

Отчёт по качеству PСА расшифровки структуры белка гексокиназы (PDB ID – 1HKV)

Автор: Иззи Антон

Аннотация:

В данной работе было проанализировано качество расшифровки структуры белка гексокиназы, который фосфорилирует глюкозу на 1 этапе гликолиза. Помимо общих характеристик модели, которые указывают на её соответствие экспериментальным данным или общим химическим свойствам, были также исследованы частные случаи ошибки расшифровки электронной плотности на уровне отдельных аминокислотных остатков. Помимо этого, было проведено сравнение текущей модели с улучшенным вариантом из PDB_redo. И в заключении был сделан общий вывод о качестве расшифровки структуры исследуемого белка.

Введение:

Гексокиназа, или АТФ: D-гексоза-6-фосфотрансфераза, представляет собой фермент из класса трансфераз, известный своей функцией на 1 стадии гликолиза. Он осуществляет фосфорилирование глюкозы, используя АТФ как донор фосфорильной группы. Общая схема реакции выглядит следующим образом (рис 1):

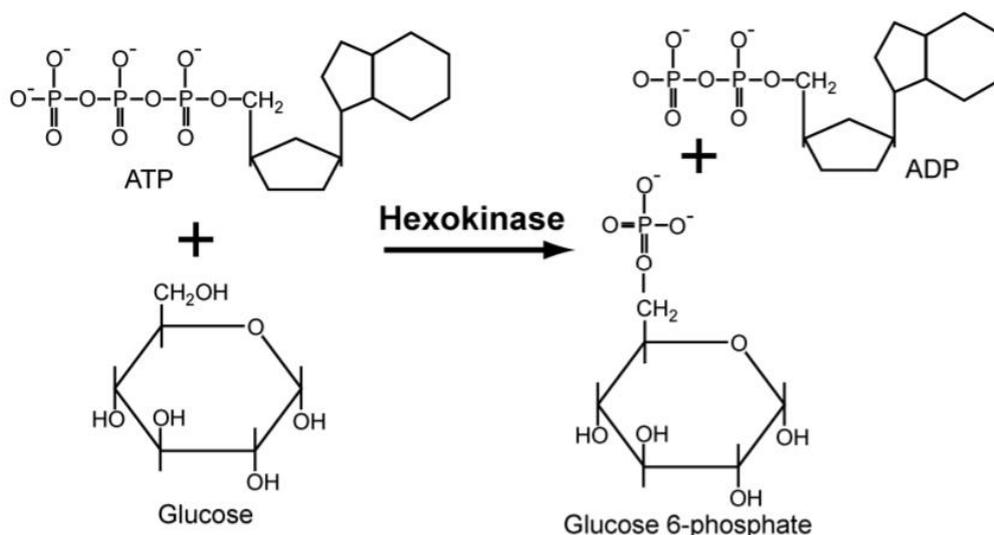


Рис 1. Схема реакции фосфорилирования глюкозы в гексокиназной реакции [1]

У гексокиназы есть изофермент, который находится в печени – глюкокиназа. Отличается от гексокиназы в основном тем, что он не ингибируется по отрицательной обратной связи глюкозо-6-фосфатом. Также совместно реагирует в физиологическом диапазоне на изменение концентрации глюкозы [2].

У гексокиназы активный центр составляют 5 остатков: Asp205, Lys169, Asn204, Glu256 и Thr168 [3]. Данные остатки располагаются глубоко в щели между двумя долями. Активный центр может связывать два лиганда – глюкозу и глюкозо-6-фосфат. При связывании глюкозы гексокиназа претерпевает некоторые конформационные изменения, которые предотвращают гидролиз АТФ, при этом идёт аллостерическое ингибирование глюкозой-6-фосфатом. Всего гексокиназа имеет два конформационные состояния – открытое и закрытое. В открытом состоянии фермент находится до связывания глюкозы, при этом АТФ связан с большой долей и находится далеко от сайта связывания глюкозы и в другом положении. При связывании глюкозы конформация меняется на закрытую, и две доли смыкаются вокруг субстрата [4].

Работа, связанная с использованной моделью (1HKV), была посвящена более детальному исследованию механизма катализа и регуляции на атомарном уровне. Было показано, что конформационные изменения в N-концевом домене фермента в ответ на связывание глюкозы, фосфата или глюкозо-6-фосфата могут влиять на связывание АТФ с C-концевым доменом [5].

Результаты и обсуждение:

При изучении модели первым делом были установлены её основные характеристики:

_ Год получения модели: 1998

_ Авторы: Aleshin, A.E., Zeng, C., Burenkov, G.P., Bartunik, H.D., Fromm, H.J., Honzatko, R.B.

_ Состав комплекса: HUMAN BRAIN HEXOKINASE TYPE I (рекомбинантный), GLUCOSE, GLUCOSE-6-PHOSPHATE

_ Метод решения фазовой проблемы: молекулярное замещение.

Осуществлялось с помощью программы X-PLOR, а в качестве стартовой модели использовалась дрожжевая гексокиназа в комплексе с глюкозой.

_ Число измеренных рефлексов: 55607

_ Разрешение: 2.79 Å

_ Полнота набора структурных факторов: 96.4 %

_ Диапазон разрешений структурных факторов: от 2.79 до 8

_ Кристаллографическая группа: P 1 21 1

_ Параметры кристаллографической симметрии: $a = 83$, $b = 122$, $c = 123$, $\alpha = 90$, $\beta = 92$, $\gamma = 90$

Ниже представлены идентификаторы качества модели 1НКВ и их сравнение с идентификаторами других структур PDB.

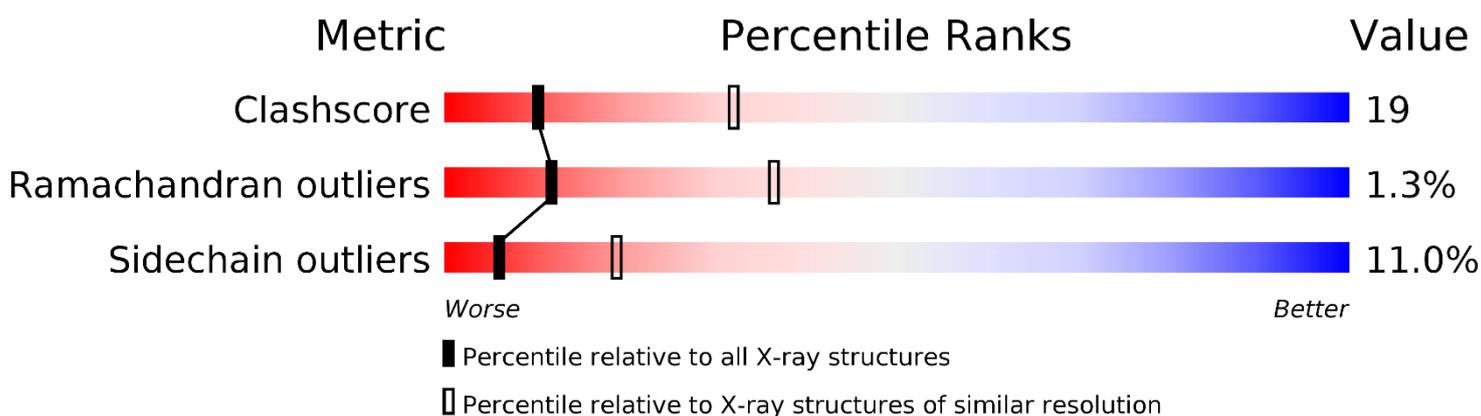


Рис. 2. Основные показатели качества модели. Здесь чёрные прямоугольники отражают перцентиль структур с таким же значением параметра относительно всех PCA-моделей, белые отражают перцентиль структур относительно PCA моделей близкого разрешения.

Помимо прочего также были проанализированы параметры, которые характеризуют качество модели в целом, например как R-factor, стандартный фактор достоверности, который представляет собой меру согласованности между кристаллографической моделью и реальным объектом (по сути тем, что наблюдаем в эксперименте), а также R-free, который считается по той же формуле, что и R-factor, но на гораздо меньшей выборке, то есть R-factor, рассчитанный для валидационной выборки рефлексов.

_ R-factor (тренировочная + тестовая выборки): 0.2

_ R-free (тестовая выборка): 0.27

При этом размер тестовой выборки составил 5653, что составляет примерно 10.17 % от всего набора измеренных рефлексов.

Можно заметить, что значение R-free не является большим, и причём мало отличается от R-value тренировочной выборки (R-factor (observed)), которая составляет 0.2), различие всего на 0.07. Можно предположить, что в таком случае качество расшифровки может быть вполне удовлетворительным (разница меньше чем 0.1), но это естественно не окончательный вывод, и поэтому необходимо провести дополнительный анализ.

Также было проанализировано качество расшифровки основных цепей белка посредством рассмотрения карт Рамачандрана, которая представляет собой карту допустимых торсионных углов. Каждой точке на карте можно приписать некоторую энергию конформации, и области низкой энергии будут соответствовать предпочитаемой конформации. Линиями обводятся предпочитаемые области, и остатки, не попавшие в эти области, считаются маргинальными. Было проанализировано как аминокислотные остатки накладываются на эти карты, какой процент их находится в энергетически невыгодных состояниях (напр. имеют такие значения торсионных углов ϕ и ψ , которые допускают перекрывание соседних остатков).

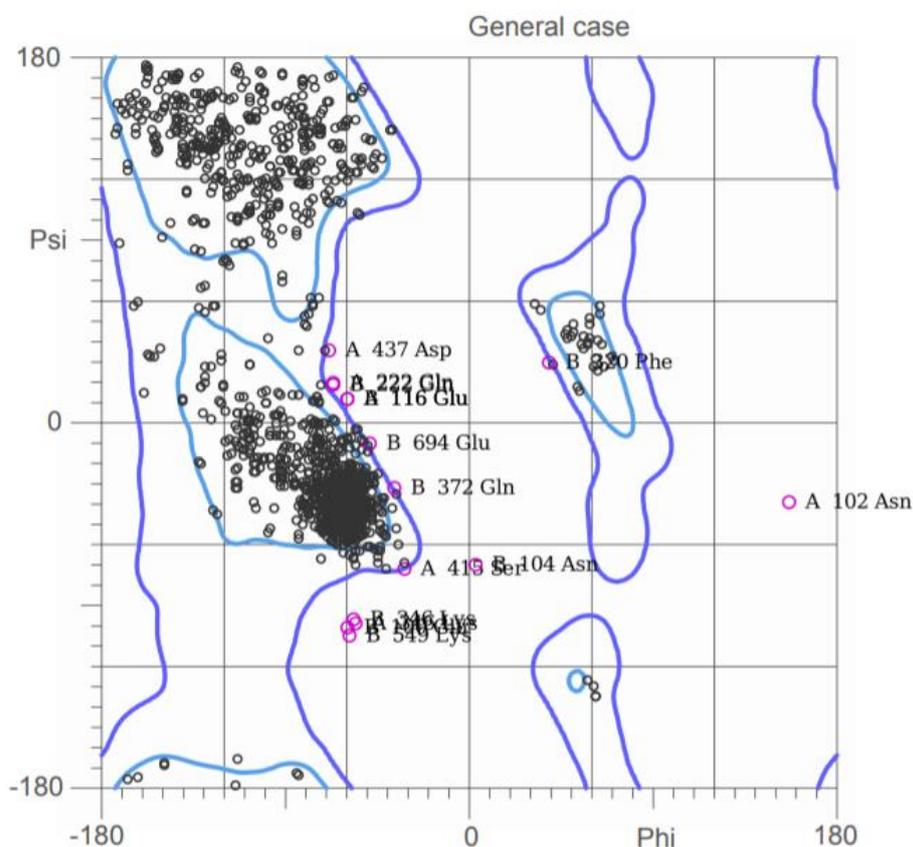


Рис 3. Карта рамачандрана модели с PDB_ID 1HKV (в общем случае)

В результате работы онлайн ресурса (ссылка будет в конце отчёта) была получена карта Рамачандрана модели с PDB_ID 1HKV (рис 2). На данной карте тёмно-синими линиями выделены области допустимых значений углов ϕ и ψ , а голубыми линиями – наиболее предпочтительные области. И на данной карте сразу можно заметить, что не всё так гладко. В глаза бросается сразу несколько маргинальных остатков, причём они 1,34 % от всех остатков (24). В предпочитаемых областях находятся 90,36 % остатков (1621). В допустимой зоне оказалось 8 % остатков (149). В общем итоговая таблица на сайте практически вся красного цвета. Уже намекает на то, что с расшифровкой структуры возможно есть некоторые проблемы.

Помимо карты Рамачандрана были также рассмотрены и другие характеристики для определения качества расшифровки аминокислотных остатков в составе комплекса:

_ Перекрытие Ван-дер-Ваальсовых радиусов атомов. Данный параметр отражает число стерических перекрытий в расчёте на 1000 атомов. Соседние атомы будут перекрываться, если Ван-дер-Ваальсовые радиусы этих атомов будут перекрываться более чем на 0,4 Å. Данный параметр характеризуется значением clashscore, который для данной структуры составляет 19,3 и попадает в 19-й перцентиль, и по критериям MolProbity является довольно хорошим результатом.

_ Ротамеры. Как и в картах Рамачандрана, данный параметр характеризует допустимые торсионные углы, но для боковых радикалов аминокислотных остатков, т.е. учитывает конформацию боковых цепей аминокислот и позволяет найти маргинальные остатки по данному критерию.

В данной модели по положению боковых цепей маргинальных остатков оказалось 188, что составляет 12,14 % от всех остатков. При том, что согласно критерию MolProbity таких остатков должно быть меньше, чем 0,3 %, такой результат является довольно плохим.

_ Также для модели хорошо было бы проанализировать RSR Z (Real Space R-value Z-score), который является центрированным и нормированным значением, характеризующим то, насколько электронная плотность атомов бокового радикала в модели отличается от действительности. Но по каким-то причинам данного параметра для исследуемой модели приведено не было. В связи с этим, было решено проанализировать другой параметр - RMS Z-score (Root Mean Square Z-score). В данном случае отличия от ожидаемого

считаются не по электронной плотности, а по длине связи или величине угла связи. Маргинальным считается тот остаток, у длины связи которого (или у величины угла связи которого) $|Z| > 5$. Для модели в целом RMSZ для длин связи составил 0.39, а для величин углов связи – 0.61. По длине связи маргинальных остатков не было, а вот по углу связи нашлось две аминокислоты:

Табл. 1. Маргинальные остатки с плохим RMSZ величины угла связи

Mol	Chain	Res	Type	Atoms	Z	Observed(o)	Ideal(o)
1	B	103	GLN	N-CA-C	-6.91	92.33	111.00
1	B	516	VAL	N-CA-C	-5.07	97.31	111.00

_ Помимо прочего также была проведена оценка правильности окружения остатков, посчитанная программой WHAT_CHECK (Coarse Packing Quality Control). Идея заключается в том, что распределение различных типов атомов вокруг аминокислотных остатков определено, и данное распределение в PDB отображает таковое в природе. Качество структуры определяется свёрткой средних распределений и наблюдаемых контактов в проверяемом белке. Если у остатка score < -5, то может быть несколько случаев: 1) остаток создаёт контакты симметрии; 2) он контактирует с лигандом или ионом или 3) присутствует какая-то ошибка. В целом по структуре общая оценка правильности окружения остатков составила -0.251.

Всего остатков со score < -5 оказалось 34. Далее приведены остатки с самыми «плохими» оценками правильности окружения.

Табл. 2. Остатки с самой плохой оценкой правильности окружения

Остаток	Coarse Packing Quality Control score
Chain A, 518 Arg	-6.906
Chain A, 700 Gln	-6.421
Chain B, 518 Arg	-6.887
Chain B, Gln 700	-6.954

Маргинальных по разным параметрам остатков оказалось немало, но все они подробному анализу подвергаться не будут. В таблице ниже приведено часть маргинальных остатков, некоторые причём и не по одному параметру:

Табл. 3. Некоторые маргинальные остатки структуры 1НКВ

Остаток	Параметр маргинальности
Chain A, 116 Glu	Плохой торсионный угол
Chain A, 102 Asn	Плохой торсионный угол + Недопустимый ротамер
Chain B, 100 Glu	Плохой торсионный угол + Недопустимый ротамер
Chain B, 222 Gln	Плохой торсионный угол + Недопустимый ротамер
Chain B, 516 Val	RMSZ >5
Chain B, 103 Gln	RMSZ >5 + Недопустимый ротамер
Chain B, 346 Lys	Оценка правильности окружения < -5 (WHAT_CHECK) + Плохой торсионный угол
Chain A, 407 Arg	Оценка правильности окружения < -5 (WHAT_CHECK) + Недопустимый ротамер
Chain A, 518 Arg	Оценка правильности окружения < -5 (WHAT_CHECK) + Недопустимый ротамер

Chain A, Glu 249	Оценка правильности окружения < -5 (WHAT_CHECK)
------------------	---

Из них отмеченные серым остатки были далее подвергнуты более детальному анализу с визуализацией в 3D.

1. 102-й аспарагин в цепи А

Данный остаток был охарактеризован как маргинальный по критерию карты Рамачандрана и по недопустимому ротамеру. Исходя из карты, маргинальным данный остаток является по углу ϕ – углу между $C\alpha$ атомом и NH_2 -группой, который составляет у него 157.9° , при том что для попадания в благоприятную область данный угол должен находиться примерно между -40° и -130° .

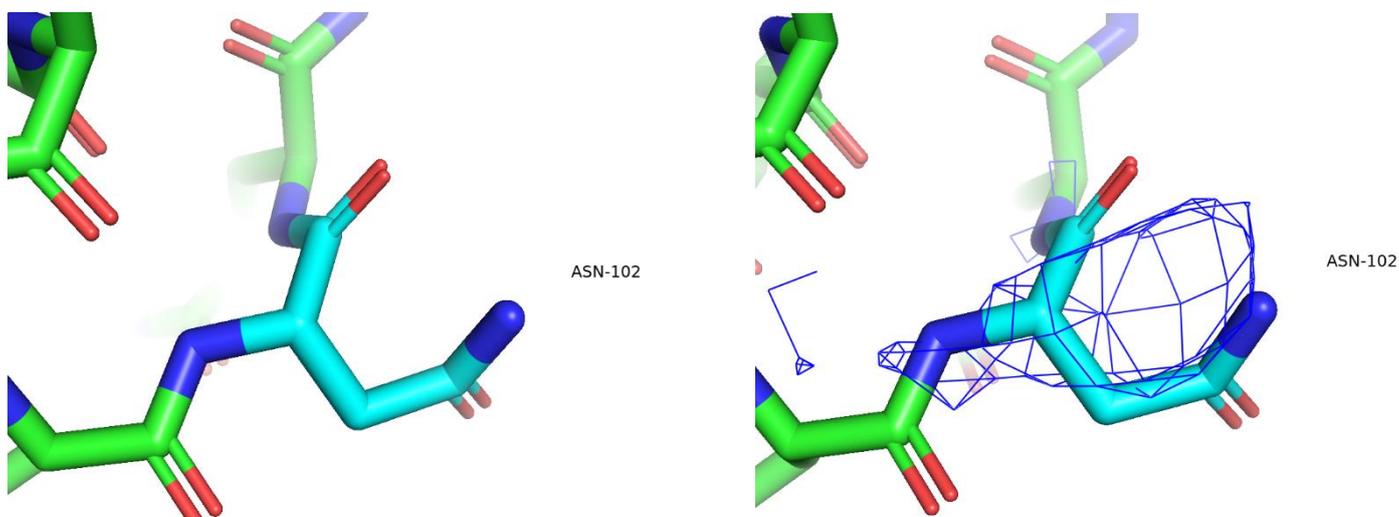


Рис. 4. Остаток ASN-102 в цепи в отображении без электронной плотности и с плотностью на уровне подрезки 1

Как можно заметить, электронная плотность данного остатка довольно плохо ложится как на радикал аспарагина 102, так и на остов. Поэтому столь странные характеристики данного остатка вполне могут вытекать из не совсем правильной интерпретации электронной плотности.

2. 100-й глутамат в цепи В

Данный остаток также был определён как маргинальный по критерию карты Рамачандрана и по недопустимому ротамеру. По карте маргинальным он оказался по углу ψ – углу между $C\alpha$ атомом и $C=O$. У него он составил -101.7° . Для попадания в нормальную область он должен составлять от -60° до -10° или от 110° до 165° .

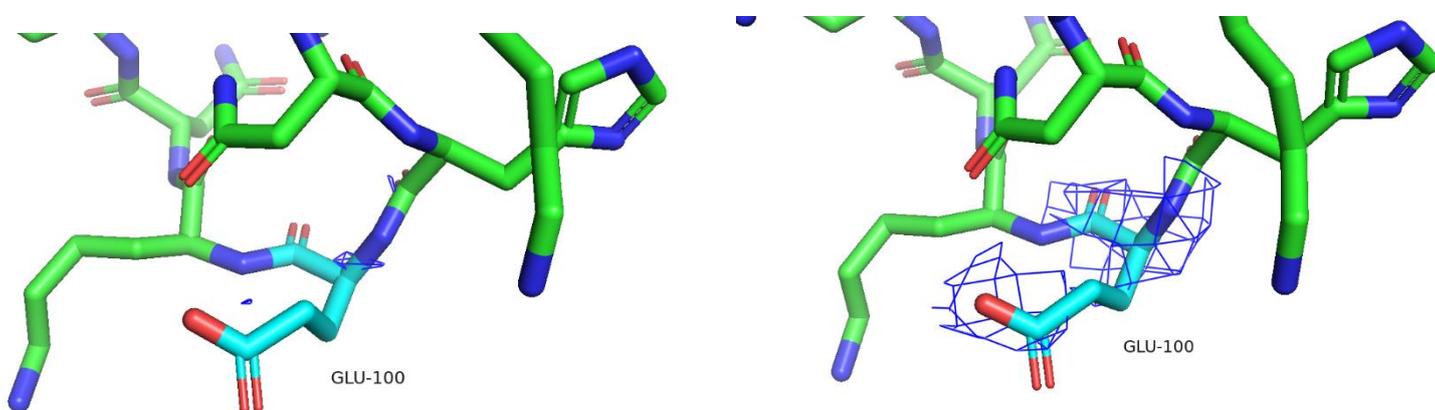


Рис. 5. Остаток GLU-100 и его электронная плотность на уровне подрезки 1 (слева) и на уровне подрезки 0.5 (справа).

Как можно заметить из картинок, электронная плотность на уровне подрезки 1 на данном остатке практически отсутствует. На уровне подрезки 0.5 она, конечно, есть, да и та довольно плохого качества и то не покрывает не весь остаток. Так что такие нетипичные параметры вполне могли возникнуть из-за неправильной интерпретации электронной плотности.

3. 222-й глутамин в цепи В

Данный остаток был определён как маргинальный по критерию карты Рамачандрана и по недопустимому ротамеру. Маргинальным он оказался по углу ψ – углу между $C\alpha$ атомом и $C=O$. У него он составил 19.6° , а для попадания в благоприятную область он должен быть примерно от -60° до -5° или 110° до 170° .

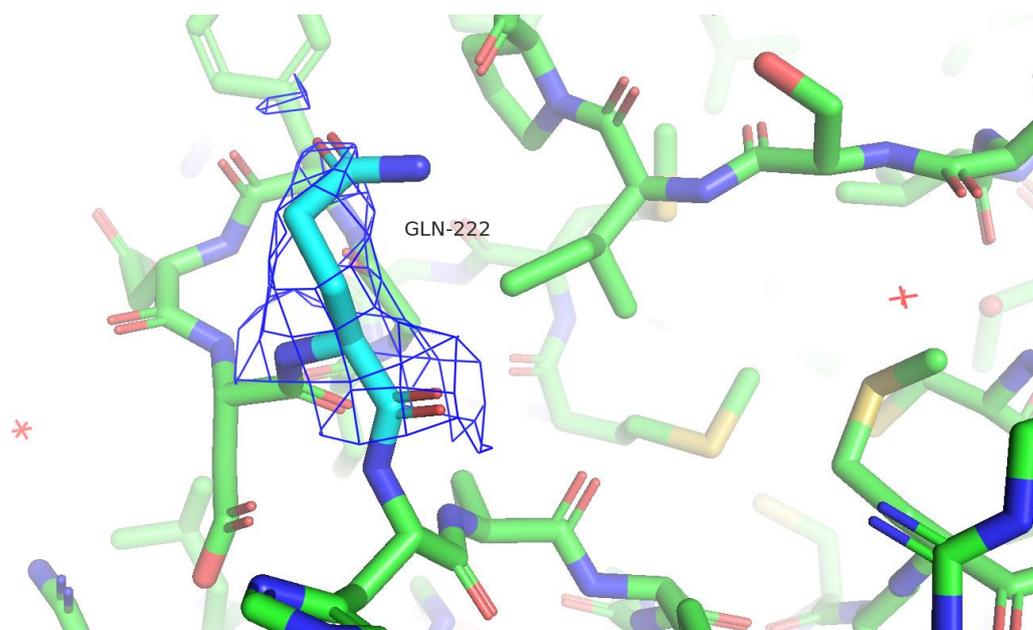


Рис. 6. Отображение остатка GLN-222 и его электронной плотности на уровне подрезки 1 и $\text{carve} = 1.5$

В целом можно заметить, что данный остаток неплохо покрыт электронной плотностью (в рамках данной модели, где качество в принципе не особо впечатляющее), разве что за исключением последних атомов радикала, при этом остов имеет вполне закономерные очертания. Вполне возможно, что такие нетипичные характеристики могут возникать из-за особенностей структуры белка, хотя и с некоторой натяжкой. Так же может создаться ощущение, что такой изгиб радикала может быть вследствие того, что он образует водородную связь с соседними остатками, однако, к сожалению, там куда он смотрит расположены лишь неполярные остатки.

4. 103-й глутамин в цепи В

Данный остаток был определён как маргинальный по плохому значению Z -score величины угла связи и как имеющий недопустимый ротамер. У данного остатка $RMSZ$ составлял -6.91 . Наблюдаемый угол составлял 92.33° , при том что в идеале он должен быть 111.00° . Подобных остатков было немного – всего 2, и они более не имеют других описанных критериев маргинальности. Всё же было интересно посмотреть, каким образом данный критерий будет находить отражение в структуре при визуализации.

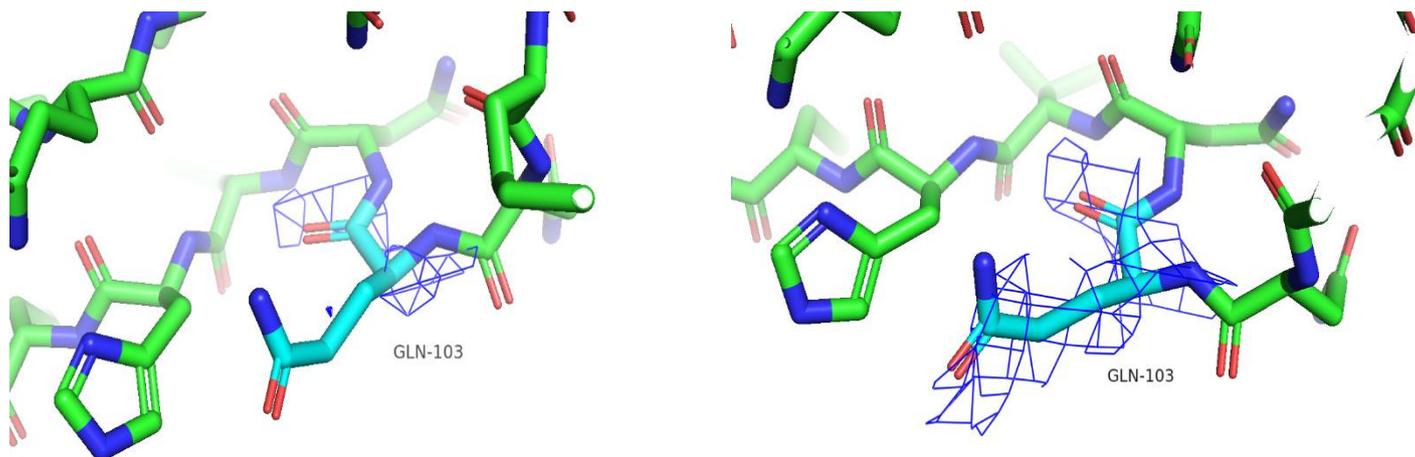


Рис. 7. Отображение остатка GLN-103 на уровне подрезки 1 (слева) и на уровне подрезки 0.5 (справа) и $\text{carve}=1.5$

В данном случае трудно будет что-то конкретное сказать об углах остатка по электронной плотности. Она практически отсутствует на уровне подрезки 1 и по ней невозможно определить очертания атомов, не говоря уже об углах связей. На уровне же 0.5 качество всё равно довольно плохое. Так что наблюдаемые характеристики вполне могут следовать из неверной расшифровки электронной плотности ввиду её посредственного качества.

5. 518-й аргинин в цепи А

Данный остаток был определён как маргинальный по оценке качества окружения, которая составила у него $-6.906 (<-5)$, а также по недопустимому ротамеру.

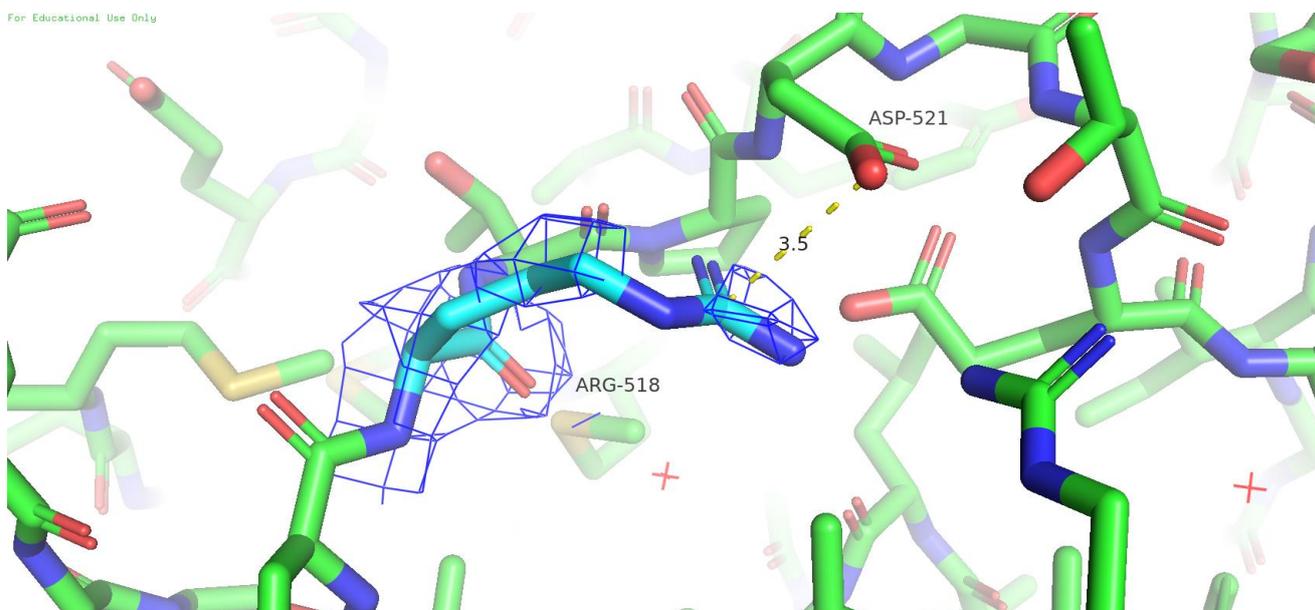


Рис. 8. Arg-518 с отображением электронной плотности на уровне подрезки 1 и $\sigma_{\text{cut}} = 1.5$ и его окружение

Данный остаток скорее всего не может образовывать водородных связей. Как вариант вполне возможно электростатические взаимодействия с остатком аспарагиновой кислоты – 521, хотя расстояние кажется «на грани». Электронная плотность с одной стороны вроде как и описывает очертания молекулы, а с другой стороны не все атомы покрывает, да и для уровня подрезки 1 как мне кажется качество довольно плохое. То есть вроде и хочется сказать, что наблюдаемые характеристики могут вытекать из особенности структуры, но всё же есть склонность считать, что в данном случае снова имеет место быть ошибка расшифровки.

6. 249-й глутамат в цепи А

Данный охарактеризован как маргинальный из-за плохой оценки качества окружения, которая составила у него -5.775 . И несмотря на то, что у него не обнаружено других параметров маргинальности, данный остаток оказался довольно любопытным. Он является полярным, и как можно видеть на картинке, он располагается далеко не на поверхности белка, а вполне смотрит внутрь, причём в сторону гидрофобных остатков, и без какой-либо возможности образовывать водородные связи.

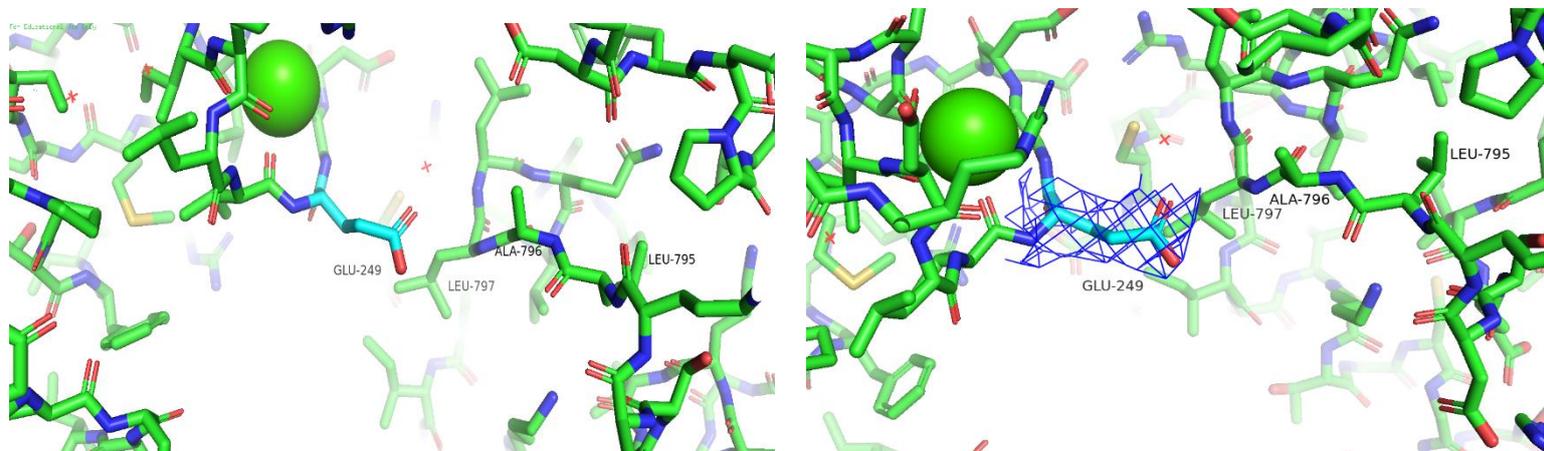


Рис. 9. GLU-249 с его окружением (слева) и его электронная плотность на уровне подрезки 1 и $\text{carve} = 1.5$ (справа)

И судя по электронной плотности, вписан данный остаток довольно неплохо, качество вполне нормальное (в рамках данной модели), так что здесь маловероятна ошибка расшифровки ЭП. Так что как мне кажется можно сказать, что в данном случае это может быть особенностью структуры, и этот остаток в таком странном положении возможно несёт какую-нибудь функциональную нагрузку.

Сравнение исследуемой модели PDB с моделью из PDB_redo

База данных PDB_redo содержит оптимизированные по различным параметрам версии моделей PDB.

При наложении двух моделей друг на друга можно увидеть, что они очень сильно друг на друга похожи, и практически идеально накладываются.

For Educational Use Only

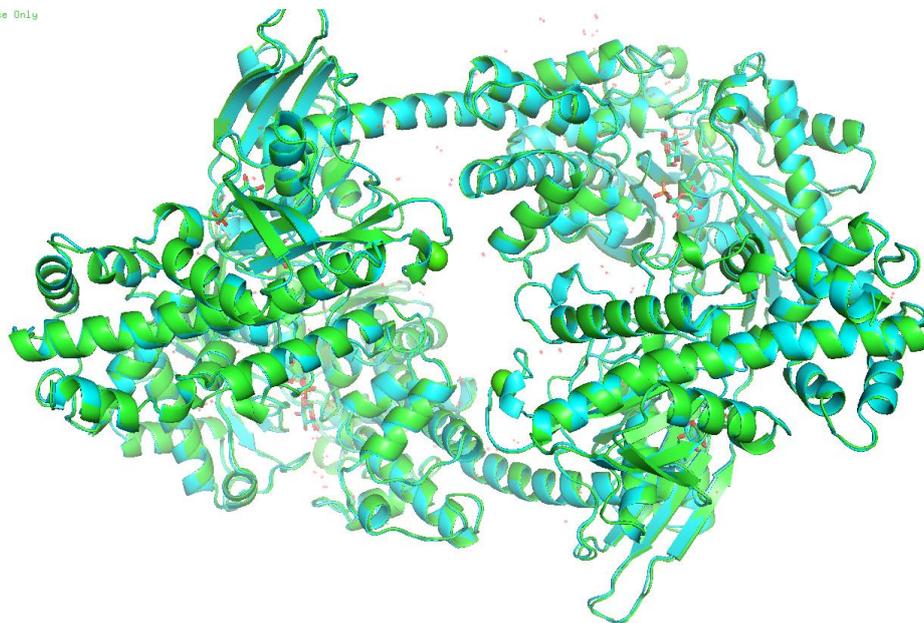


Рис. 10. Наложение друг на друга модели из PDB (циановая) и из PDB_redo (зелёная)

С одной стороны может действительно показаться что всё идеально, однако при более детальном рассмотрении можно заметить, что присутствуют некоторые различия во вторичной структуре, по типу не совсем совпадающих альфа-спиралей, причём в некоторых местах у PDB-модели за место спиральных участков стоят просто петли (рис 10)

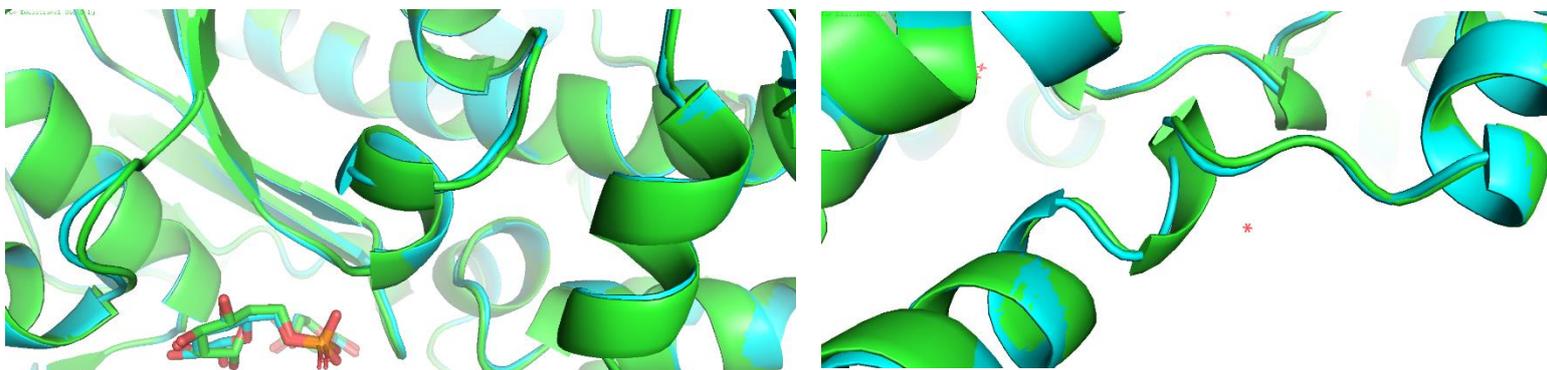


Рис. 11. Различия в структурах из PDB и PDB_redo

Далее приведена таблица со сравнениями параметров качества у модели из PDB и модели из PDB_redo.

Табл. 4. Сравнение качества структур PDB и PDB_redo

Параметр качества	PDB	PDB_redo
R	0,2745	0,1630
R-free	0,3319	0,2018
RMS Z-score длин связей	0,318	0,525
RMS Z-score величин углов	0,596	0,777
Качество карты Рамачандрана	6	22
Нормальность ротамеров	8	40
Число хорошо упакованных остатков	42	55
Достоверность водородных связей	0	18

Также в выводе PDB_redo были приведены боксплоты сравнения основных параметров моделей из PDB и PDB_redo.

Суммарно в модели PDB_redo было изменено 32 ротамера, убрано 23 молекулы воды, было изменено положение 7 пептидов, и аж 922(!) остатка стали лучше вписываться в свои электронные плотности, что в очередной раз характеризует весьма посредственное качество расшифровки электронной плотности.

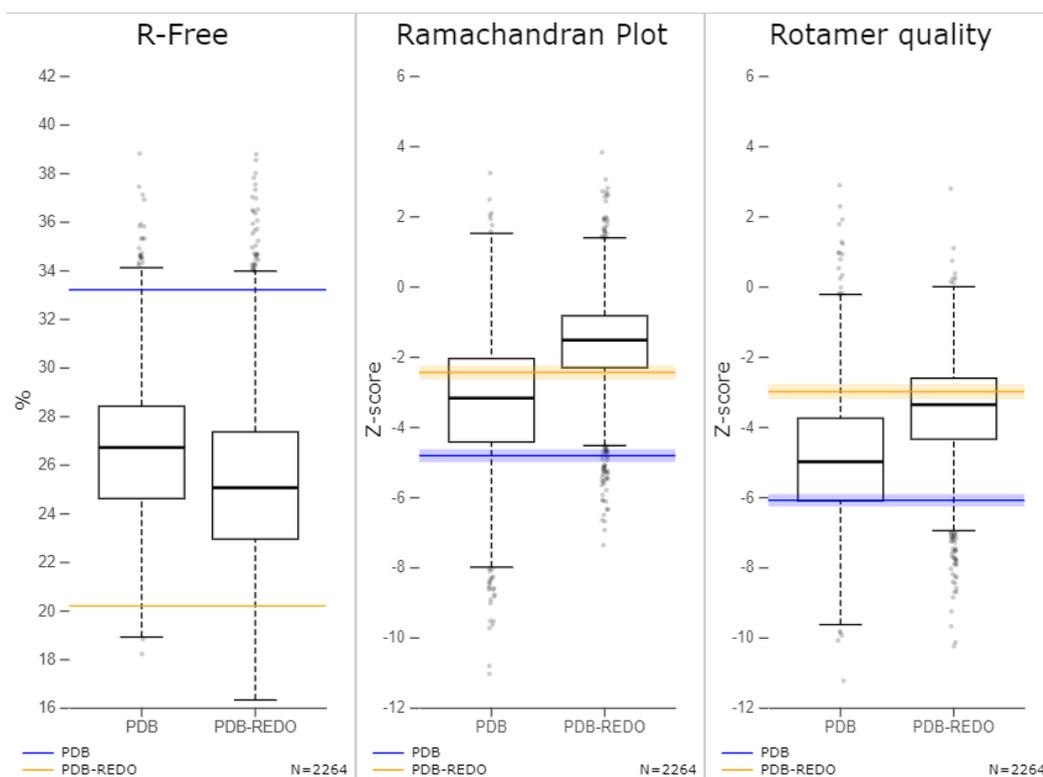


Рис. 12. Сравнение основных показателей качества расшифровки структуры для моделей из PDB и PDB_redo.

В целом можно заметить, что для модели из PDB_redo R-free уменьшился, однако Z-score карты Рамачандрана и качества ротамеров заметно вырос.

Заключение:

Подводя итоги анализа данной модели, хочется сказать, что на мой взгляд данная модель была расшифрована довольно не совсем хорошо. Было найдено довольно много недочётов. Многие остатки были определены как маргинальные по различным параметрам, и в большинстве исследованных частных случаях причина была в неправильной расшифровке электронной плотности, нежели в структурной особенности. При оптимизации модели с помощью ресурса PDB_redo было усовершенствовано довольно большое количество параметров, особенно в глаза сразу же бросилась достоверность водородных связей, которая у исходной модели 0, а у улучшенной 18. Да и к тому же помним, что целых 922 остатка лучше вписались в свою электронную плотность, что, как мне кажется, также подтверждает выдвинутое предположение о не совсем хорошем, я бы сказал даже, плоховатом, качестве исследуемой структуры.

Список литературы:

1. Ryuji Yokokawa, Shoji Takeuchi, Member, IEEE, Takahide Kon, Masaya Nishiura, Reiko Ohkura, Masaki Edamatsu, Kazuo Sutoh, and Hiroyuki Fujita, Hybrid Nanotransport System by Biomolecular Linear Motors. JOURNAL OF MICROELECTROMECHANICAL SYSTEMS, VOL. 13, NO. 4, AUGUST 2004
2. Denise Pollard-Knight & Athel Cornish-Bowden, Mechanism of liver glucokinase. Molecular and Cellular Biochemistry volume 44, pages71–80(1982)
3. Koshland D. Induced Fit and Hexokinase.Fig.7 Active Site of Glucokinase/Hexokinase.2010
4. Kitto B.G, Caras J., Caras P. Interactive Concepts in Biochemistry. Structural Tutorials.2008; Chp: 15.1, 15.5.
5. Alexander E Aleshin¹, Chenbo Zeng^{1†}, Gleb P Bourenkov², Hans D Bartunik², Herbert J Fromm¹ and Richard B Honzatko^{1*}, The mechanism of regulation of hexokinase: new insights from the crystal structure of recombinant human brain hexokinase complexed with glucose and glucose-6-phosphate. Structure 15 January 1998, 6:39–50

Ссылки на использованные ресурсы:

1. PDB: <https://www.rcsb.org/>
2. MolProbity: <http://molprobity.biochem.duke.edu/>
3. WHAT_CHECK: <https://swift.cmbi.umcn.nl/servers/html/>
4. PDB_redo: <https://pdb-redo.eu/>