

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

ФАКУЛЬТЕТ БИОИНЖЕНЕРИИ И БИОИНФОРМАТИКИ

**Отчёт по качеству расшифровки РСА структуры белка WRN (хеликаза синдрома
Вернера, PDB код 6YHR)**

Белоусовой Евгении Александровны

Москва, 2020

Аннотация

В данном отчете я оценила качество интерпретации экспериментальных данных РСА о белке WRN в целом и в некоторых частностях.

Введение

Хеликаза синдрома Вернера (WRN) относится к семейству хеликаз RecQ, которые раскручивают ДНК в направлении 3'→5'. У больных синдромом Вернера наблюдаются разнообразные изменения организма, характерные для нормального человеческого старения: потеря волос, катаракта, остеопороз, атеросклероз, повышенный риск развития рака. На клеточном уровне тоже происходят регрессивные изменения: медленный рост, хромосомные aberrации и нестабильность генома в целом, повышенная чувствительность к повреждениям ДНК.

WRN имеет длину 1432 аминокислот, вес 162 кДа. Она состоит из двух коровых доменов (D1 и D2), которые гомологичны белку RecA *Escherichia coli*, а также трех дополнительных доменов: цинк-связывающего домена, WH-домена и HRDC-домена. WRN уникальна в семействе RecQ тем, что содержит еще домен с 3'→5' экзонуклеазной активностью на N-конце. WRN, по-видимому, участвует в процессе репарации ДНК, хотя конкретный механизм его работы изучен плохо. Помимо репарации и репликации есть данные об участии WRN в некоторых других процессах, например, в аресте репликативной вилки и некоторых специфичных типах репарации.

Недавно также появились данные о корреляции уровня экспрессии WRN с развитием рака типа «microsatellite instability-high». Похоже, что WRN участвует в развитии этих заболеваний именно как хеликаза (а не как нуклеаза). Поэтому на сегодняшний день WRN – это важная потенциальная мишень для лекарств от этих болезней [1].

Результаты и обсуждение

Я рассматривала запись PDB 6YHR. Общий вид структуры представлен на рисунке 1. Структуру получили следующие авторы: Newman, J.A.; Gavard, A.E.; Savitsky, P.; von Delft, F.; Arrowsmith, C.H.; Edwards, A.; Bountra, C.; Gileadi, O. в 2020 году. Ссылка на статью – первая в списке литературы. Белок был наработан в культуре *E. coli* с гексагистидиновым тэгом на C-конце и с сайтом протеазы вируса TEV, затем кристаллизован. В РСА-эксперименте была получена структура с разрешением 2.2 Å, а фазовая проблема решалась методом молекулярного замещения, в качестве уже известной модели использовалась структура хеликазы RecQ1 (PDBid 2WWY). Число измеренных рефлексов: 35356, полнота наборов структурных факторов: 98.4 %, диапазон разрешений структурных факторов: 2.2 - 76 Å. В состав получившегося комплекса входит сам белок (единственная цепь A), молекула АДФ, 2 атома цинка, молекулы воды. Параметры кристаллографической ячейки (Å): a = 54.595, b = 90.627, c = 138.233, alpha, beta, gamma = 90 градусов, и кристаллографическая группа P 21 21 21. Насколько я поняла, некристаллографические симметрии в структуре отсутствуют.

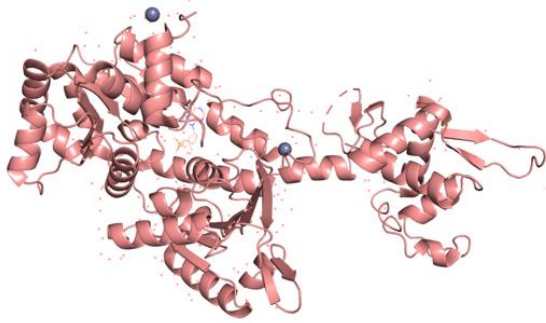


Рисунок 1. Общий вид структуры 6YHR.

Теперь рассмотрим общие индикаторы качества модели. Значение R-фактора: 0.197, а R_{free} : 0.236, при этом процент, который занимает тестовая выборка = 4.73. На рисунке 2 представлена карта Рамачандрана, 97.4% всех остатков попали в предпочтительный регион, а 99.8% - в разрешенный регион. И только один остаток оказался маргинальным.

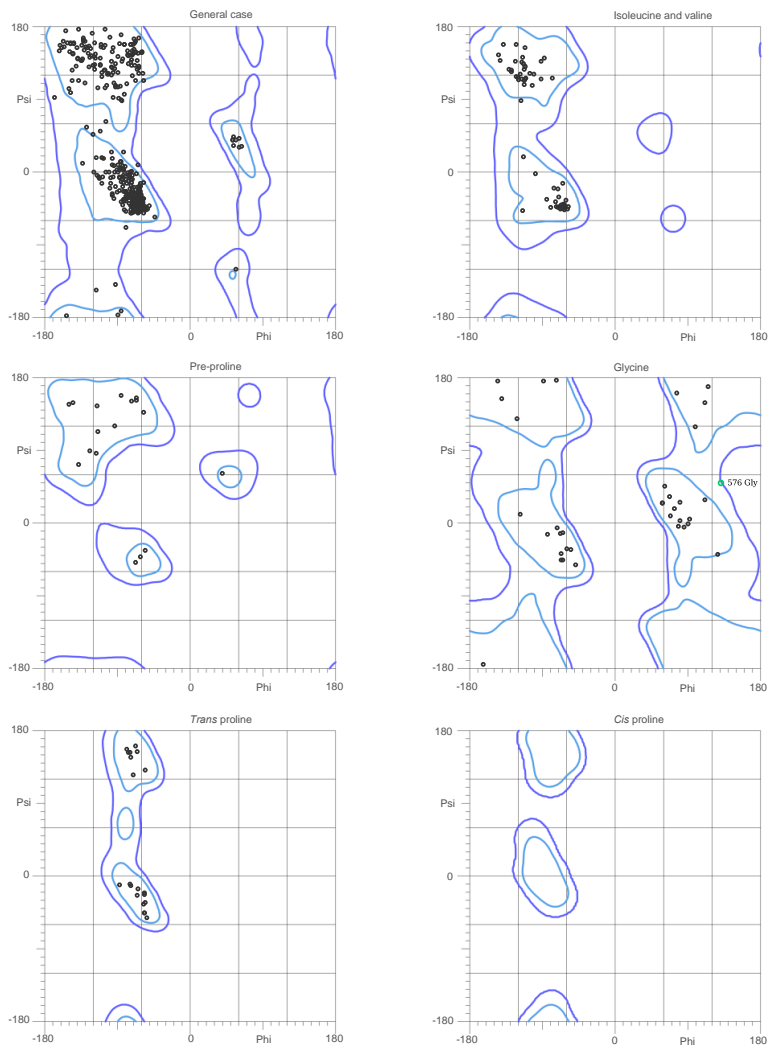


Рисунок 2. Карты Рамачандрана для некоторых случаев, полученные с помощью MolProbity [3].

В таблице 1 представлены некоторые другие индикаторы качества модели, полученные с помощью MolProbity [3].

Таблица 1.

| Название индикатора | Его смысл | Среднее значение |
|--|--|------------------|
| Clashscore | Среднее число случаев аномального ($> 0.4 \text{ \AA}$) перекрывания Ван-дер-Ваальсовых радиусов на 1000 атомов. | 3.13 |
| % молекул воды перекрыванием $> 0.4 \text{ \AA}$ | | 3.59 |
| % poor rotamers | % ротамеров с встречаемостью $< 0.3 \%$ | 4.29 |
| % favored rotamers | % ротамеров с встречаемостью $> 98 \%$ | 90.56 |
| % цис-пролинов | | 0 |
| % углов с встречаемостью $< 0.1\%$ | | 0 |

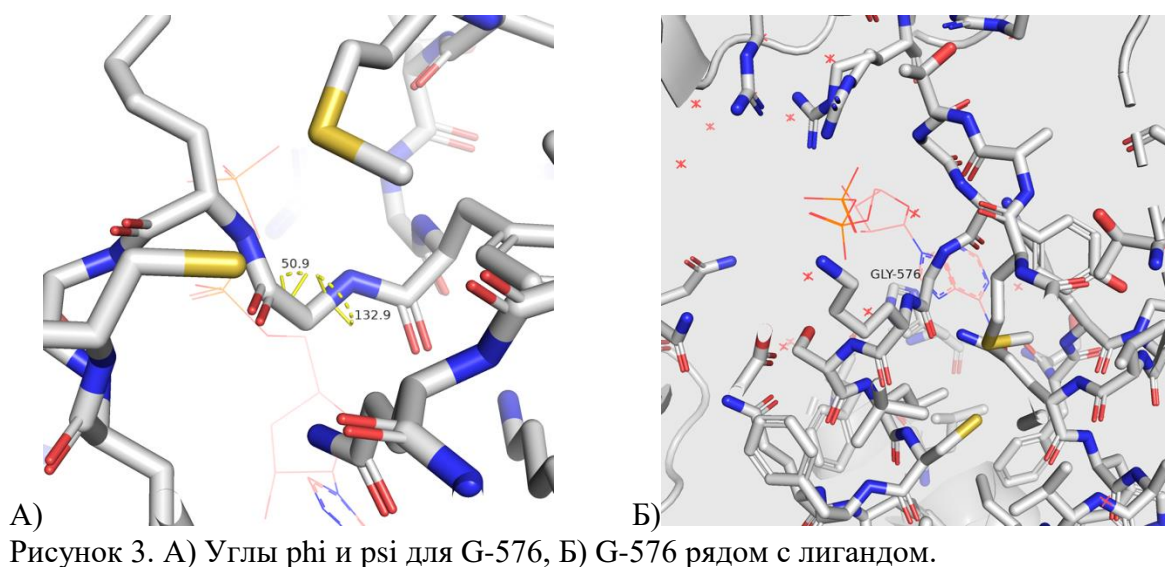
В таблице 2 перечислены маргинальные остатки, которые я отобрала по разным критериям.

Таблица 2.

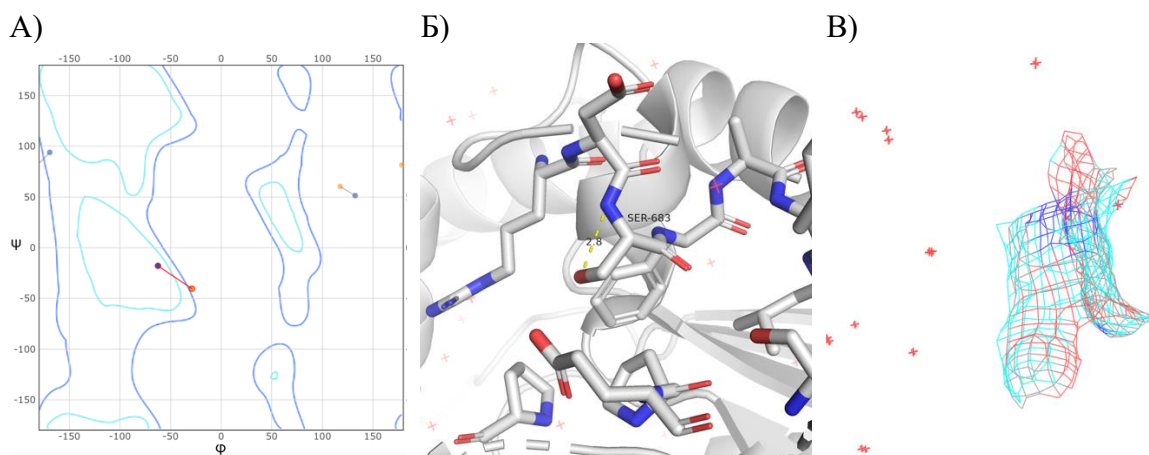
| Остаток | Критерий маргинальности |
|----------|--|
| G-576 | Аутлаер на карте Рамачандрана. |
| N-876 | |
| K-924 | |
| S-759 | |
| W-529 | Аутлаеры по значению Z-score RSR фактора. |
| K-555 | |
| R-884 | |
| C-580 | Этот остаток не выходит на поверхность и, похоже, не образует водородных связей. |
| НОН-1319 | Молекула воды, не фиксированная водородной связью. |

Далее я проведу более подробный анализ пяти маргинальных и особо подозрительных остатков и попытаюсь сделать свой вывод о том, действительно ли у остатка есть причины быть таким странным, или это, скорее, ошибка интерпретации эксперимента.

1. Для **G-576**, единственного в структуре аутлаера на карте Рамачандрана, $\phi=132.8$ градуса, $\psi=51.0$ (рисунок 3 А)). Остаток находится практически на границе приемлемых углов для глицина. Тут нетипичен, скорее, угол ϕ , будь он чуть меньше, глицин бы попал в разрешенную область. Очень заманчиво было бы приписать этому остатку какую-нибудь особую функцию, потому что он находится рядом с лигандом (рисунок 3 Б)). Но никаких интересных взаимодействий у него не обнаруживается, к тому же, в отчете PDB_redo для него получаются другие углы: 118.0 и 60.0, что уже соответствует разрешенному региону на карте (Δ RSCC), разница в RSCC исходной структуры и структуры из PDB_redo, не значимый для этого остатка). Так что, мой вывод: скорее, неправильная расшифровка, чем интересная особенность.



2. **S-683** является аутлаером по RSRZ (Z-score RSR фактора), а также аутлаером по карте Рамачандрана в выдаче PDB_redo (рисунок 4 А)). Это означает, что определенная экспериментально ЭП не очень хорошо соответствует модели. При этом в PDB_redo delta(RSCC) указывает на то, что в PDB_redo структуре ЭП остатка лучше соответствует модели, но при этом остаток становится аутлаером на карте Рамачандрана. Однако в структуре PDB_redo серин 683 визуально ничем не отличается от такового в изначальном PDB (рисунок 4 Б), В)). Есть все та же одна водородная связь ОН-группы серина с его же азотом длиной 2.8 Å. Мне кажется, что этот остаток все же, действительно, нетипичен, раз программа готова пойти на то, чтобы вынести его за пределы разрешенного региона карты Рамачандрана, компенсируя несоответствие экспериментальной ЭП модели. Но определить причину его нетипичности мне не удастся. На рисунке 4 под буквой В) показано несовпадение ЭП S-683 в исходной структуре и в структуре PDB_redo, так я попыталась продемонстрировать маргинальность остатка по RSRZ в исходной структуре. Я постаралась подобрать ракурс, с которого виднее всего различия в ЭП. Конечно, это не лучший способ визуализации, но лучшего мне придумать не удалось.



3. **D-679** также аутлаер по RSRZ, а в выдаче PDB_redo – аутлаер на карте Рамачандрана (рисунок 5 А)). И, опять же, в PDB_redo ЭП структуры значимо лучше соответствует экспериментальной (судя по $\Delta(RSCC)$). Но при этом в структуре PDB_redo 2 из 3 более или менее правдоподобных водородных связей с окружением теряются (рисунок 5 В)). Так что мой вывод: такая ЭП для D-679 – это, скорее, особенность остатка, чем ошибка расшифровки.

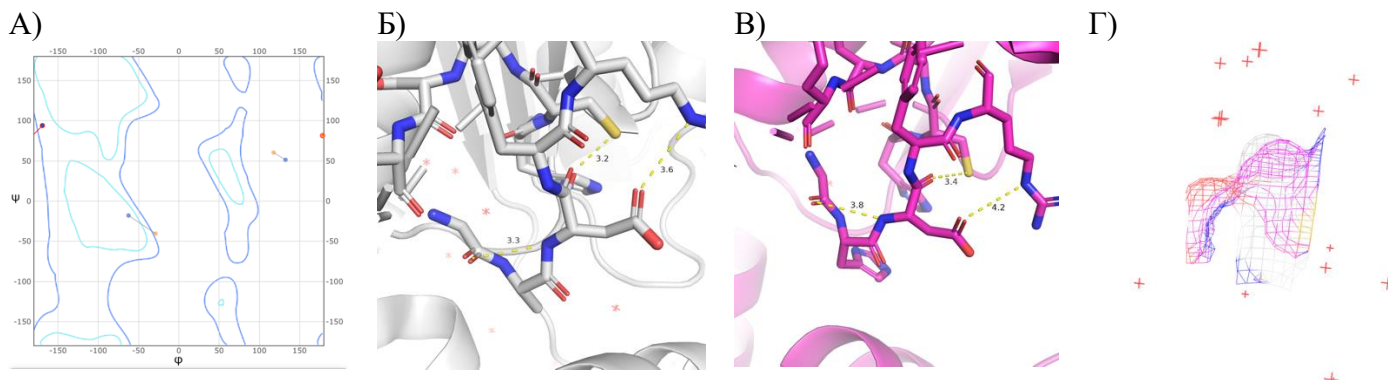


Рисунок 5. А) Карта Рамачандрана из выдачи PDB_redo [4]. D-679 выделен красным, синяя точка – положение остатка в исходной структуре, а оранжевая – в структуре PDB_redo, Б) D-679, взаимодействия, В) Соответствующие взаимодействия в структуре PDB_redo, Г) ЭП D-679 в исходной структуре (серая) и в структуре PDB_redo (розовая).

4. **Q-623** – маргинальный ротамер глутамина. Кажется, что углы 113.2 и 112.1 слишком большие (рисунок 6 А)). Для боковой группы этого остатка я не нашла водородных связей, что неудивительно, потому что его атомы с двух сторон выходят на поверхность (рисунок 6 Б)). Если мы посмотрим на поверхность окружения этого атома, то увидим, что боковая цепь этого глутамина как бы «зажата» гидрофобными соседями (рисунок 6 В)). Интересно, что этот остаток – один из трех, для которых RSCC был значимо хуже определен в PDB_redo. Возможно, в модели, используемой в PDB_redo, этот ротамер глутамина не был маргинальным, что скомпенсировалось плохим соответствием экспериментальной ЭП. Мой вывод: особенность, а не ошибка.

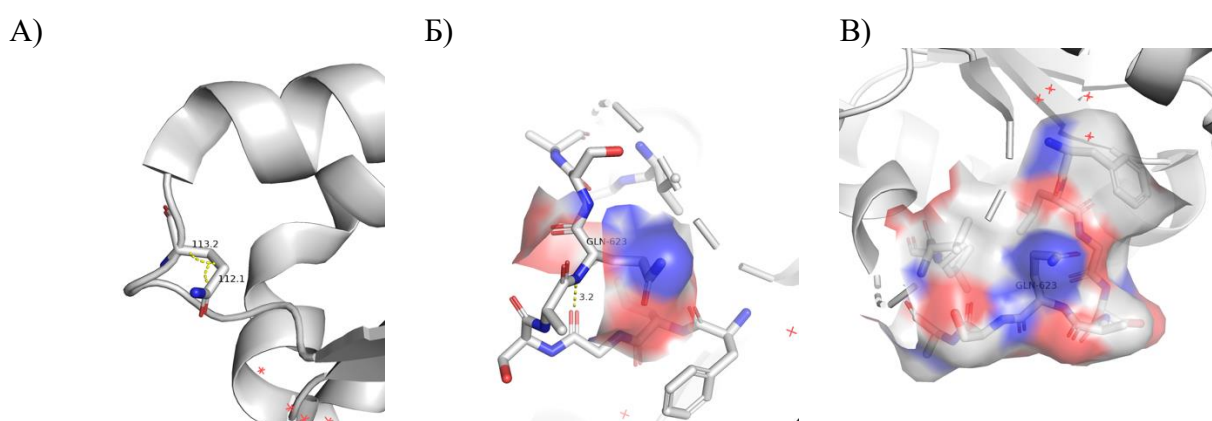


Рисунок 6. А) Торсионные углы Q-623, Б) Поверхность Q-623, В) Поверхность окружения Q-623.

5. **I-673** – аутлаер и по RSRZ, и по ротамерам. На рисунке 7 А) представлены углы боковой цепи. На первый взгляд не видно причин, по которым с этим изолейцином должно происходить что-то странное, он просто участвует в «молекулярном тетрисе», как ему и положено. Но, возможно, причина его маргинальностей в том, что его боковая цепь находится очень близко к поверхности белка и это как-то неблагоприятно на нее влияет

(рисунок 7 Б)). Вывод в этом случае сделать затруднительно, если говорить упрощенно, мы знаем, что для этого плохого остатка плохо определена ЭП. Вероятно, это – результат некоторого компромисса. Я бы сказала, что здесь, скорее, ошибка эксперимента, учитывая, что данный остаток находится рядом с «unmodelled region» (625-626 остатки).

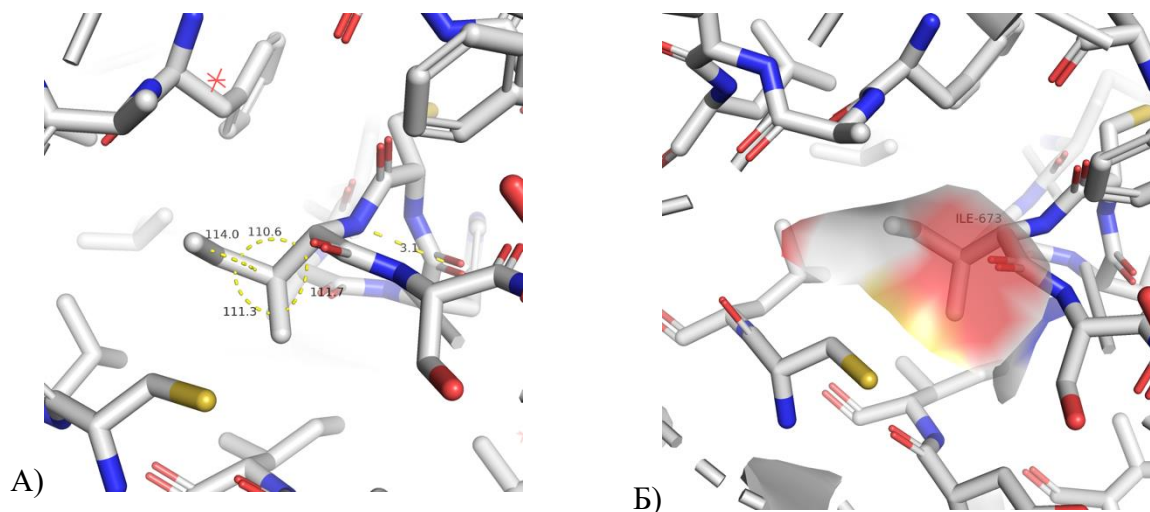


Рисунок 7. А) Торсионные углы I-673, Б) Поверхность I-673.

Для определения и описания наиболее подозрительных остатков я использовала выдачу PDB_redo [4]. В целом, общие показатели, такие, как R-фактор, R_free, разрешение, процент тестовой выборки в исходной структуре и в PDB_redo совпадают или почти совпадают. Но, судя по моим выводам о каждом маргинальном остатке, PDB_redo не всегда лучше интерпретирует экспериментальные данные (например, остаток D-679). Однако для подавляющего большинства остатков $\Delta(RSCC)$ указывает на то, что в PDB_redo структуре ЭП остатка лучше соответствует модели либо $\Delta(RSCC)$ не значим.

Заключение

Я оцениваю данную валидацию результатов PCA эксперимента по структуре WRN как хорошую, исходя из, во-первых, общих показателей (разрешение в сочетании с полнотой данных, R-фактор и R_free), во-вторых, некоторых частных. Например, в рассмотренных мной маргинальных остатках так или иначе достигался компромисс между соответствием экспериментальной ЭП модели и обоснованностью остатка в именно таком его состоянии.

Список литературы:

1. Newman, Joseph A et al. "Structure of the helicase core of Werner helicase, a key target in microsatellite instability cancers." *Life science alliance* vol. 4,1 e202000795. 16 Nov. 2020, doi:10.26508/lsa.202000795
2. PDB запись <https://www.rcsb.org/structure/6YHR>
3. MolProbity <http://molprobity.manchester.ac.uk>
4. PDB_redo <https://pdb-redo.eu>