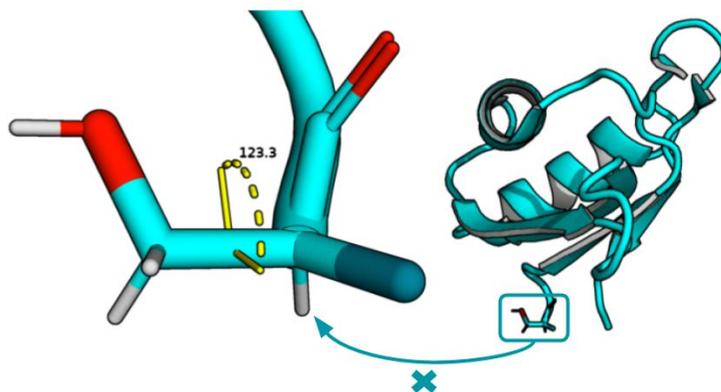


Рис.7. Ротамер Ser257

2. Ser336

Ситуация, аналогичная первому пункту. Указанный остаток является последним в цепи, поэтому-то и вполне ожидаемо свободен по конформации. Единственным отличием в данном пункте является наличие альтернативных расположений: у A-Ser336 совершенно неприемлемый $\chi = 354.1^\circ$, а у B-Ser336 χ -угол поменьше, 267.3° , и он-то уже попадает в разрешенную (от *англ.* allowed) область. Однако, стоит отметить, что в PDB Full Validation Report B-Ser336 также считается выбросом с плохим ротамером.

3. B-Lys286

Здесь опять-таки встречается явление альтернативных положений остатков. В Таблице 3 приведены χ -углы Lys286 в положениях “А” и “В”.

Таблица 3. Торсионы боковых цепей Lys286

altloc	occupancy	χ_1	χ_2	χ_3	χ_4	ротамер
A	0.49	192.4°	152.8°	185.8°	175.6°	предпочтительный
B	0.51	85.8°	184.8°	83°	124.6°	плохой (outlier)

Стоит вспомнить, что B-Lys286 также был отмечен в clash-наложении с молекулой воды. Так что, несмотря на его формально бóльшую по значению occupancy, я бы также, как и авторы, присвоила флаг “А” лизину с occupancy в 0.49. Он представляет собой хороший, предпочтительный ротамер и не замечен в наложениях. Оба ротамера и окружающая их электронная плотность представлены на Рис. 8.

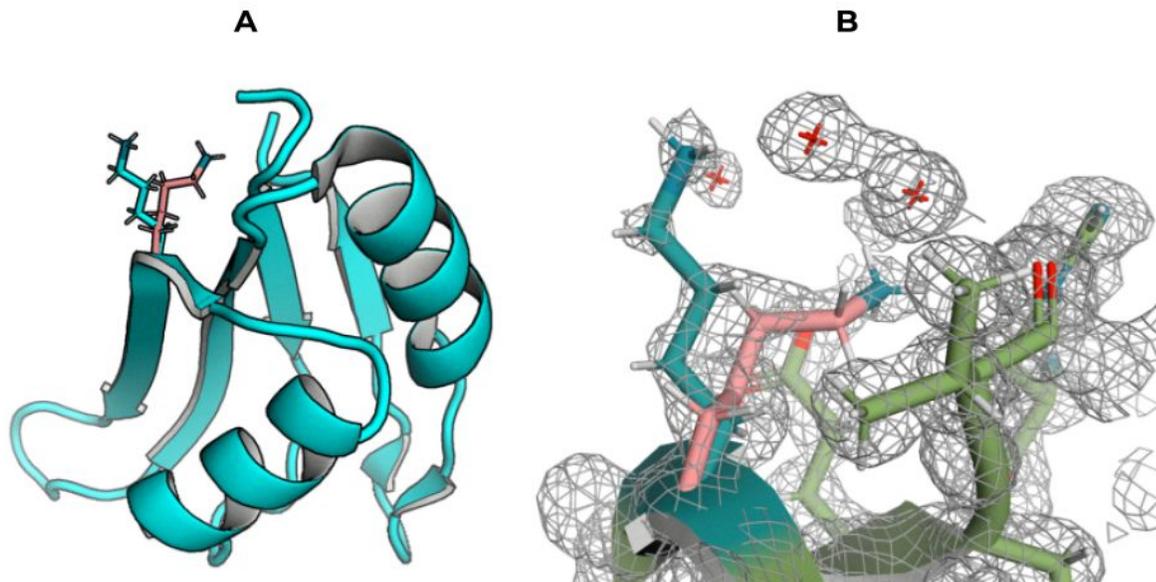


Рис.8. Альтернативные положения Lys286.

А: Общий вид белка, отдельно отрисованы А- и В-Lys286 (последний выделен розовым); В: Отображение электронной плотности на уровне подрезки 1.0 вокруг Lys286 (синим — А-Lys286, розовым — В-Lys286, зеленым — соседствующие в пространстве остатки).

Перейдем к рассмотрению другого показателя качества для отдельного остатка, коим является RSR — пространственный R-фактор, определяющий соответствие электронной плотности, выявленной в эксперименте, и электронной плотности конечной модели. По RSR можно делать выводы о том, насколько хорошо остаток вписывается в электронную плотность.

Для выделения маргинальных остатков используется RSRZ — попросту Z-скор конкретного RSR в распределении RSR на остатках того же типа из структур со схожим показателем разрешения. Остатки с RSRZ, превышающим 2, можно считать маргиналами.

На **Рис. 9**, частично заимствованном с изменениями из PDB Full Validation Report, показана суммарная информация о качестве структуры и ее вписании в электронную плотность. Для наглядности также приведено изображение структуры 5W0H с выделенными маргинальными по RSR остатками.

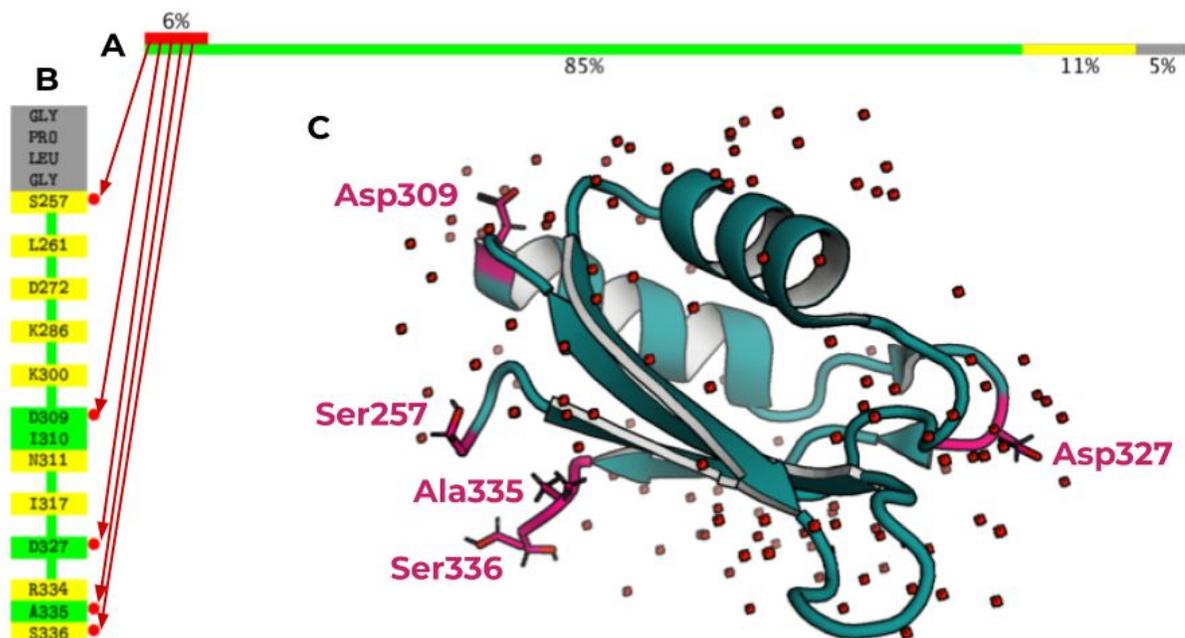


Рис.9. Суммарная информация о качестве остатков структуры 5W0H

A: Процентное соотношение остатков с различным числом нарушенных геометрических критериев;

B: Ход последовательности с выделенными проблемными остатками; (зеленым: 0 нарушенных критериев, желтым: 1 критерий; серым: не встроенные в структуру остатки; красный: маргиналы по RSR);

C: Структура белка 5W0H с выделенными розовым остатками-маргиналами по RSR.

Итак, на основе этой репрезентации можно заключить, что 85% структуры не вызывает подозрений с точки зрения геометрических критериев. Однако, стоит обратить внимание, что 6% остатков обладают высоким RSR. Суммарная информация о маргинальных остатках структуры приведена в **Таблице 4**.

Таблица 4. Маргинальные остатки структуры 5W0H

#	Остаток	Критерий
1	Ser257	плохой ротамер, RSRZ > 2
2	Leu261	участие в clash-наложении
3	Asp272	плохой ротамер*
4	Lys286	плохой ротамер

5	Lys300	участие в clash-наложении
6	Asp309	RSRZ>2
7	Asn311	участие в clash-наложении
8	Ile317	участие в clash-наложении
9	Asp327	RSRZ>2
10	Arg334	участие в clash-наложении
11	Ala335	RSRZ>2
12	Ser336	плохой ротамер, RSRZ>2

* Asp272 не был выявлен MolProbity как плохой ротамер, ибо попадает в разрешенную (от англ. allowed) область, но, видимо, в ее крайний регион.

Прежде чем перейти к анализу RSR-маргиналов, стоит отметить интересную особенность. Первый и последний остатки цепи (Ser257 и Ser336), и только они, стали “двойными” маргиналами: и по ротамерному критерию, и про RSR. Это может быть особенностью крайних участков белка, особенно в свете того факта, что рассматривается в качестве структуры небольшой мотив, изъятый из структуры фактора сплайсинга.

Таким образом, в белке выявлено 3 остатка (Asp309, Asp327, Ala335) приемлемой геометрии, которые обладают RSRZ>2. Далее будет приведен анализ каждого из этих случаев.

Стоит отметить то, что остаток Ala335 представлен в структуре в двух альтернативных расположениях (А и В). На **Рис. 10** приведен остаток Ala335 и окружающая его электронная плотность на разных уровнях подрезки.

Обращает на себя внимание неодинаковое поведение с увеличением уровня подрезки разных альтернативных расположений остатка. Так, А-Ala335 раньше теряет электронную плотность вокруг боковой цепи.

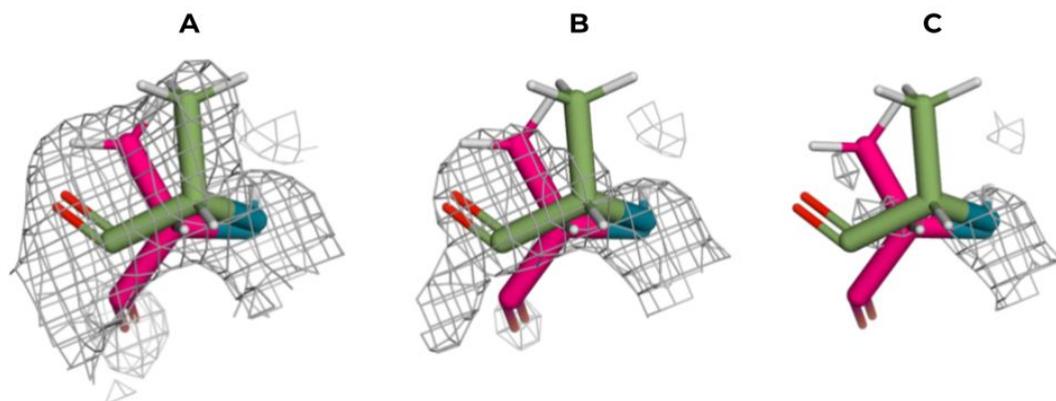


Рис.10. Ala335 и окружающая его электронная плотность

В порядке увеличения уровня подрезки А: 1.0, В: 2.0, С: 3.0; Зеленым показано альтернативное расположение $altloc=A$, розовым — $altloc=B$.

Но и В-Ala335 не очень устойчив к повышению числа σ : на уровне 3.0 он тоже не окружен электронной плотностью. Наблюдаются какие-то обрывочные остатки ЭП (электронной плотности), ничему не соответствующие. На самом деле, у хорошо вписанных аланинов ЭП вокруг $C\beta$ атома сохраняется и при гораздо более высоких (4-5) уровнях подрезки.

Также маргиналами по RSR являются два аспартата — Asp309 и Asp327. Их ЭП на разных уровнях подрезки приведена на **Рис. 11**. Вдобавок для сравнения будет проиллюстрирован остаток Asp314, который хорошо вписывается в ЭП и не является маргиналом по RSR. Достаточно наглядно из этого рисунка видно, что при нормальном вписании аспартат сохраняет ЭП вокруг обоих кислородных атомов карбоксильной группы вплоть до высоких значений уровня подрезки.

Asp309 же на уровне 3.5σ теряет ЭП вокруг одного из кислородов. А в случае с Asp327 один из кислородов не располагается внутри поверхности ЭП даже на уровне 0.5σ : мы лишь видим вблизи него некое облако, в которое он не вписывается. На уровне 2.0σ с карбоксильной группы ЭП уходит совсем. Данные наблюдения вполне согласуются с заключением о том, что Asp309 и Asp327 действительно являются маргиналами, и, как следствие, недочетами расшифровки структуры.

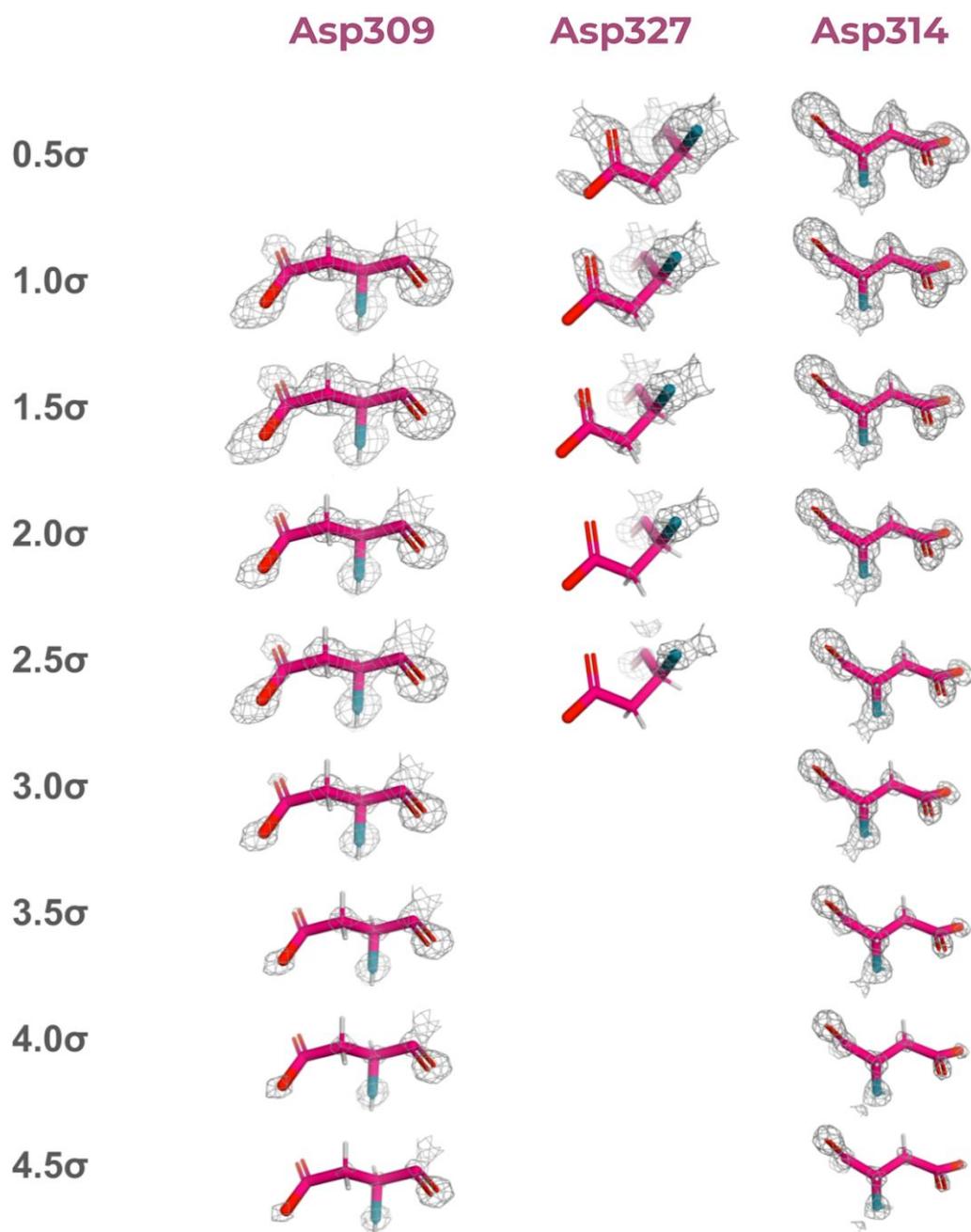


Рис.11. Asp309, 327 и 314 и окружающая их электронная плотность

2.3. Сравнение структур из PDB и PDB-REDO

PDB-REDO — это сервис, позволяющий провести процедуру оптимизации моделей кристаллографических структур, привлекая к процессу множество специализированных программ [11]. В ходе оптимизации

происходит изменение параметров, улучшающих модель, а также имеют место частичные перестройки структуры (такие как, например, удаление молекул воды и исправление боковых цепей остатков). По завершении работы PDB-REDO выдает файл с новой моделью и автоматически сгенерированный отчет о ее валидации.

Сравнение моделей PDB и PDB-REDO приведено на **Рис. 12**. Даже на общем плане заметно, что в некоторых местах действительно произошла перестройка. Самый яркий пример отмечен стрелкой.

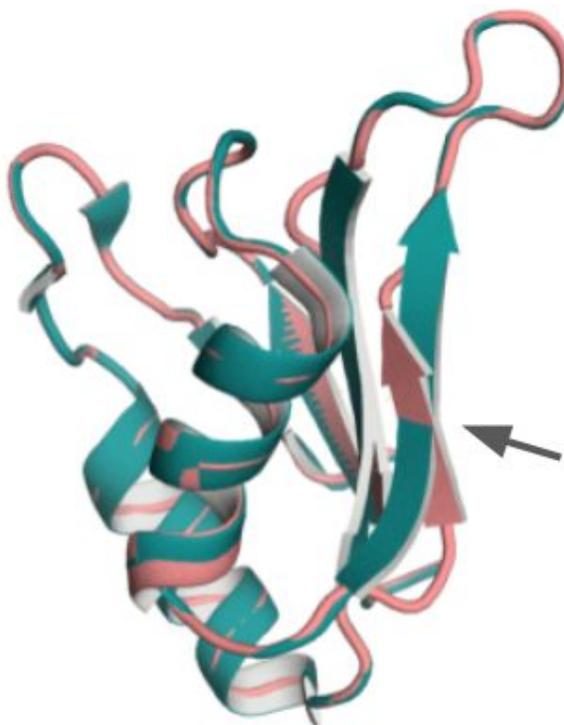


Рис.12. Сравнение моделей PDB и PDB-REDO

Голубым: исходная PDB модель, розовым: оптимизированная PDB-REDO модель; стрелка указывает на место перестройки.

После оптимизации изменились факторы R и R_{free} : они немного увеличились. R -фактор стал составлять 0.165 (ранее — 0.163), а R_{free} — 0.174 (ранее — 0.172 по данным PDB-REDO, противоречащим ранее указанным в PDB: 0.168). Таким образом, можно сказать, что модель

действительно стала лучше соответствовать экспериментальным данным, однако, повысилась вероятность переоптимизации.

Из значительных изменений, внесенных PDB-REDO, стоит отметить удаление 28 молекул воды. Далее, 14 остатков, как сообщается, стали вписаны в ЭП лучше (среди улучшений вписывания в ЭП стоит отметить таковое у Ser257, Thr296), 13 же остатков — хуже. Для Ser257 улучшение проиллюстрировано на **Рис. 13**. Он перестал быть маргиналом по ротамеру.

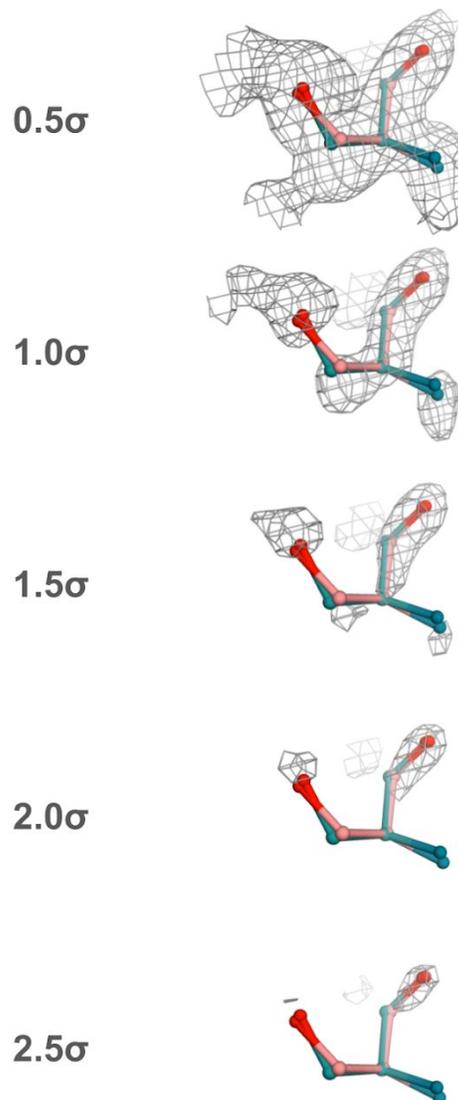


Рис.13. Улучшение вписывания Ser257 в ЭП

Голубым: исходная PDB модель, розовым: оптимизированная PDB-REDO модель.

Интересно, что по одному из параметров, которые сообщает PDB-REDO, ситуация кардинально ухудшилась. Это серьезность clash-перекрытий (“bump severity”). Если исходная модель была по этому параметру в 78-ом перцентиле, то теперь она находится в 20-м. Добавилось лишь одно перекрытие, но такое значительное снижение ранга вызвано тем, что у атомов, в нем участвующих, низкие B-факторы. Если бы это было не так, можно было бы предположить, что именно высокое значение B-фактора “размыло” электронную плотность, что осложнило вписывание атомов и, в конце концов, привело к перекрытию.

В качестве еще одного формата выдачи результата PDB-REDO предлагает *.sce-файл для визуализации в YASARA [12]. Эта программа позволяет раскрасить новую модель в соответствии с серьезностью сдвигов атомов, проведенных в ходе оптимизации PDB-REDO.

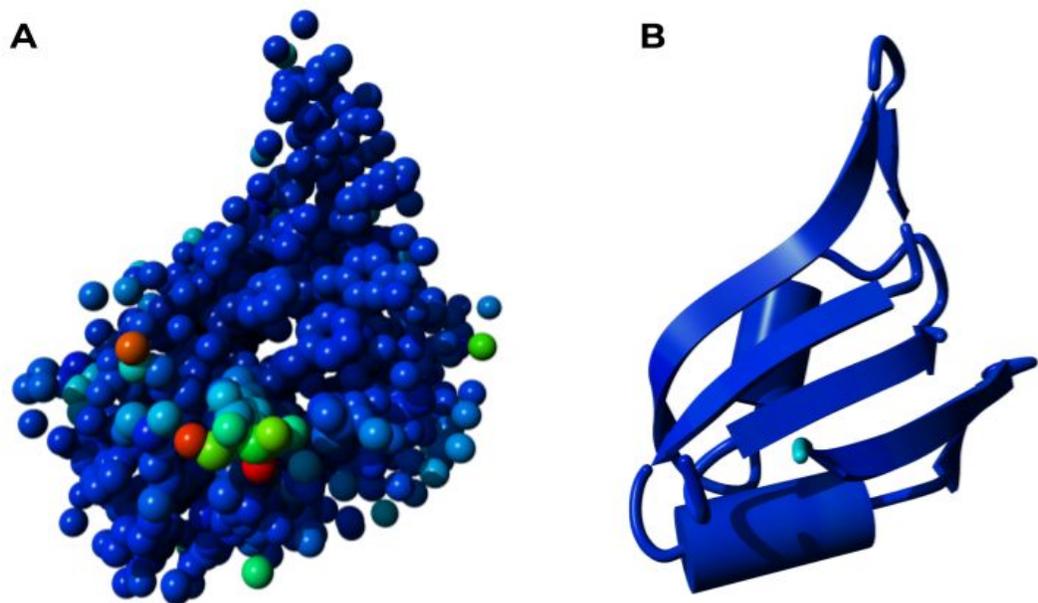


Рис.14. Результат оптимизации PDB-REDO в YASARA

А: Шариковая модель, более теплые цвета соответствуют большим сдвигам; В: Cartoon-визуализация

На **Рис. 14** приведена подобная сцена (от *англ.* scene) в YASARA. С помощью этого инструмента были выявлены наиболее затронутые оптимизацией атомы и аминокислоты: Ser336 (исходно двойной маргинал по

ротамеру и RSR), Lys260, Lys329, Met323, Gln315, Ile317 (исходно участие в clash-наложении), Lys286 (исходно плохой ротамер) и Lys276.

Разберем три наиболее интересных примера, а именно — те случаи, в которых в результате оптимизации сильно поменялся маргинальный остаток. Исправило ли это положение дел?

1. Ser336 был маргиналом по RSR. A-Ser336 был маргиналом по ротамеру, B-Ser336 попадал в разрешенную область. После PDB-REDO A-Ser336 так и остался маргиналом, а B-Ser336 попал в предпочтительную область (от *англ.* favored). Однако при этом B-Ser336 образовал clash с B-Ala335.
2. A-Ile317 участвовал в clash-наложении с Leu261. Это не исправилось. Также B-Ile317 теперь столкнулся clash-ем с HOH420.
3. B-Lys286 был плохим ротамером, участвовал в clash-наложении с HOH412. Теперь же он попал в разрешенную (от *англ.* allowed) область, но clash с атомом кислорода воды стал более серьезным.

Таким образом, сильные изменения, затронувшие исходно нехорошие атомы, в большинстве случаев не то что не исправили проблемы, а, возможно, и усугубили их.

IV. ЗАКЛЮЧЕНИЕ

1. Качество расшифровки структуры 5W0H, по моему мнению, можно считать неплохим. В пользу этого говорят значения факторов R и R_{free} ; отсутствие маргиналов по конформации остова (по Рамачандрану); отсутствие перевернутых Asn, Gln, His; Clashscore имеет значение, соответствующее 73-му перцентилю среди структур подобного качества. Неправильных ротамеров было всего 3, но MolProbity счел это плохим результатом, но лишь потому, что длина белка очень небольшая, и из-за этого процент плохих ротамеров оказывается высоким. Интегральный же скор MolProbity имел значение, попадающее в 43-ий перцентиль. Не было обнаружено ни некорректных связей, ни некорректных углов, ни значительных ($>0.25\text{\AA}$) отклонений в положениях C β атомов.
2. Судить о действительной маргинальности 3 плохих ротамеров сложно. Два из них — это первый и последний остатки цепи, чем и может быть обусловлено отклонение торсионов их боковых цепей от привычных значений. Третий же плохой ротамер (имеется в виду остаток B-Lys286) имеет альтернативное расположение A-Lys286 с occupancy в 0.49 (небольшая разница с 0.51), который не является маргиналом по ротамеру.
3. Стоит признать наличие в структуре пяти маргиналов по RSR. Опять же принимая во внимание, что два из них находятся на концах цепи.
4. PDB-REDO, вопреки ожиданиям, в целом ухудшил состояние модели, особенно в тех местах, где были произведены наибольшие сдвиги атомов. Да, стоит признать, что 14 остатков стали лучше вписаны в ЭП, в том числе начальный Ser257, переставший быть маргиналом по ротамерному критерию. Но при этом большинство проблем не было решено, у 13 остатков ухудшилось вписывание в ЭП, появилось 3 связи ненормальной длины (“worst” по данным MolProbity). Более того, добавилось clash-наложений, что привело к понижению интегрального

MolProbity-скора. После оптимизации PDB-REDO модель оказалась в 37-ом перцентиле (вместо изначального 43-го).

5. По моему мнению, неожиданный исход оптимизации PDB-REDO связан с особенностями белка, являющегося небольшим мотивом, изъятым из структуры субъединицы фактора сплайсинга.

СПИСОК ЛИТЕРАТУРЫ

1. Yoshimi A, Abdel-Wahab O. Molecular Pathways: Understanding and Targeting Mutant Spliceosomal Proteins. *Clin Cancer Res.* 2017;23: 336–341.
2. Dvinge H, Kim E, Abdel-Wahab O, Bradley RK. RNA splicing factors as oncoproteins and tumour suppressors. *Nat Rev Cancer.* 2016;16: 413–430.
3. Jung H, Lee D, Lee J, Park D, Kim YJ, Park W-Y, et al. Intron retention is a widespread mechanism of tumor-suppressor inactivation. *Nat Genet.* 2015;47: 1242–1248.
4. Glasser E, Agrawal AA, Jenkins JL, Kielkopf CL. Cancer-Associated Mutations Mapped on High-Resolution Structures of the U2AF2 RNA Recognition Motifs. *Biochemistry.* 2017;56: 4757–4761.
5. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. Phaser crystallographic software. *J Appl Crystallogr.* 2007;40: 658–674.
6. DeLano WL, Others. Pymol: An open-source molecular graphics tool. *CCP4 Newsletter on protein crystallography.* 2002;40: 82–92.
7. Bank RPD. RCSB PDB: Homepage. RCSB PDB. 2019. Available: <https://www.rcsb.org/>
8. Williams CJ, Headd JJ, Moriarty NW, Prisant MG, Videau LL, Deis LN, et al. MolProbity: More and better reference data for improved all-atom structure validation. *Protein Science.* 2018. pp. 293–315. doi:10.1002/pro.3330
9. Rose AS, Bradley AR, Valasatava Y, Duarte JM, Prlic A, Rose PW. NGL viewer: web-based molecular graphics for large complexes. *Bioinformatics.* 2018;34: 3755–3758.
10. Bank RPD. RCSB PDB - 5W0H: Structure of U2AF65 (U2AF2) RRM2 at 1.11 Angstrom Resolution. [cited 8 Dec 2019]. Available: <https://www.rcsb.org/3d-view/5W0H?preset=validationReport>
11. Joosten RP, Long F, Murshudov GN, Perrakis A. The PDB_REDO server for macromolecular structure model optimization. *IUCrJ.* 2014;1: 213–220.
12. Krieger E, Vriend G. YASARA View—molecular graphics for all devices—from smartphones to workstations. *Bioinformatics.* 2014;30: 2981–2982.