

Отчет о качестве РСА расшифровки структуры человеческого пропердина

Автор: Станислав Тихонов

Аннотация

В данной работе проведен анализ качества расшифровки структуры 6S0B, полученной методом РСА. Проведен анализ качества структуры в целом, анализ маргинальных аминокислотных остатков и сравнение структур из PDB и PDB_redo.

Введение

Я выбрал белок пропердин, потому что он смешно называется. Помимо того, что он имеет такое название, данный белок еще участвует в опсонизации антигенов в рамках системы комплемента, таким образом способствуя усилению иммунного ответа. Структура, которую я анализировал (PDB ID: 6S0B) отличается от предыдущих структур этого белка более высоким разрешением [1].

Результаты и обсуждение

Общая информация о модели

Структура 6S0B содержит белок пропердин, состоящий из двух цепей (А и В), в комплексе с СТС-доменом компонента комплемента С3 (см Рис. 1). Кроме этого, пропердин О-гликозилирован дисахаридом, и в структуре еще присутствуют 2 ковалентно несвязанных типа моносахаридов.

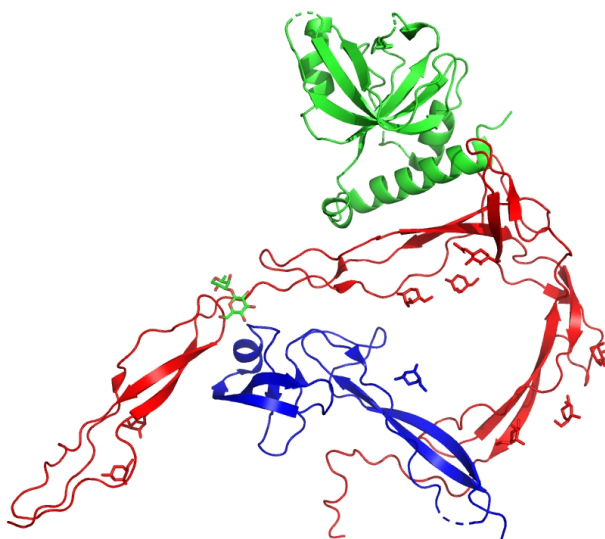


Рисунок 1. Общий вид структуры. Показаны цепь А (красная), цепь В (синяя) и СТС-домен компонента комплемента С3 (зеленый)(обозначен в структуре как цепь С).

Структура была загружена в PDB 14.06.2019 авторами van den Bos, R.M., Pearce, N.M., Gros, P. Статья, в которой авторы представили эту структуру и проанализировали взаимодействия

пропердина с компонентами комплемента была опубликована в журнале *Frontiers in Immunology* 04.09.2019 [1].

Для решения фазовой проблемы использовался метод молекулярного замещения (MR).

В файле структурных факторов представлены 21287 рефлексов. При этом в статье заявляется, что для этой структуры должно быть 21145 уникальных рефлексов, что совпадает с записью в файле *pdb*. В файле структурных факторов не уникальных рефлексов нет, но эти количества рефлексов различаются несильно, так что не думаю, что это может быть серьезной проблемой.

Разрешение рефлексов имеется от 2.312 Å до 63.081 Å с полнотой данных всего 68.6%, что нехорошо. Возможно можно было взять разрешение ближе к 2.5 Å но получить более высокую полноту. Итоговое разрешение структуры 2.31 ангстрем.

Ниже представлены параметры кристаллографической ячейки:

Таблица 1. Параметры кристаллографической ячейки.

Длина стороны (Å)	Угол между сторонами (°)
a = 71.471	$\alpha = 90$
b = 71.497	$\beta = 90$
c = 134.182	$\gamma = 90$
Кристаллографическая группа: P 21 21 21	

Значения индикаторов качества модели в целом

На сайте PDB представлены следующие процентильные оценки качества данной PCA модели по сравнению с другими структурами из PDB:

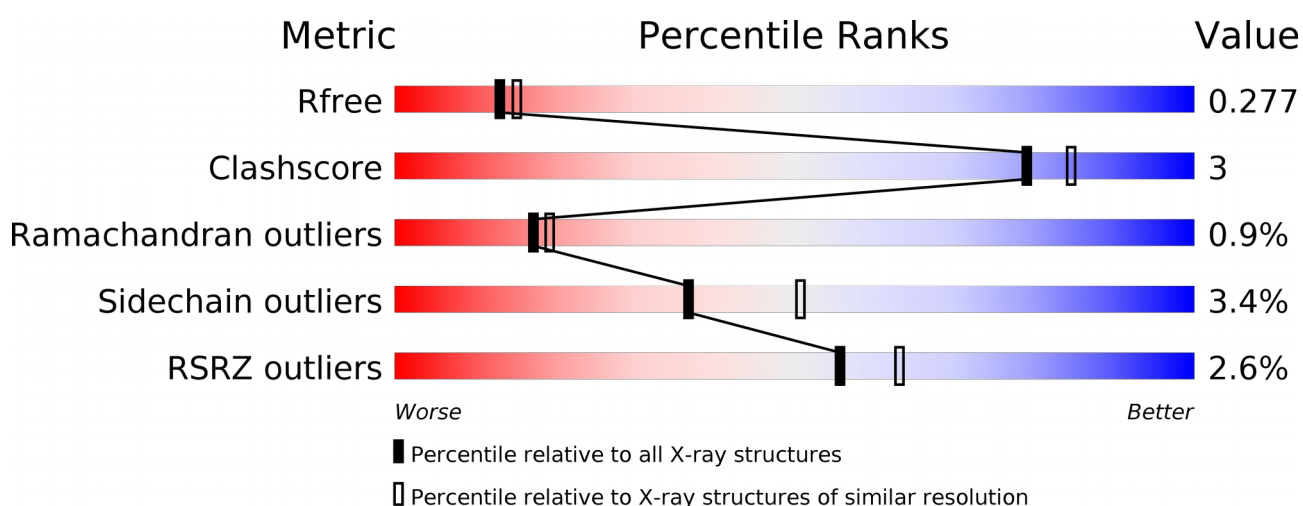


Рисунок 2. Процентильные оценки качества модели по сравнению с другими структурами из PDB.

R-фактор для рабочей выборки равен 0.230, R-free равен 0.277. Здесь видно, что не произошло переоптимизации. Сами значения R-факторов получились довольно большими, но достаточно хорошими для того, чтобы в принципе работать с этой структурой.

Карта Рамачандрана для аминокислотных остатков, отличных от глицина, пролина, валина и изолейцина, полученная с помощью сервиса MolProbity [2], показана ниже:

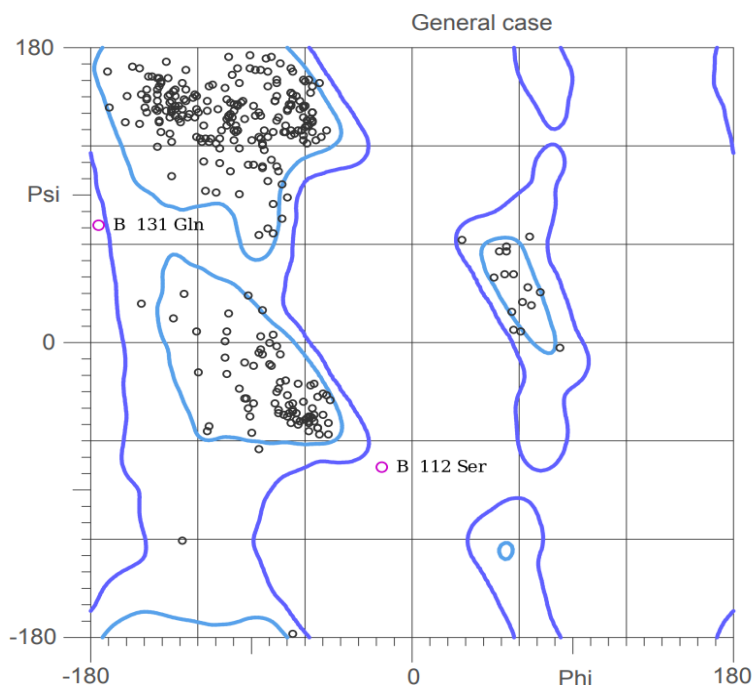


Рисунок 3. Карта Рамачандрана для всех аминокислотных остатков за исключением глицина, пролина, валина и изолейцина.

В разрешенную область не попали 4 остатка: Ser-112B (причем оба альт-лока), Gln-131B, Gly-396A и Ile-1520C. Процент таких остатков от всех составил 0.88%. В предпочтительных областях оказалось 95.6% остатков. Структура содержит 13 неротамеров (3.65%), и всего 87.64% предпочтительных ротамеров. Несмотря на большое количество вышеупомянутых маргинальных остатков, clashscore структуры (количество недопустимых наложений атомов на 1000 после добавления атомов водорода) составил всего 2.65, что позволяет отнести эту структуру к 100му перцентилю (самому лучшему) среди структур с похожим разрешением.

В процессе добавления атомов водорода MolProbity автоматически произвел инверсию боковой цепи для Gln-374A и Gln-1643C, что говорит о том, что они вероятнее всего были неправильно вписаны в структуру.

По данным автоматически сгенерированной PDB валидации в структуре имеется 12 (2.6%) RSRZ выбросов (т. е. остатков, пространственный R-фактор которых отличается от среднего для базы данных больше чем на 2 стандартных отклонения), что тоже не очень хорошо.

Таким образом, в целом модель имеет приемлемое, но не очень хорошее качество, а вышеупомянутые маргинальные остатки либо несут в себе некий биологический смысл, либо являются ошибками в расшифровке структуры.

Анализ маргинальных остатков

В таблице ниже приведены показатели качества некоторых маргинальных остатков:

Таблица 2. Некоторые маргинальные остатки.

Маргинальный остаток	Показатели качества
Ser-112B	Выброс на карте Рамачандрана (оба альт-лока) Неблагоприятная конформация боковой цепи

Маргинальный остаток	Показатели качества
	(оба альт-лока)
Gln-131B	Выброс на карте Рамачандрана Плохое значение RSRZ (2.9)
Ile-1520C	Выброс на карте Рамачандрана
Gly-396A	Выброс на карте Рамачандрана
Gln-374A	Подозрение на необходимость инверсии боковой цепи Перекрытие Ван-дер-Ваальсовых радиусов с His-375A на 0.74 Å (худшее перекрытие)
Gln-1643C	Подозрение на необходимость инверсии боковой цепи
Ser-90B	Неблагоприятная конформация боковой цепи Перекрытие Ван-дер-Ваальсовых радиусов с Cys-89B на 0.44 Å
Arg-330A	Неблагоприятная конформация боковой цепи Перекрытие Ван-дер-Ваальсовых радиусов с Gly-1660C на 0.5 Å
Cys-132B	Худшее значение RSRZ (6.6)

Значения RSRZ и перекрытия Ван-дер-Ваальсовых радиусов взяты из автоматически сгенерированной базой данных PDB валидации.

Далее следует анализ 5 маргинальных остатков из данной таблицы.

Ser-112 цепи B

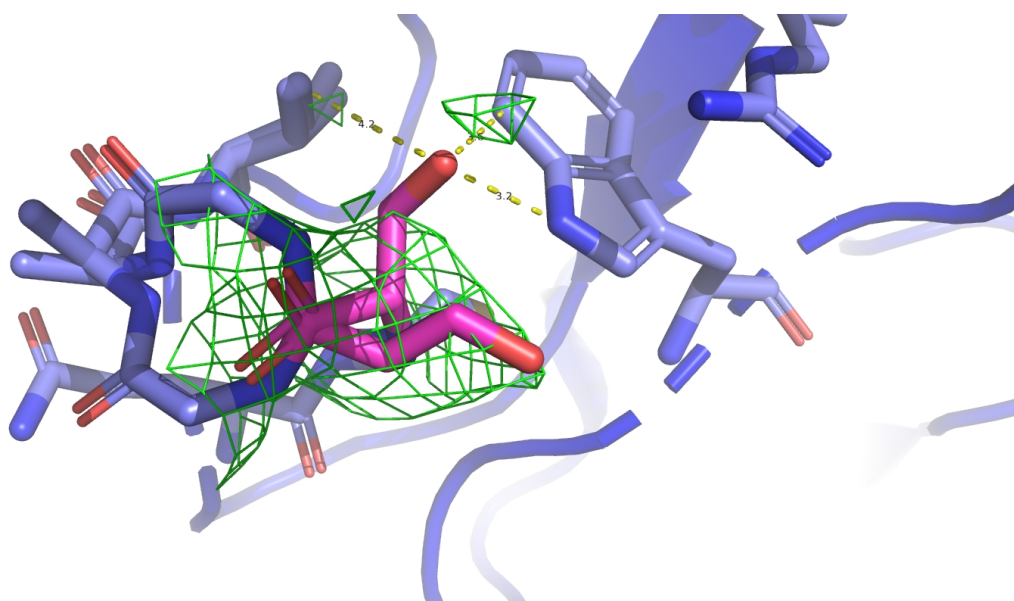


Рисунок 4. Окружение Ser-112 цепи B (выделен цветом magenta). Сеткой обозначена электронная плотность остатка уровня 1 с подрезкой на уровне 0.5.

На рисунке выше видно, что данный серин располагается в петле очень близко к гидрофобному участку, но одновременно на поверхности белка, тем самым он

“отгораживает” гидрофобную часть белка от гидрофильного окружения. Мне кажется, что такую невыгодную конформацию ему приходится принимать для того, чтобы убрать ОН группу максимально далеко от гидрофобных остатков, но при этом образовать водородную связь с триптофаном. Невыгодная конформация может быть еще объяснена просто необходимостью вписать остаток в электронную плотность с неидеальным разрешением и качеством. То есть в случае получения более качественной структуры этот серин, возможно, и не был бы маргинальным.

Gln-374 цепи A

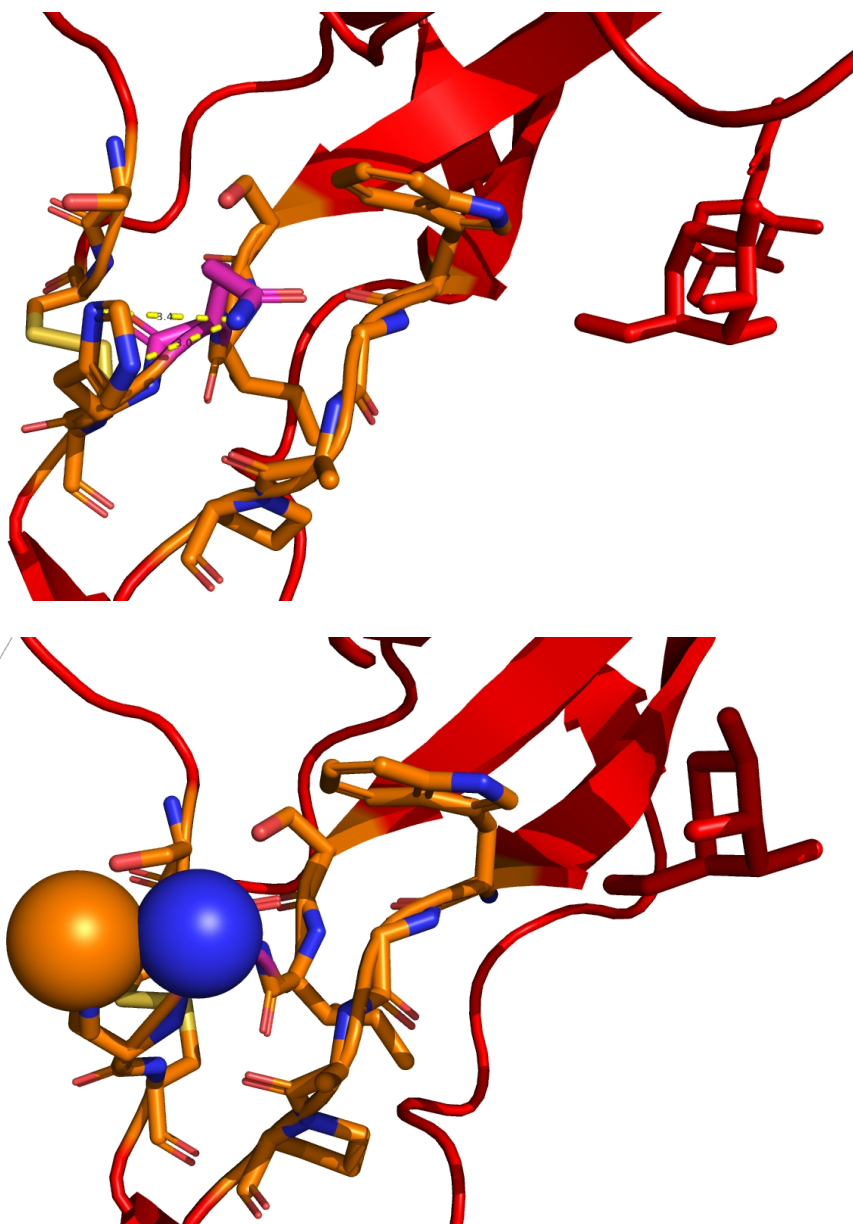


Рисунок 5. Водородные связи Gln-374 цепи A (покрашен в цвет magenta) с His-375 цепи A, которые могут образоваться, если сделать инверсию боковой цепи для Gln-374 (сверху), и перекрытие Ван-дер-Ваальсовых радиусов CE1 His-375 и NE2 Gln-374 (снизу).

Ситуация с глутамином 374 цепи A – это типичная ошибка авторов структуры, которые просто не посмотрели на этот остаток (в статье он не упоминается, так как, видимо, не очень важен для функции белка). Здесь очевидно, что глутамин 374 мог бы образовывать водородные связи с азотами гистидина 375, если бы смотрел на него кислородом, а не NH₂ группой (рис. 5, картинка сверху). Из-за того, что он смотрит на гистидин NH₂ группой, то

атом углерода в кольце гистидина сильно толкается с протонами на азоте глутамина (рис. 5, картинка снизу). Если бы глутамин был повернут к гистидину кислородом, то из кислорода не торчали бы два протона и такого перекрытия Ван-дер-Ваальсовых радиусов не было бы. Разумеется, в реальной молекуле глутамин смотрит на гистидин кислородом, но алгоритм вписывания структуры в электронную плотность этого не учел (так как, скорее всего, смотрел на длины связей, углы, и совпадение плотности модели наблюдаемой плотности, что здесь, в общем-то, соблюдено)).

Arg-330 цепи А

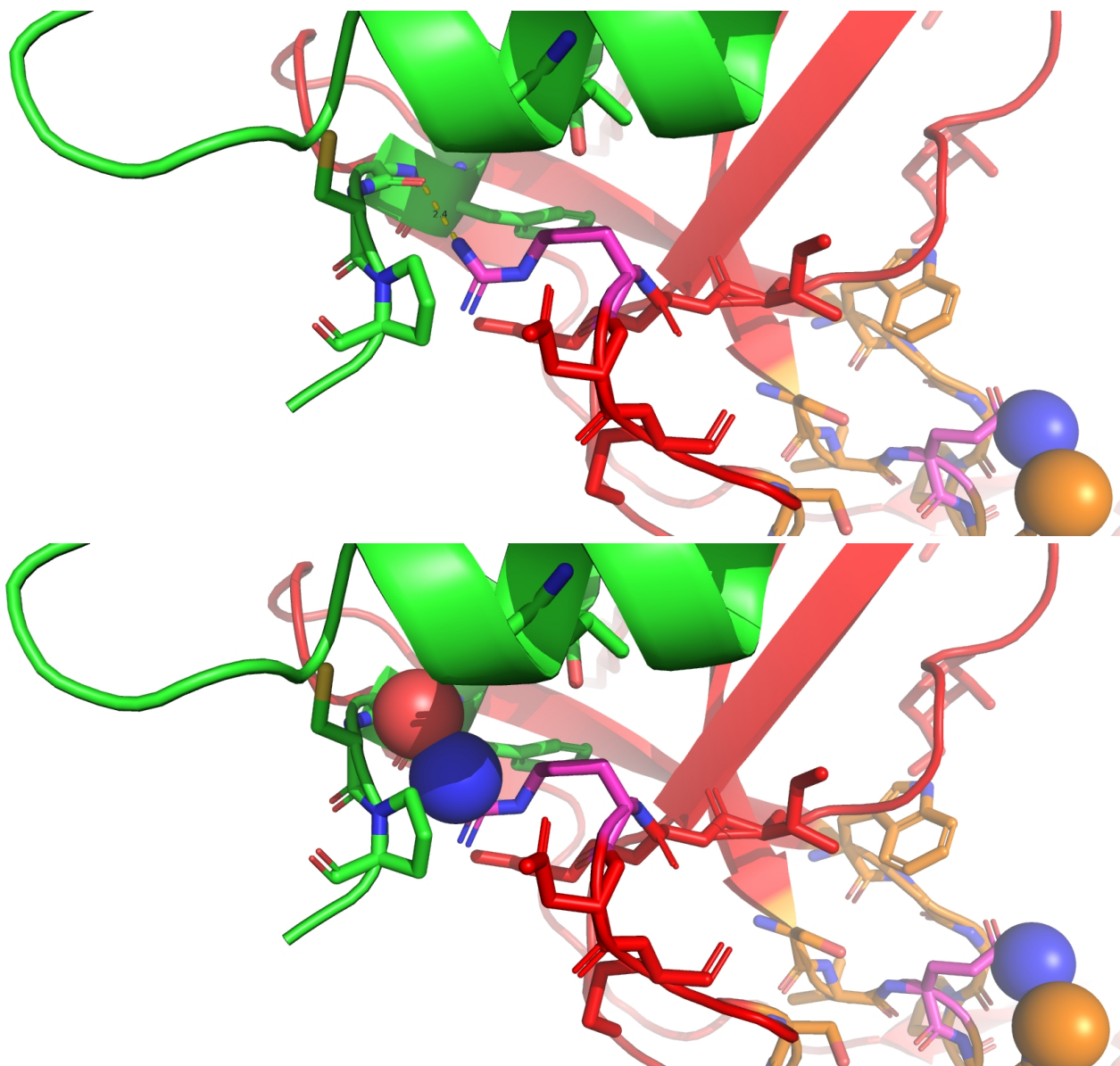


Рисунок 6. Arg-330 цепи А образует водородную связь с глицином 1660 цепи С (СТС-домена компонента комплемента С3)(сверху), он же перекрывается радиусом атома азота с атомом кислорода глицина (снизу).

Данный аргинин, скорее всего, просто ввиду остальной структуры белков слишком близко расположен к глицину 1660 СТС-домена компонента комплемента С3. В статье написано, что аргинин образует водородную связь с глицином (мы это видим) [1], и она всего 2.4 Å в длину из-за близкого положения аминокислот. Невыгодная конформация боковой цепи аргинина

может объясняться тем, что его боковая цепь пытается “сжаться” и отодвинуть азот от кислорода глицина.

Ser-90 цепи В

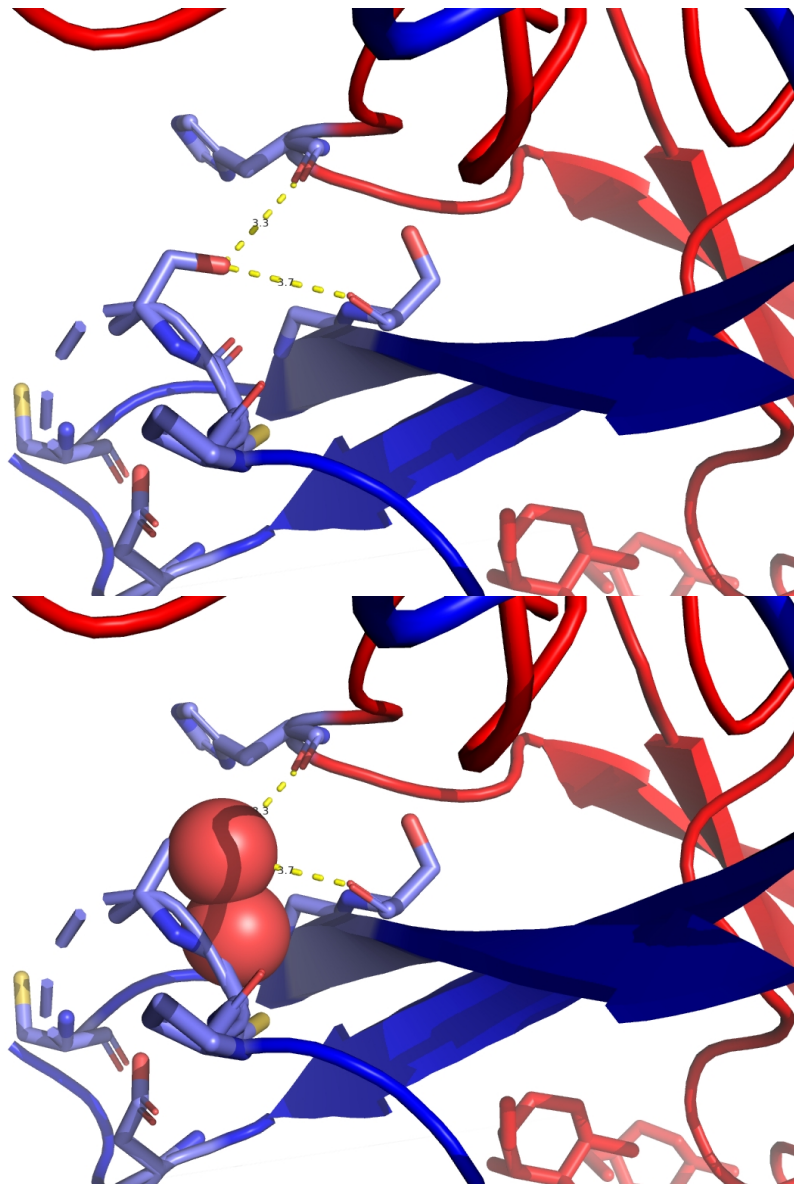


Рисунок 7. Возможные водородные связи Ser-90 цепи В с основными кислородами других остатков (сверху), перекрытие Ван-дер-Ваальсовых радиусов кислорода боковой группы серина 90 и основного кислорода цистеина 89 (снизу).

В данном случае серину выгодно образовывать водородную связь с одним из остатков, показанных на Рис. 7, но при этом он толкается с основным кислородом цистеина. Такое перекрытие радиусов может быть оправдано только в случае водородной связи серина 90 именно с серином 97 цепи В (правый остаток на рисунке), так как в случае связи с верхним остатком серин 90, скорее всего, мог бы избежать перекрытия радиусов. Но, видимо, водородная связь именно с правым остатком в целом выгоднее для структуры, даже несмотря на перекрытие радиусов. В статье авторы тоже говорят о водородной связи серина 90 с серином 97 (который справа)[1].

Cys-132 цепи В

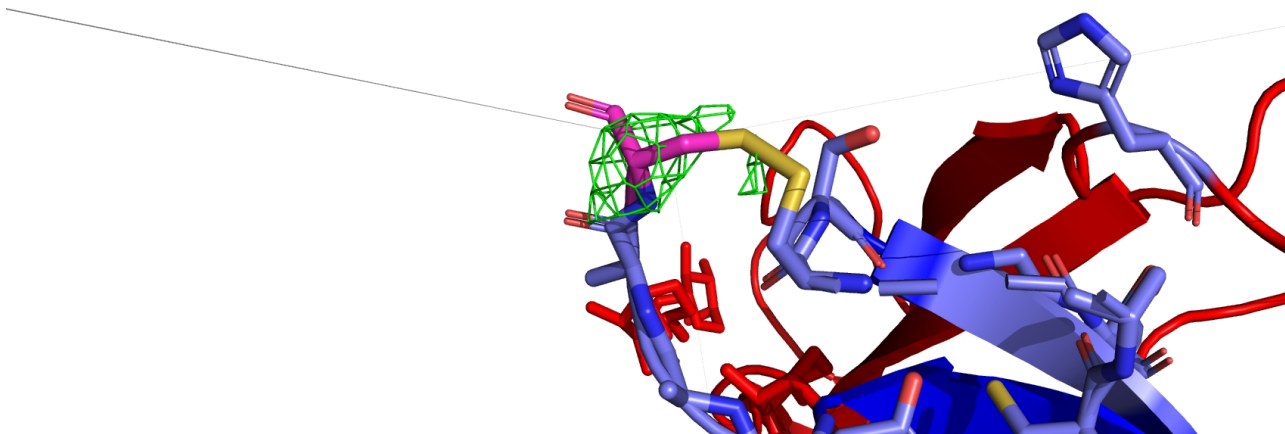


Рисунок 8. Электронная плотность С-терминального цистеина 132 цепи В (показан цветом magenta) уровня 1, связанного дисульфидным мостиком с цистеином 93 цепи В.

Этот цистеин находится на С-конце цепи В. В статье говорится, что у конструкций пропердина, которые они экспрессировали, должен быть еще цистеин 133 (они экспрессировали несколько конструкций), но по какой-то причине конструкция той структуры, которую я анализирую, получилась обрезанной на цистеине 132. В необрезанных конструкциях цистеин 132 образует дисульфидную связь с цистеином 170, что считается канонической структурой пропердина, но в данной структуре цистеин 132 связан дисульфидным мостиком с цистеином 93. Авторы статьи пишут, что это сопровождается увеличенной неупорядоченностью С-конца цепи. Все это (а особенно увеличенная неупорядоченность), я полагаю, могло повлиять на качество структуры на С-конце, что могло привести к такому низкому качеству вписывания цистеина 132 в электронную плотность (RSRZ 6.6, худшее значение по структуре).

Анализ структуры из PDB_redo

Вписанная в электронную плотность структура из PDB_redo [3] оказалась немного лучше по Rfree (0.2674 вместо 0.2762), и по z-score карты Рамачандрана (22 перцентиль вместо 21) и качества ротамеров (44 перцентиль вместо 37):

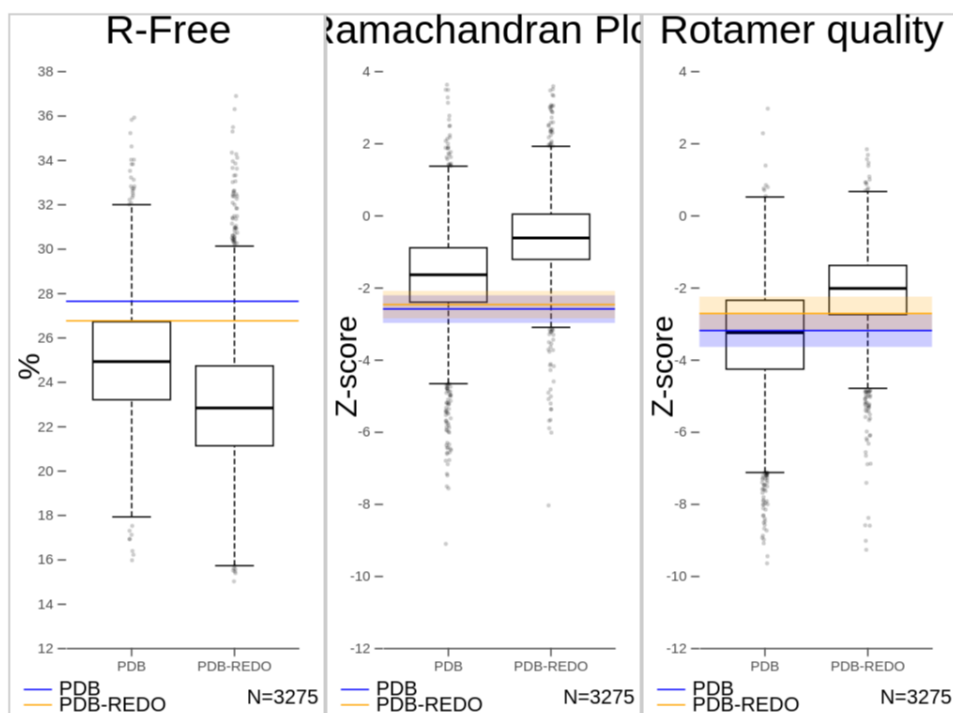


Рисунок 9. Сравнение моделей из PDB

Из рассмотренных мной маргинальных остатков ситуация улучшилась для Arg-330 цепи А – его гуанидиновая группа развернулась и теперь он меньше перекрывается Ван-дер-Ваальсовыми радиусами с глицином 1660:

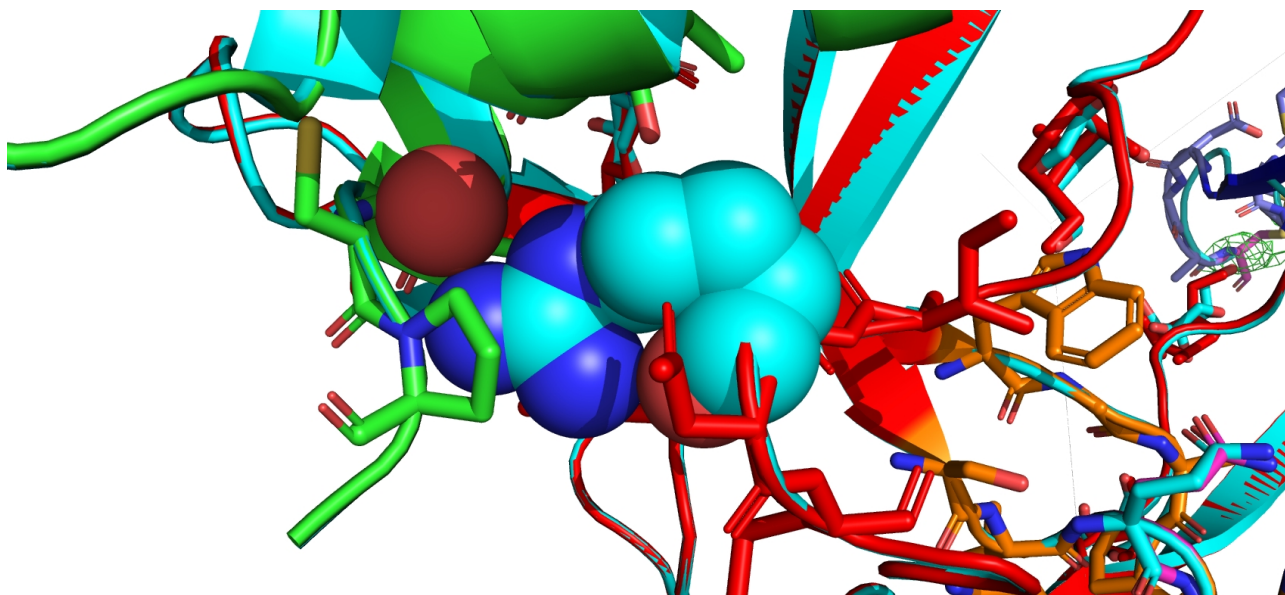


Рисунок 10. Перекрывание Ван-дер-Ваальсовых радиусов аргинина 330 цепи А и глицина 1660 цепи С в структуре из PDB_redo (гуанидиновая группа развернута, перекрывание уменьшилось).

Из этого можно сделать вывод, что маргинальный аргинин 330 цепи А – это все-таки недоработка алгоритма авторов оригинальной структуры PDB, а не биологически осмысленный маргинальный остаток.

Аналогичная ситуация состоит и с серином 90 цепи В. Вот он в структуре из PDB_redo (структура из PDB_redo показана циановым):

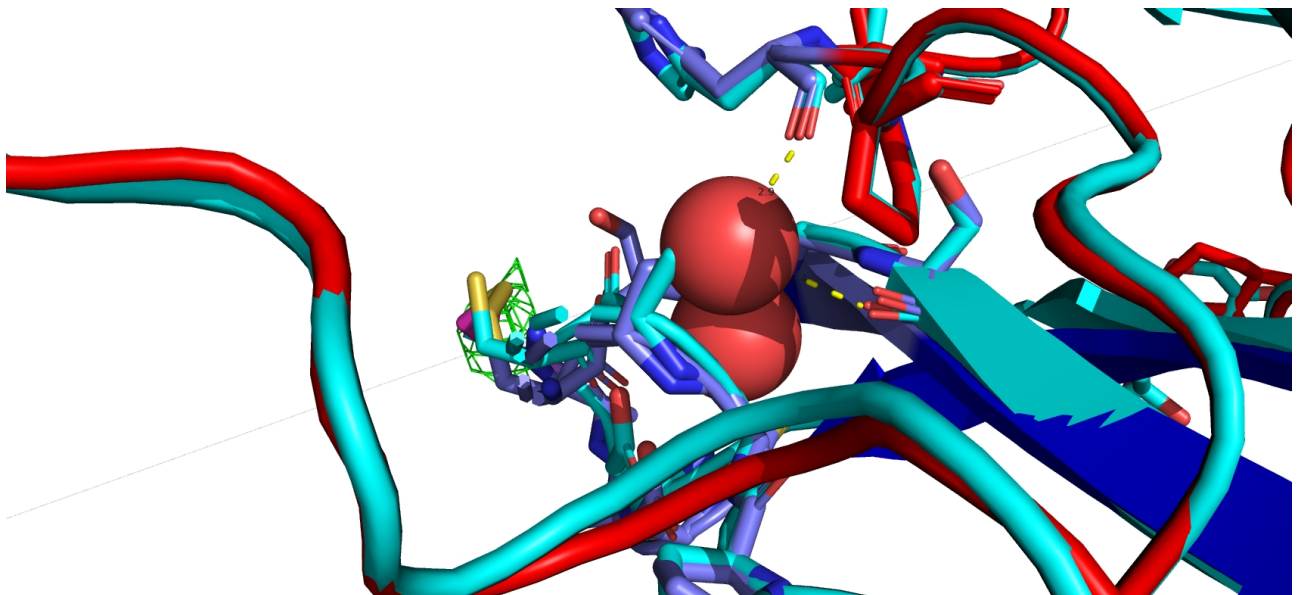


Рисунок 11. Ser-90 цепи В теперь меньше перекрывается радиусами с остовным кислородом цистеина 89. Водородная связь в этой структуре, скорее всего, должна быть с верхним остатком (гистидином 457 цепи А), а не с правым (серин 97 цепи В).

На картинке видно не очень хорошо, но визуально можно определить, что перекрывание Ван-дер-Ваальсовых радиусов здесь меньше. Потенциальная водородная связь с верхним остатком стала длиной 2.9 Å (вместо 3.3 Å), а с левым – 3.7 Å (вместо 3.4 Å). Таким образом, если бы авторы в оригинальной статье анализировали структуру из PDB_redo, то они бы не писали про связь Ser-90 с Ser-97 (правая), а писали бы про связь Ser-90 с His-457 (верхняя). Так как водородная связь теперь с гистидином 457 цепи А, перекрывание Ван-дер-Ваальсовых радиусов больше не требуется для успешной водородной связи, поэтому оно и меньше.

Заключение

Изначальная структура из PDB имеет приемлемое качество для анализа, хотя полнота набора гармоник довольно низкая и в структуре много маргинальных остатков. После анализа этих остатков и структуры из PDB_redo стало ясно, что многие маргинальные остатки на самом деле не несут в себе биологического смысла, а являются лишь следствием плохого алгоритма вписывания структуры в электронную плотность (Arg-330 и Ser-90) и халатности авторов структуры (Gln-374). Если кому-то потребуется работать с этой структурой, то это в принципе не лишено смысла, но лучше пользоваться вариантом из PDB_redo с исправленными боковыми цепями инвертированных глутаминов. В шутку про эту структуру можно сказать следующее: как белок называется, такая у него и структура.

Литература и ссылки

[1] Gros, Piet, et al. "Insights into enhanced complement activation by structures of properdin and its complex with the C-Terminal domain of C3b." *Frontiers in Immunology* 10 (2019): 2097.

[2] <http://molprobit.biochem.duke.edu/index.php?MolProbSID=par9q56bsukhcvmi9mvgvjgs11&eventID=144>

[3] <https://pdb-redo.eu/>