

Отчет по качеству расшифровки PCA структуры белка WRN из *Homo sapiens* (PDB ID 2FBX)

Нефедова Анастасия

Аннотация

В данной работе был проведен анализ качества расшифровки экзонуклеазного домена белка WRN в комплексе с ионом магния - PDB ID 2FBX. Структура была получена методом рентгеноструктурного анализа. В отчете были рассмотрены параметры оценки качества модели и проанализированы некоторые маргинальные остатки структуры.

Введение

Белок WRN относится к семейству RecQ ДНК-хеликаз, обладающих АТФазной активностью и способностью расплетать ДНК в направлении 3' → 5', в связи с чем представители человеческих RecQ хеликаз участвуют в репарации, репликации и рекомбинации. Среди представителей своего семейства WRN отличается тем, что имеет 3' → 5' экзонуклеазный домен и обладает 3' → 5' экзонуклеазной активностью, при этом субстратом является двуцепочечная ДНК с 3'-рецессивным концом.

Белок WRN кодируется геном *WRN*, мутации в котором вызывают синдром Вернера, приводящего к преждевременному старению больного, появлению атеросклероза, злокачественных опухолей. На клеточном уровне нарушения работы гена *WRN* выражаются в обширных хромосомных делециях, повышенной степени гомологичной рекомбинации, длительному синтезу ДНК в S-фазе и нарушении поддержания теломер. [1]

В статье [1] получили расшифровку структуры экзонуклеазного домена WRN (WRN-ехо), а также охарактеризовали его биохимическую и функциональную активность. Авторы получили структуры со следующими PDB ID: 2FBT (нативный WRN-ехо), 2FBY (WRN-ехо в комплексе с Eu^{3+}), 2FBV (WRN-ехо в комплексе с Mn^{2+}), 2FBX (WRN-ехо в комплексе с Mg^{2+}) и 2FCO (WRN-ехо в комплексе с Mn^{2+} -dGMP). Было показано, что *in vitro* Mn^{2+} поддерживает более высокую экзонуклеазную активность WRN, чем Mg^{2+} , однако наиболее вероятно присутствие Mg^{2+} в качестве лиганда в структуре.

Для валидации я взяла структуру с WRN-ехо в комплексе магнием - PDB ID 2FBX.

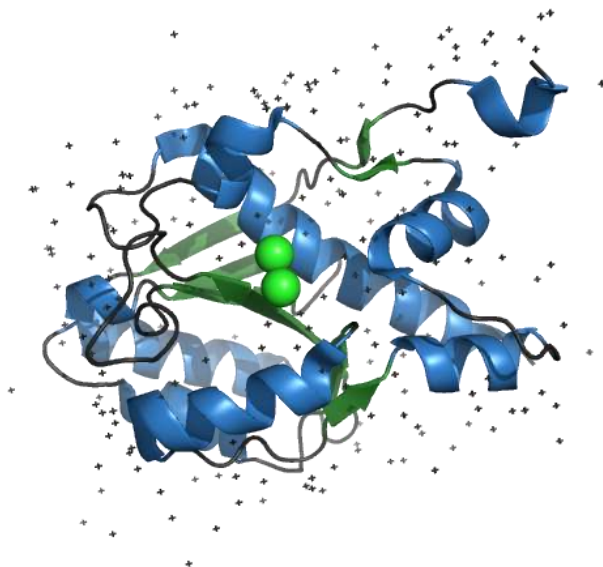


Рис.1. WRN-ехо (PDB ID 2FBX). Раскраска по типу вторичной структуры.

Результаты и обсуждение

1. Общая информация о модели

Структура была получена методом PCA и находится в PDB банке [2] с 10 декабря 2005 года и была получена для статьи [1], вышедшей в 2006 году, авторы Perry, J.J. et al.

Длина последовательности структуры – 205 а.о. Для получения итоговой структуры авторы сначала экспрессировали N-терминальный экзонуклеазный домен WRN₁₋₃₃₃, а затем путем ограниченного протеолиза получили функциональный WRN-ехо минимальной длины – WRN₃₈₋₂₃₆.

В структуре с данным PDB присутствует одна цепь, содержащая альфа-спирали 1-7, бета-листы 0-7, петли и ион магния. Биологическая единица представляет собой мономер.

Разрешение структуры 2FBX - 2.20 Å, диапазон разрешений – от 2.20 до 19.70 Å, полнота набора гармоник 96.9 %, число рефлексов составляет 17851. Для решения фазовой проблемы использовался метод аномального рассеяния на белковом кристалле WRN-ехо, замещенным селенметионинами. [1]

Кристалл имеет кристаллографическую группу $P3_221$. Параметры кристаллографической ячейки приведены в табл.1. Количество молекул в ячейке равно 6. Некристаллографических симметрий в асимметрической ячейке нет.

Табл.1. Параметры кристаллографической ячейки.

Длина (Å)	Угол (°)
a = 80.806	$\alpha = 90$
b = 80.806	$\beta = 90$
c = 93.174	$\gamma = 120$

2. Значения индикаторов качества модели

В таблице 2 приведены параметры оценки качества модели. Хорошим R-фактором считаются его значения < 0.25 . R-фактор характеризует соответствие модели ее экспериментальным значениям. У рассматриваемой структуры он равен 0.242, что является допустимым для того, чтобы считать соответствие хорошим. R-free служит контролем переоптимизации, хорошими считаются значения менее 0.2. В данном случае 0.262 – удовлетворительное значение, близкое к R-фактору, а значит, нет подозрений на переоптимизацию. Среднеквадратичные отклонения длин и углов являются довольно маленькими, это говорит о качестве модели.

Табл.2. Параметры качества модели.

Параметр	Значение
R-value	0.242
R-free	0.262
RMSD bond length (Å)	0.006
RMSD bond angles (°)	1.1

Посмотрим также на оценку маргинальных остатков. С помощью сервера MolProbity [3] были получены параметры оценки, табл. 3. С помощью MolProbity в структуру были добавлены водороды (1560 атомов), оптимизированы 49 водородных связей.

Табл.3. Индикаторы качества структуры 2FBX, определенные сервисом MolProbity.

All-Atom Contacts	Clashscore, all atoms:	13.49		79 th percentile* (N=456, 2.20Å ± 0.25Å)
	Clashscore is the number of serious steric overlaps (> 0.4 Å) per 1000 atoms.			
Protein Geometry	Poor rotamers	12	6.86%	Goal: <0.3%
	Favored rotamers	152	86.86%	Goal: >98%
	Ramachandran outliers	1	0.52%	Goal: <0.05%
	Ramachandran favored	186	96.37%	Goal: >98%
	Rama distribution Z-score	-2.65 ± 0.54		Goal: abs(Z score) < 2
	MolProbity score	2.51		50 th percentile* (N=10167, 2.20Å ± 0.25Å)
	Cβ deviations >0.25Å	0	0.00%	Goal: 0
	Bad bonds:	0 / 1588	0.00%	Goal: 0%
Bad angles:	1 / 2142	0.05%	Goal: <0.1%	
Peptide Omegas	Cis Prolines:	1 / 5	20.00%	Expected: ≤1 per chain, or ≤5%

Clashscore показывает число недопустимых перекрытий (пересечений Ван-дер-Ваальсовых радиусов более, чем на 0,4 ангстрема) на 1000 атомов. Для данной модели значение clashscore входит в 79 перцентиль, т.е. у 79% структур такого же разрешения clashscore хуже, чем у данной, это хорошее значение.

Для оценки конформации боковых цепей используют ротамеры. Ротамеры – это наборы разрешенных для данного остатка положений боковых цепей, которые задаются допустимыми значениями торсионных углов. По MolProbity, число плохих ротамеров не должно превышать 0.3 %, чтобы считать модель хорошей по этому параметру. Для 2FBX сервис обнаружил 12 плохих ротамеров, что составляет 6.86% остатков, это высокое значение.

Оценка конформации остова оценивается с помощью карты Рамачандрана, которая отражает допустимые углы вращения между атомами остова белка. На рис.2 представлена карта Рамачандрана для всех аминокислотных остатков белка. 96.4% остатков находятся в предпочтительной области, 99.5% - в допустимой. Есть один аутлайер – 87 пролин. Он является цис-пролином.

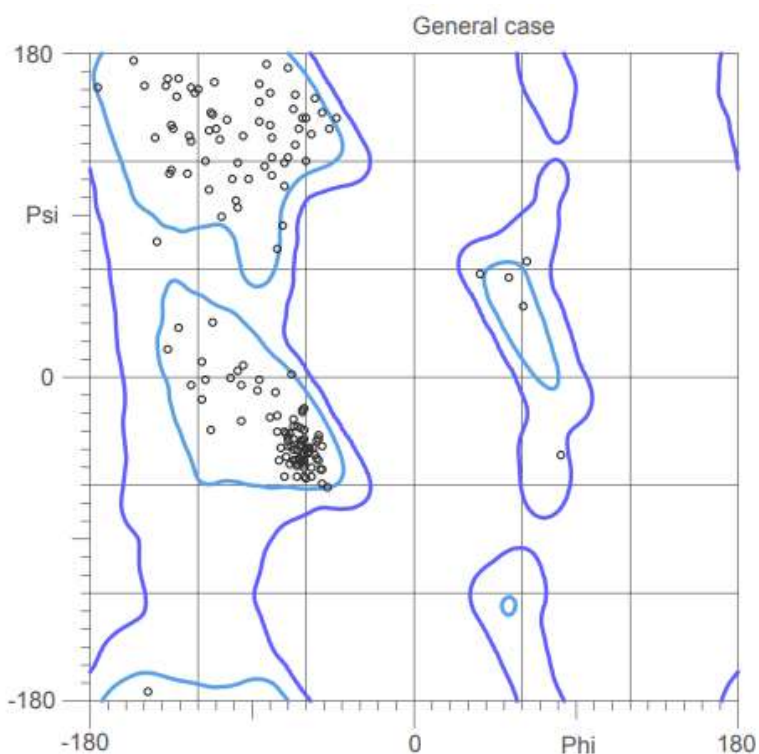


Рис.2. Карта Рамачандрана для структуры 2FBX. Точками обозначены аминокислотные остатки, голубые линии ограничивают участки предпочтительных областей, синие – допустимых областей.

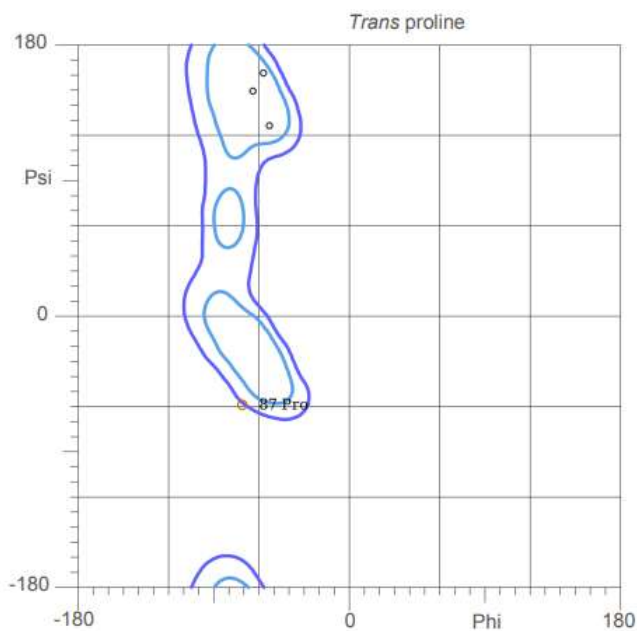


Рис.3. Карта Рамачандрана для транс-пролинов структуры 2FBX. Точками обозначены аминокислотные остатки, голубые линии ограничивают участки предпочтительных областей, синие – допустимых областей.

Также для структуры не было найдено отклоняющихся β -атомов, ковалентных связей и валентных углов (в PDB report сказано, что плохих углов нет, MolProbity нашел одну, видимо, это связано с условиями работы оценивающих качество алгоритмов).

MolProbity score – это интегральная оценка, которая комбинирует clashscore, показатели по ротамерам и карте Рамачандрана. Структура 2FBX оказалась средней по этому суммирующему параметру – 50% моделей оказались хуже.

Из рис.4. видно, что модель по ряду параметров хуже средних значений по всему банку PDB. Заметим, что присутствие всего одного аутлайера по карте Рамачандрана сдвинуло модель PDB в “плохую” сторону относительно среднего. Для данной структуры один аутлайер – это 1%, и связано это с небольшой длиной последовательности. Значение R-free, хоть и ниже среднего по PDB, но вполне удовлетворительное и не сильно отличается от R-фактора. Минусом структуры является относительно большое количество “плохих” ротамеров и перекрываний по Ван-дер-Ваальсовым радиусам. Но некоторые ротамеры, например, можно доисправить вручную для улучшения модели. В целом, качество структуры 2FBX можно назвать удовлетворительным.

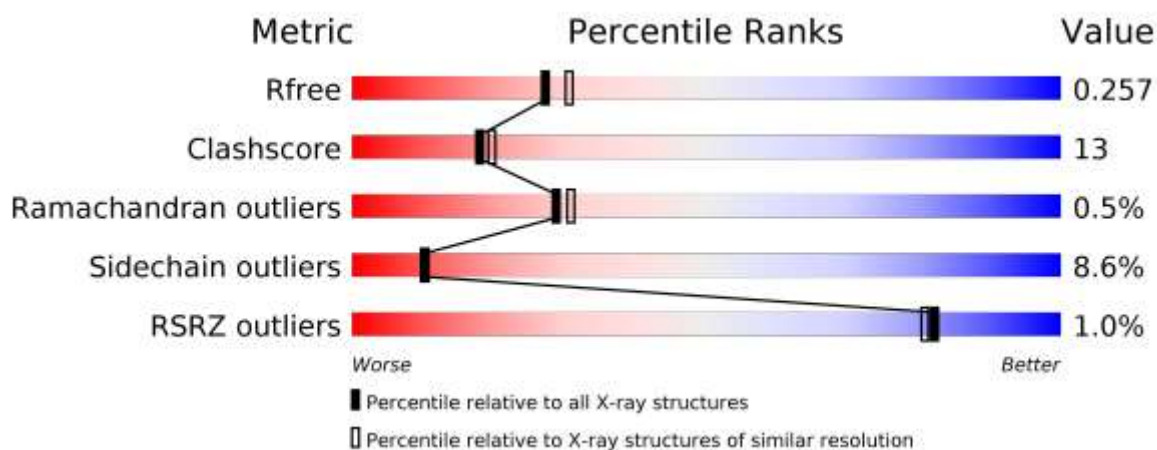


Рис.4. Оценка качества структуры 2FBX относительно банка PDB. Красный цвет означает смещение в худшую сторону, синий – в лучшую. Черная метка показывает положение модели относительно всех PCA кристаллов, белая – относительно структур с тем же разрешением. Источник: [2]

3. Маргинальные остатки

Разберем поподробнее некоторые маргинальные остатки структуры 2FBX, то есть остатки с нетипичными значениями параметров.

Параметр RSR характеризует соответствие модели аминокислотных остатков экспериментальным данным. Оценка изображается в виде RSRZ – Z-score, показывающий, насколько RSR для остатка данной модели отличается от среднего RSR для остатков такого же типа всех структур в PDB с таким же разрешением. Маргинальным считается остаток с RSRZ > 2.

На рис. 5 приведена визуализация качества восстановления остатков. Остатки покрашены по маргинальности относительно геометрических параметров, в которых есть хотя бы один outlier: зеленым – 0 таких параметров, желтым – 1, оранжевым – 2. Красными точками помечены остатки с RSRZ > 2. Остатки, которых нет в финальной модели, указаны серым (в статье [1] сказано, что последние 6 С-терминальных остатков не разрешились, серые остатки на N-конце – это, вероятно, 6His тег).

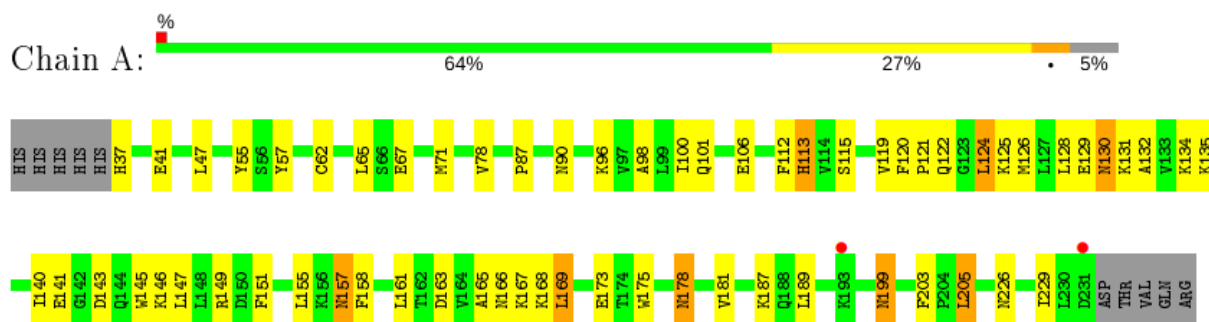


Рис. 5. Маргинальные остатки. Источник: [4]

В табл.4. приведены маргинальные остатки по критерию RSRZ, их всего 2, доля составляет 1%, что ниже среднего по банку PDB (рис.4).

Табл.4. Маргинальные остатки 2FBX согласно критерию RSRZ. Источник: [4]

Остаток	Аминокислота	RSRZ
231	Asp	3.6
193	Lys	2.2

Также в структуре есть 15 остатков по данным из PDB report (12 остатков по MolProbity) с необычными для себя (не ротаметрическими) торсионными углами, это 9% (6,8% по MolProbity) аутлайеров, что достаточно много и гораздо хуже среднего значения по PDB (рис.4).

У 11 остатков аспарагина, глутамина и гистидина необходимо инвертировать боковую цепь для улучшения водородных связей и уменьшения клэшей.

С помощью сервиса WHAT_CHECK [5] была рассчитана молекулярная поверхность для каждого остатка, при этом было обнаружено всего несколько гидрофобных остатков с большой доступной площадью (значительно превышающей среднее значение для данного остатка по всей структуре), два таких остатка – Trp 145 и Ile 229 приведены в табл.5.

Табл.5. Список маргинальных остатков (по нескольким критериям). Жирным отмечены остатки, которые будут подробнее проанализированы. Источники: [4, 5]

Номер остатка	Аминокислота	Признак маргинальности
145	Trp	Перекрывание атомов 1:A:145:TRP:O 1:A:146:LYS:CB 1:A:145:TRP:O 1:A:146:LYS:HB3 Повышенная доступная молекулярная поверхность
90	Asn	Боковая цепь – неротамер Подозрение на инверсию боковой цепи
199	Asn	Боковая цепь – неротамер Подозрение на инверсию боковой цепи Перекрывание атомов 1:A:199:ASN:HD22 1:A:199:ASN:C
113	His	Боковая цепь – неротамер Подозрение на инверсию боковой цепи Перекрывание атомов 1:A:113:HIS:NE2 1:A:205:LEU:HD13 1:A:55:TYR:OH 1:A:113:HIS:HD2
178	Asn	Боковая цепь – неротамер Подозрение на инверсию боковой цепи Перекрывание атомов 1:A:178:ASN:HD21 1:A:189:LEU:H 1:A:178:ASN:ND2 1:A:189:LEU:H
87	Pro	Цис-пролин, аутлайер карты Рамачандрана 87 Pro (-71.2, -59.9)
229	Ile	Повышенная доступная молекулярная поверхность Перекрывание атомов 1:A:226:ASN:O 1:A:229:ILE:HG22

4. Анализ 5 маргинальных остатков

Был произведен более пристальный анализ 5 маргинальных остатков из табл.5.

1) Trp 145

Маргинальные признаки этого остатка – перекрывание атомов с атомами лизина 146 и излишняя доступная молекулярная поверхность. На рис.6. изображена электронная плотность триптофана 145, видно, что остаток хорошо покрыт плотностью.

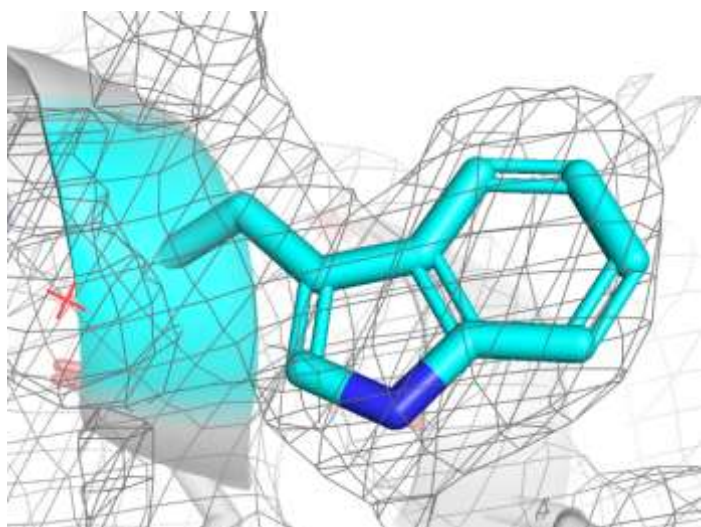


Рис. 6. Электронная плотность триптофана 145

Этот остаток входит в состав альфа-спирали (рис.7). Если посмотреть на водородные связи остова этой альфа-спирали, видно, что C=O группа триптофана 145 имеет не совсем правильное положение (должна смотреть вниз), из-за чего образует водородную связь в 5.5 Å, то есть не образует водородную связь, однако этот участок считается альфа-спиралью.

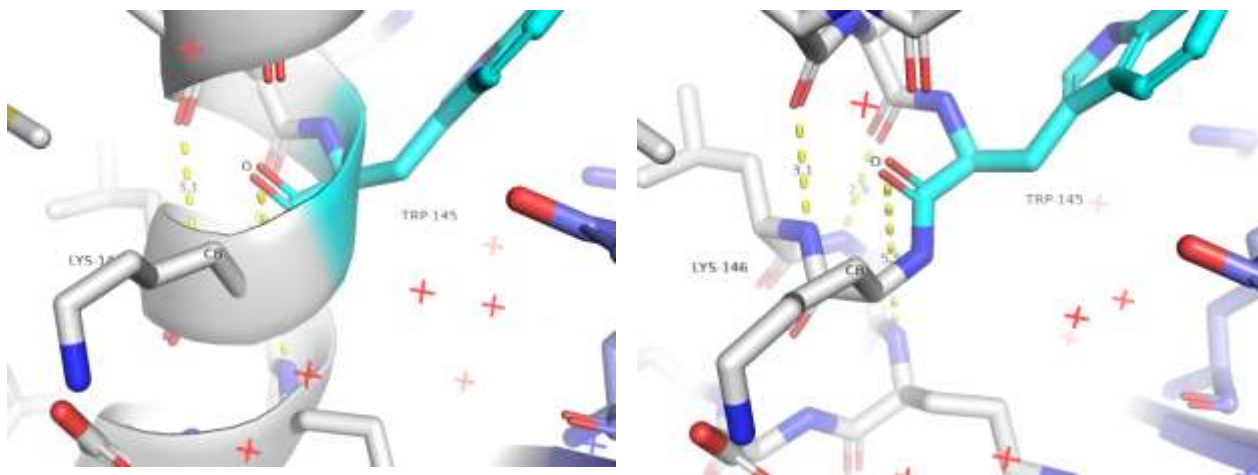


Рис. 7. Участок альфа-спирали, куда входит триптофан 145. Желтым показаны водородные связи.

Если посмотреть на этот остаток в контексте положения в структуре (рис.8), то видно, что он находится на периферии 2FBX, рядом находится соседняя по ячейке молекула, и с одной стороны (сверху на рис.8) он действительно доступен для растворителя. Я думаю, что можно считать этот остаток маргинальным, потому что положение его в альфа-спирали. Возможно, он занял такое положение в ячейке из-за того, что находится на поверхности молекулы рядом с другой молекулой-соседом.

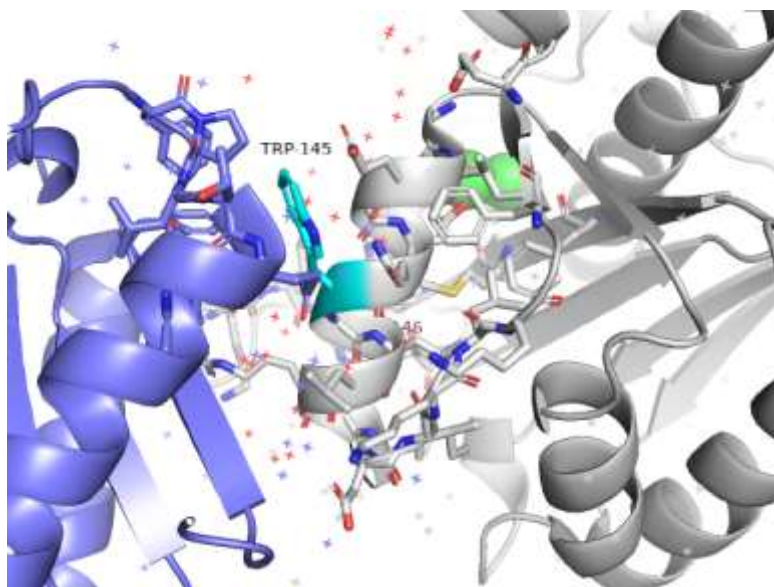


Рис.8. Участок с триптофаном 145 в структуре ячейки. Синим показана соседняя по ячейке молекула.

Интересно, что в улучшенной модели из PDB_redo положение триптофана 145 остается тем же, но альфа-спираль на месте этого остатка прерывается. (рис. 9) Думаю, что было правильно исключить триптофан из вторичной структуры.

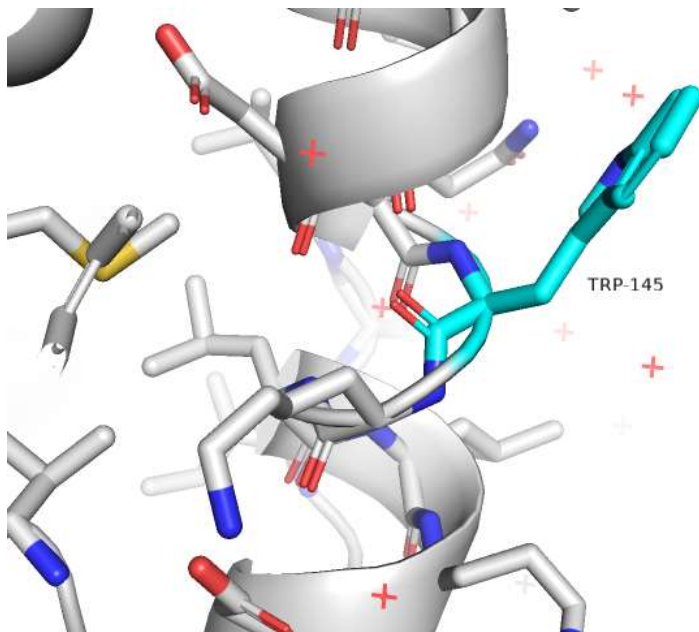


Рис. 9. Участок с триптофаном 145 из улучшенной модели PDB_redo.

2) Asn 199

Из рис. 10 видно, что аспарагин 199 действительно занимает не предпочтительное положение. Сравнивая с положением этого же остатка из модели PDB_redo, где он может образовать водородную связь с остовом серина 201, становится понятно, что в исходной модели боковая цепь аспарагина 199 инвертирована, и этот остаток является маргинальным.

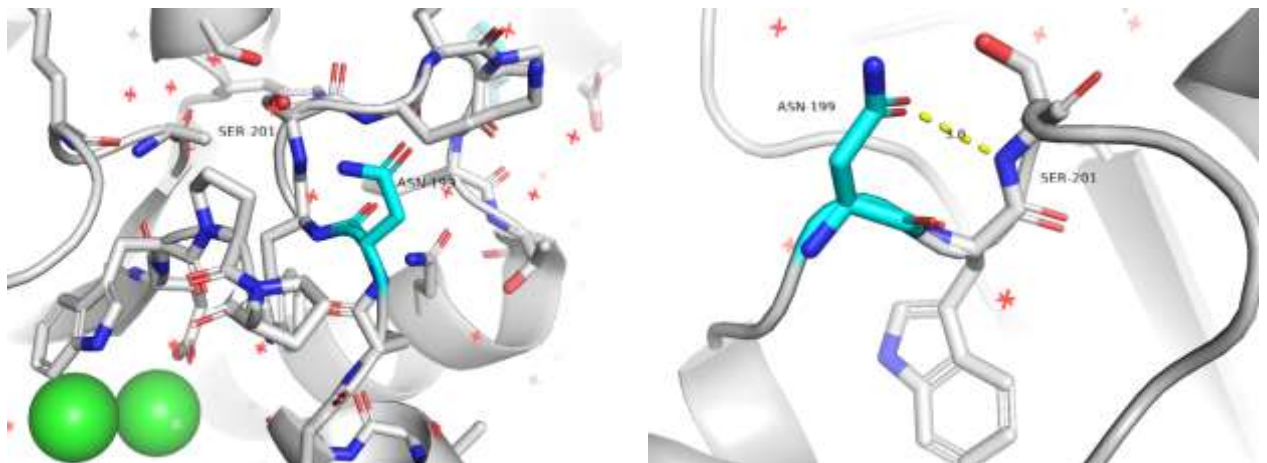


Рис. 10. Участок аспарагина 199 (голубым цветом). Слева – модель PDB, справа - модель PDB_redo. Желтым цветом показана водородная связь.

3) His 113

Для данного остатка ситуация практически аналогичная. Правильное положение (рис.11 справа – из PDB_redo) позволяет гистидину образовать

водородную связь с серином 201. Гистидин в исходной модели является маргинальным и справляется инвертированием боковой цепи.

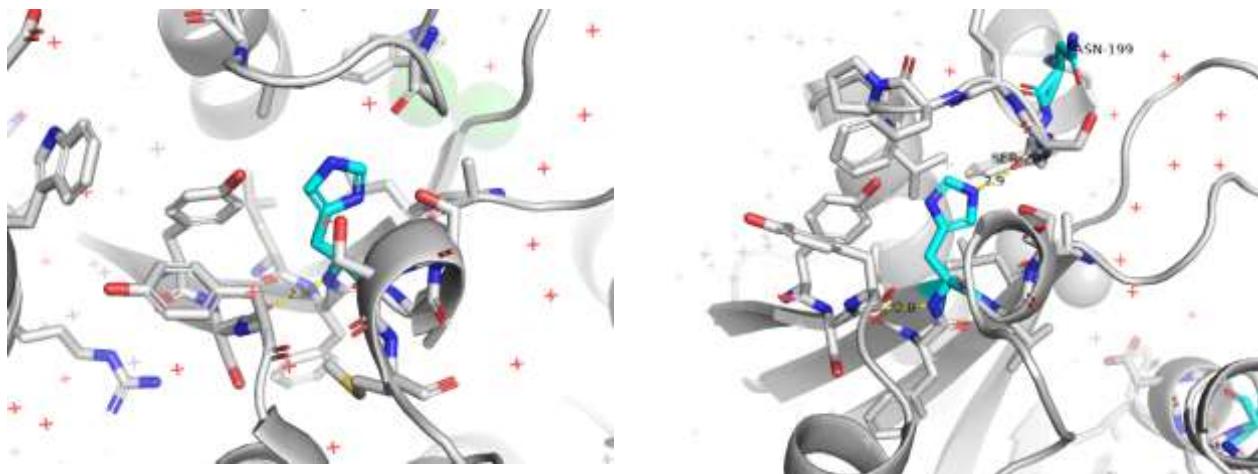


Рис. 11. Участок гистидина 113 (голубым цветом). Слева – модель PDB, справа - модель PDB_redo. Желтым цветом показана водородная связь.

4) Ile 229

Данный остаток является слишком доступным для растворителя, а также его атомы перекрываются с аспарагином 226. Изолейцин 229 попадает в участок модели, плохо покрываемой электронной плотностью. Поскольку остаток находится на периферии, не удалось зафиксировать здесь электронную плотность, поэтому трудно утверждать что-либо о маргинальности изолейцина 229. На мой взгляд, из-за его гидрофобности он мог быть расшифрован неправильно.

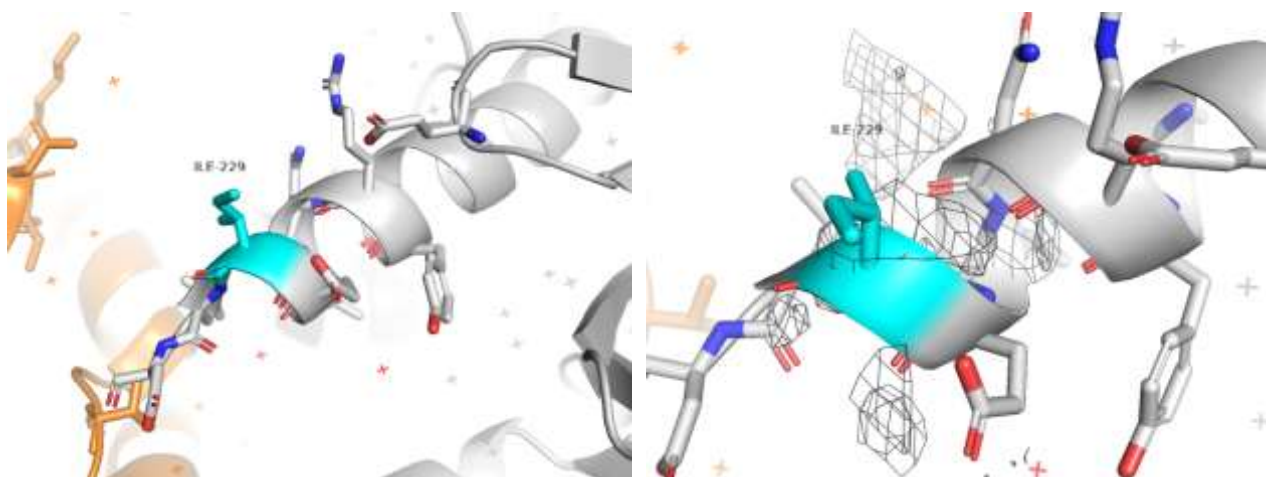


Рис. 12. Участок изолейцина 229 (голубым цветом). Справа показана электронная плотность.

В плохой расшифровке электронной плотности лежит и причина высокого RSRZ аспартата 231, который находится рядом.

5) Asn 178

Причина маргинальности аспарагина 178 лежит в инвертированности боковой цепи. При правильном положении этот остаток образует водородную связь с лейцином 189 (рис.13), а в исходной модели он является маргинальным.

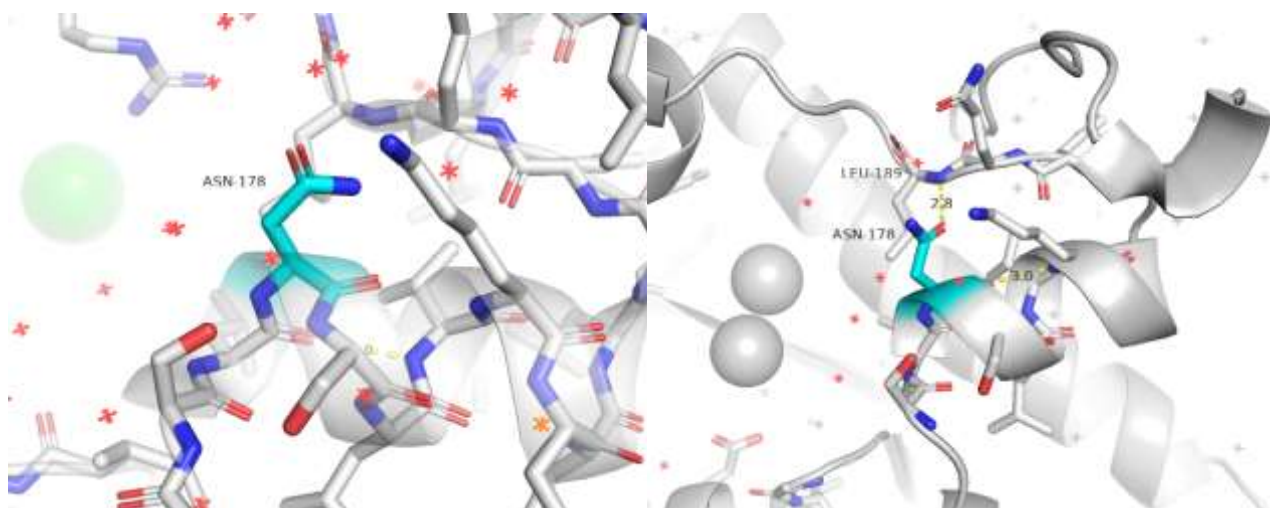


Рис. 13. Участок аспарагина 178 (голубым цветом). Слева – модель PDB, справа - модель PDB_redo. Желтым цветом показана водородная связь.

5. Сравнение модели из PDB с моделью из PDB_redo

В базе данных PDB_redo содержатся оптимизированные версии моделей из банка PDB. Сравним параметры улучшенной модели 2FBX [6] с исходной.

Из табл.6 и рис. Видим, что улучшились значения R-фактора и R-free, улучшилось положение некоторых остатков относительно допустимых торсионных углов. Было улучшено 9 ротамеров, удалено 46 молекул воды.

При анализе маргинальных остатков проводилось сравнение с моделью из PDB_redo, положения многих остатков действительно улучшились.

Табл. 6. Сравнение R-фактора и R-free модели в PDB и PDB_redo.

Параметр	Значение в PDB модели	Значение в PDB_redo модели
R	0,2359	0,1882
R-free	0,2573	0,2160

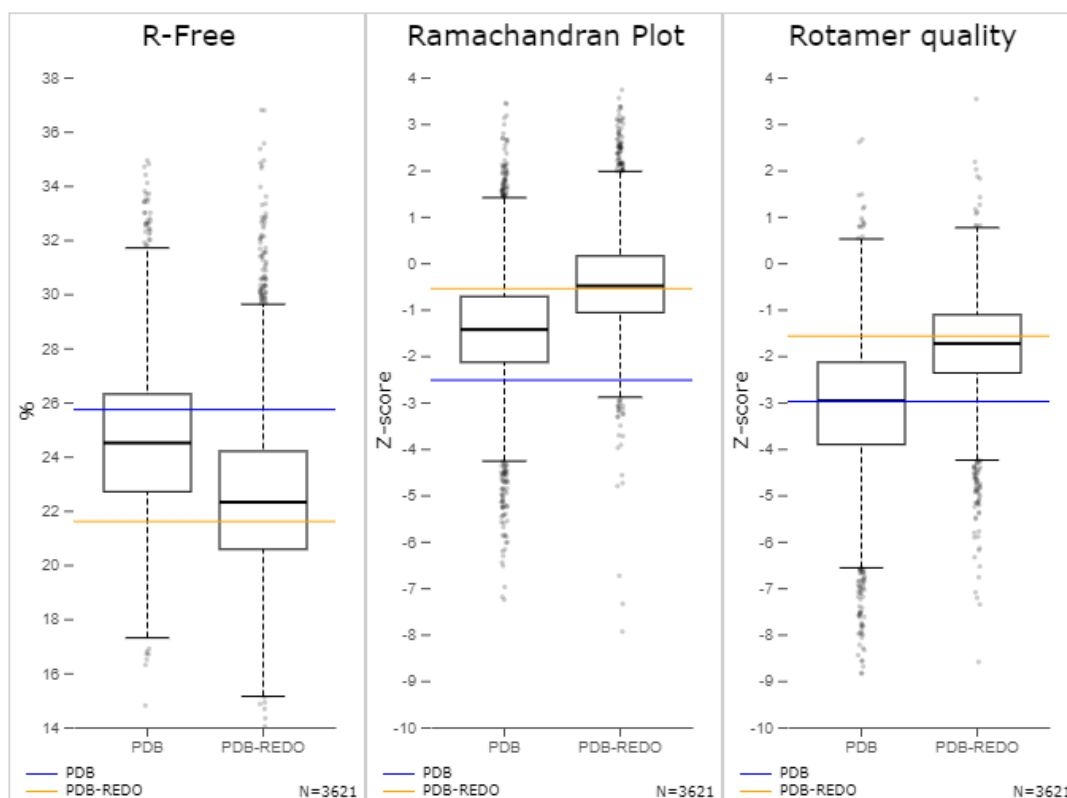


Рис.14. Сравнение основных параметров модели в PDB и PDB_redo.

Выводы

Рассмотренная структура экзонуклеазного домена белка WRN в комплексе с ионом магния (PDB ID 2FBX) имеет достаточно высокое разрешение, однако почти все параметры оценки качества структуры имеют значения ниже среднего по PDB – в основном, это объясняется наличием достаточно большого количества “плохих” ротамеров и клэшей, но эта проблема частично решается оптимизацией модели в PDB_redo. В целом, я бы сказала, что качество исходной модели удовлетворительным, но требующим доработки.

Список литературы

1. Perry, J., Yannone, S., Holden, L. *et al.* WRN exonuclease structure and molecular mechanism imply an editing role in DNA end processing. *Nat Struct Mol Biol* **13**, 414–422 (2006). <https://doi.org/10.1038/nsmb1088>
2. <https://www.rcsb.org/structure/2FBX>
3. <http://molprobity.biochem.duke.edu/>
4. https://files.rcsb.org/pub/pdb/validation_reports/fb/2fbx/2fbx_full_validation.pdf
5. <https://swift.cmbi.umcn.nl/servers/html/>

6. <https://pdb-redo.eu/db/2fbx>