

Отчёт по качеству РСА-расшифровки структуры карликового тетрамера Ndc80 (PDB_ID = '5TCS')

Потапенко Елена, 4 курс ФББ

Аннотация

В данной работе анализируется качество модели 5TCS карликового тетрамера Ndc80 дрожжей, полученной рентгеноструктурным анализом. Рассматриваются основные показатели качества модели в целом и отдельные проблемные (маргинальные) остатки.

Введение

Кинетохоры — это сложные комплексы, соединяющие центромерную ДНК и микротрубочки веретена деления. Согласно статье [1], в которой впервые опубликована модель 5TCS, Ndc80c — это гетеротетрамерный комплекс, устанавливающий связь вдоль главной оси кинетохоры. Он состоит из белков Ndc80, Nuf2, Spc24 и Spc25, образующих два гетеродимерных субкомплекса, каждый из которых имеет глобулярную головку и стержень. Стержни двух субкомплексов соединяются концами, образуя один длинный стержень с глобулярными головками на обоих концах. Такое соединение (тетрамерный переход) является консервативным для всех эукариот; предполагается, что это соединение имеет некую функциональную роль, кроме механического соединения двух концов в более длинный стержень. Авторы предполагают, что эта роль может быть связана с регуляцией. Для изучения структуры соединения авторы статьи синтезировали два укороченных (карликовых) комплекса, в которых оставлены без изменений глобулярные концы и область тетрамерного перехода, но укорочены альфа-спиральные стержни. 5TCS – структура более короткого из этих двух комплексов.

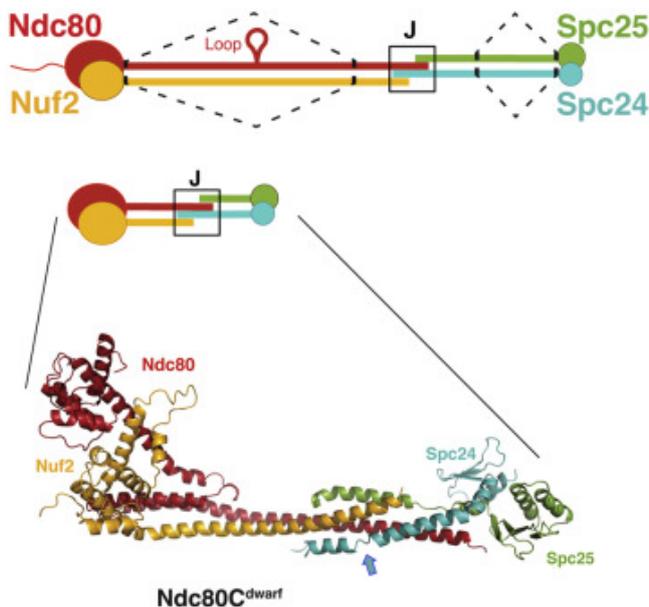


Рисунок 1. Комплекс Ndc80 состоит из 4 мономеров: Ndc80, Nuf2, Spc25, Spc24. Сверху на рисунке приведена схема строения обычного комплекса, чуть ниже — схема строения карликового комплекса. Снизу показан вид карликового комплекса в виде cartoons [1].

Результаты и обсуждение

1. Общая информация о модели

Структура была получена авторами Valverde, R.; Harrison, S.C из Медицинского Института Говарда Хьюза (ННМИ) и опубликована [1] в 2016 году вместе со структурой 5TD8, которая тоже является карликовой версией комплекса Ndc80с, но менее укороченной относительно природного комплекса, чем структура 5TCS.

Одна асимметрическая ячейка содержит один комплекс, состоящий из 4 уникальных белковых цепей. Кристаллическая ячейка прямоугольная ($\alpha = \beta = \gamma = 90$), имеет линейные размеры в ангстремах: $a = 169.38$, $b = 186.59$, $c = 122.04$.

Группа симметрии $C 2 2 2_1$. Буква C в обозначении группы симметрии означает, что группа базоцентрированная. Три цифры после буквы означают оси второго порядка. Двойки без индексов — это простые поворотные оси второго порядка, идущие параллельно двум осям кристаллической ячейки, двойка с индексом один — это винтовая ось второго порядка. Сам комплекс асимметричен, поэтому некристаллографических симметрий нет.

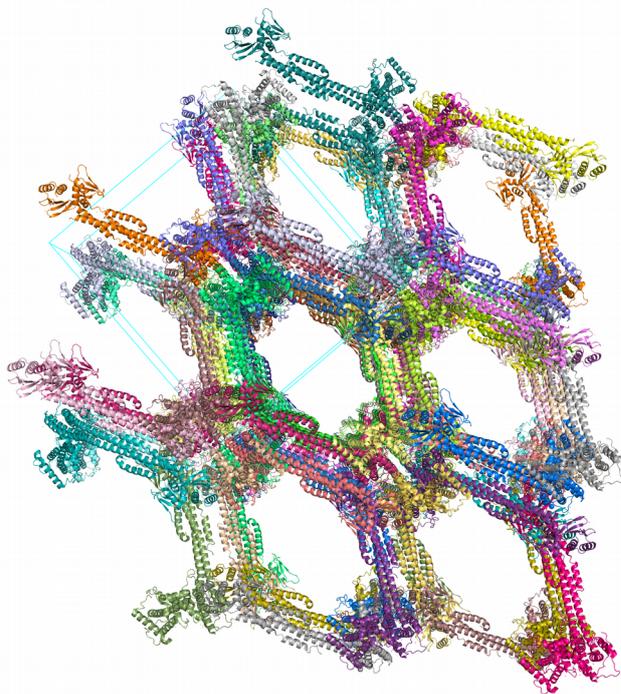


Рисунок 2. Вид кристалла вдоль (короткой) оси c. Голубыми линиями показана ячейка.

Всего было измерено 52775 уникальных рефлексов в диапазоне разрешений от 2.703 до 125.415 ангстрем (по данным файла со структурными факторами), полнота данных 99.3% (т. е. теоретически в этом диапазоне можно измерить 53157 рефлексов). Несмотря на хорошее значение полноты, общее разрешение структуры указано не по гармонике с лучшим разрешением (2.703), оно равно 2.83 ангстрем.

Фазовая проблема была решена методом SAD (single-wavelength anomalous diffraction, одноволновое аномальное рассеяние).

2. Значения индикаторов качества модели в целом

Параметр R-фактор этой модели равен 0,239, а R-free = 0,266 (контрольных рефлексов было 2234, что составляет примерно 5% от всех измеренных). Оба значения достаточно хорошие, R-free немного больше, чем R, поэтому можно считать, что модель не переоптимизирована. На рисунке 3 показано положение значения R-free относительно этого же параметра у других моделей со схожим разрешением. Видно, что R-free модели близок к медианному значению этого параметра среди моделей с похожим разрешением, однако относительно всех моделей в банке значение R-free этой модели не очень хорошее (верхний квартиль). Возможно, это связано с тем, что уже многие структуры в банке имеют разрешение лучше, чем 2.83Å, а модели с лучшим разрешением имеют меньшие значения и R-фактора, и R-free (так как экспериментальная электронная плотность лучше соответствует модели).

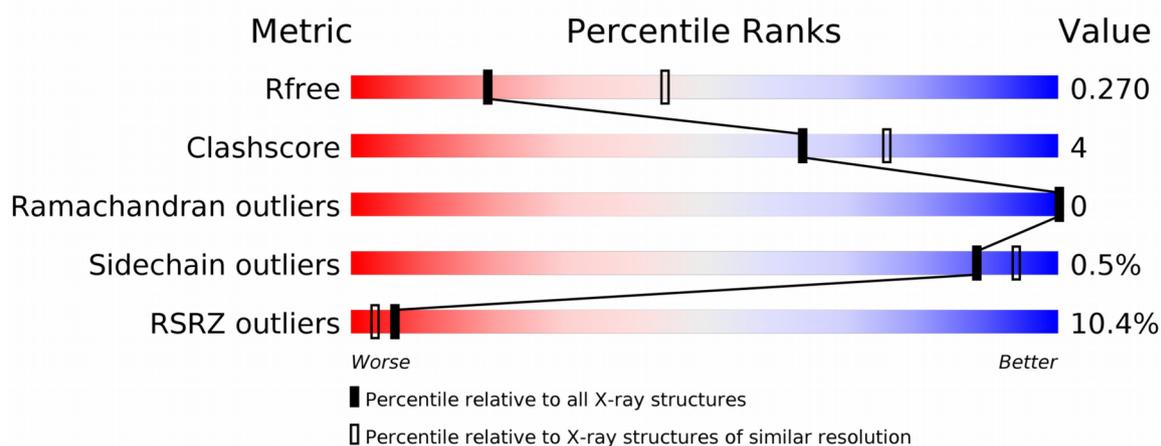


Рисунок 3. Некоторые показатели качества модели относительно характеристик других моделей. Рисунок взят из pdb report.

Карты Рамачандрана были построены с помощью сервиса MolProbity [2]. Модель не содержит остатков, маргинальных по углам остова: все 680 аминокислотных остатков лежат в разрешённой области, из них 628 (92.4%) лежат в предпочтительной области. Показатель доли остатков в предпочтительной области достаточно низкий: сервис MolProbity считает хорошим показателем 98% остатков в предпочтительной области и более, 92.4% сильно меньше. Некоторые из этих остатков будут рассмотрены подробнее.

Карта Рамачандрана для всех остатков представлена на рисунке 4. Отдельные карты были построены также для пролина (как циклической аминокислоты), для остатков, идущих в цепи перед пролином, для валина и изолейцина (они крупные, поэтому имеют немного другой диапазон разрешенных углов). Как и ожидалось, аутлаеров найдено не было (0% ramachandran outliers указано в отчёте pdb, см. рисунок 3).

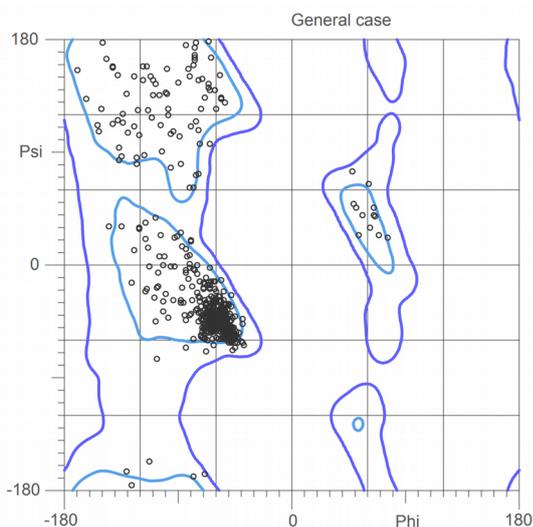


Рисунок 4. Карта Рамачандрана для аминокислотных остатков модели. По осям — углы ψ и ϕ , каждая точка соответствует одному остатку из модели. Голубым обозначены предпочтительные области, фиолетовым — разрешённые области. Видно, что все точки лежат внутри разрешённых областей.

Ниже приведены некоторые другие оценки качества модели:

- ClashScore (число недопустимых наложений атомов в пересчёте на тысячу атомов модели. Недопустимым в данном случае считается наложение Ван-дер-Ваальсовых атомов на 0.4 ангстрема и больше) = 4.24, это 137 пар атомов. Почему-то эту характеристику сервис считает как 100-й (лучший) перцентиль среди моделей с разрешением $2.83\text{\AA} \pm 0.25\text{\AA}$, хотя в отчёте *pdb* это значение пусть и принадлежит лучшему квартилю среди похожих разрешений (2.86-2.82), но явно не является лучшим. Думаю, так получается из-за того, что эти перцентили считаются по разным выборкам (диапазон разрешений у MolProbity значительно шире). В любом случае, значение ClashScore у модели достаточно хорошее.
- У модели не очень хорошие характеристики по встречаемости отклонений боковой цепи (poor rotamers): 2% против ожидаемого 0.3%. Боковых цепей, почти не отклоняющихся от ротамера, наоборот, меньше, чем нужно: favored rotamers = 89% (против целевого 98%).
- У модели очень плохое (относительно всех PCA-моделей и среди моделей с похожим разрешением) значение доли RSRZ-outliers = 10.4%. Получается, что каждый десятый остаток не очень хорошо вписывается в свою электронную плотность (и может считаться маргинальным). Всего модель содержит 70 остатков с RSRZ > 2. Самое большое значение имеет Ser-150 (chainA) с RSRZ = 19.7
- Отклонений длин связей и углов связей в модели нет, C-beta отклонений (отклонений первого атома радикала от ожидаемого положения) больше 0.25 ангстрем тоже нет (согласно выдаче MolProbity).
- Во время работы сервиса MolProbity не было найдено гистидинов, аспарагинов и глутаминов, подозреваемых в неправильном (нуждающимся во флипе) положении боковой цепи.

3. Маргинальные остатки в структуре

В таблице 1 приведены (некоторые) маргинальные остатки модели. Они отобраны из таблицы с характеристиками всех остатков из выдачи MolProbity. Интересно, что в среднем у остатков, имеющих плохое значение какого-нибудь критерия, среди соседей (по последовательности) больше остатков, тоже имеющих плохие значения. Статистически это предположение я не проверяла, но оно кажется мне вполне логичным: если отклонение характеристики связано с какой-то биологической ролью (например, остаток в активном центре изогнут так, что выбивается из предпочтительной области на карте Рамачандрана), соседние ему остатки также могут быть задействованы в каких-то процессах. Если же отклонение вызвано плохим соответствием модели и электронной плотности в этом месте, скорее всего, плохое соответствие модели тоже будет распространяться не только на единственный остаток, но и на его соседей.

Я старалась отбирать остатки, подозрительные хотя бы по нескольким критериям, и старалась брать разные примеры. Точный критерий, по которому отобран каждый остаток, указан в таблице; порядок остатков в таблице случайный.

Таблица 1. Список маргинальных остатков.

Цепь	Номер и тип остатка	Критерий маргинальности.
C	194 Leu	Клэш 0.61 Å + аутлаер по не-ротамеру боковой цепи + CaBLAM disfavored* + RSRZ = 7.1
D	186 Pro	CA Geom Outlier* + клэш с 185 His + разрешенная (но не предпочтительная) область на карте Рамачандрана
B	32-35	По отдельности каждый из 4х остатков не вызывает особых подозрений: они перекрываются радиусами с другими атомами (имеют клэши), двое из них лежат в непредпочтительной области карты Рамачандрана. Но из-за того, что 4 подряд идущих остатка имеют клэши, появляется ощущение, что пептидная цепь в этом участке идёт как-то по-другому. + 35й остаток имеет RSRZ = 5.3
B	40 Tyr	Аутлаер по не-ротамеру боковой цепи + клэш 0.48 Å с Asn-32.
B	450 Mse	Клэш 0.42 Å с Gln-451 + аутлаер по не-ротамеру + разрешенная (но не предпочтительная) область на карте Рамачандрана
B	71 Gly	RSRZ = 8.7 + CaBLAM Outlier + разрешенная (но не предпочтительная) область на карте Рамачандрана
A	148 Pro	Twisted Pro + Клэш 0.5 Å с 147 His + CaBLAM Outlier + разрешенная (но не предпочтительная) область на карте Рамачандрана
A	147 His	RSRZ = 13.0 (на втором месте в списке аутлаеров по этой характеристике) + клэш 0.5 Å с 148Pro
A	150 Ser	RSRZ = 19.7 (самое большое значение среди остатков модели), высокий B-фактор, остальные параметры нормальные.

* вообще говоря, CaBLAM – это метод аннотации вторичной структуры. В данном случае он используется (в составе сервиса MolProbity) не для аннотации, а для валидации положения Ca-атома с точки зрения участия всего остатка во вторичной структуре. Соответственно, аутлаеры по этому параметру — это остатки, чья геометрия Ca атома плохо соотносится с контурами предполагаемой вторичной структуры.

Далее 5 из этих остатков будут рассмотрены более подробно.

Маргинал 1. chain B, 450 Mse.

Меня этот остаток заинтересовал (среди других маргинальных) тем, что это нестандартная аминокислота. Мне показалось, что, возможно, плохие параметры связаны именно с нестандартностью (так как, например, остатков такого типа мало в базе), а не с проблемным участком расшифровки.

Это селенометионин, он обычно обозначается Se-Met, и используется трёхбуквенный код SEM, но тут он обозначен как MSE. Интересно, что селенометионинов в структуре довольно много (около 20 штук). Можно предположить, что это не свойственно исходному белку, а специально для кристаллизации был наэкспрессирован белок с заменой всех метионинов на селенометионины. Так делают, когда для решения фазовой проблемы собираются использовать SAD (селен — более тяжёлый атом, чем сера). Эта догадка подтверждается тем, что обычных метионинов в модели вообще нет.

В белке этот селенометионин расположен как раз в области тетрамерного перехода, которая интересовала авторов структуры. Это предпоследняя аминокислота в цепи B, она находится на конце длинной альфа-спирали, но сама уже во вторичную структуру не входит. На рисунке 5 показана карта Рамачандрана, положение 450 Mse обведено красным. Видно, что соотношение углов в остове остатке близко к недопустимому.

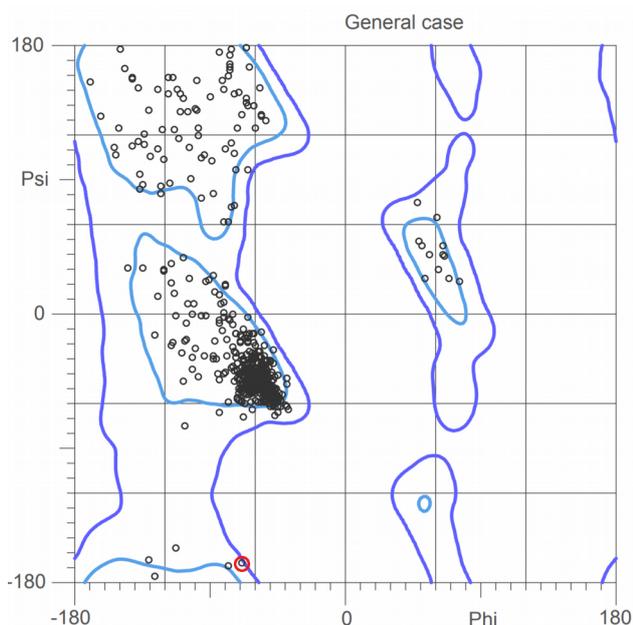
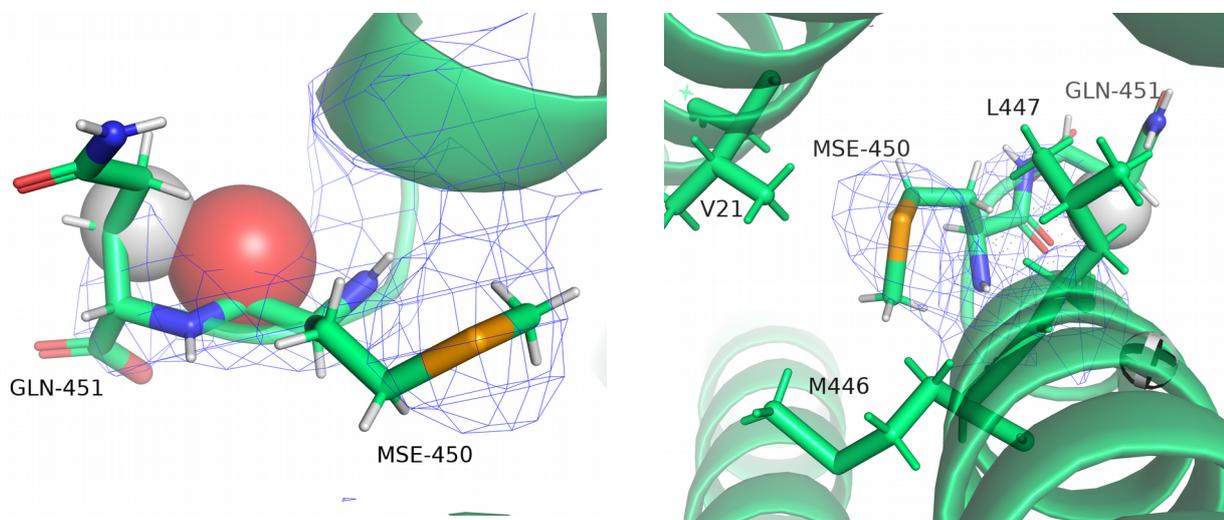


Рисунок 5. Карта Рамачандрана для всех остатков структуры, положение 450 Mse помечено красным.

На рисунке 6 показаны остатки 450 Mse и 451 Gln, а также их электронная плотность (уровень подрезки 1.7). Видно, что электронная плотность хорошо окружает атом селена (логично, он ведь довольно тяжёлый), но сползает с атомов углерода селенометионина и с

кислотной группы глутамина. Видимо, радикал глутамина (конечного в цепи) достаточно подвижный, это подтверждается и высоким значением В-фактора.

Остовный атом кислорода селенометионина и водород на С-beta глутамина показаны сферами. Видно, что их радиусы достаточно сильно перекрываются (хотя между ними, конечно, нет связи). Вероятно, в действительности этого перекрывания нет, а перекрывание получается из-за неточного определения положения атома углерода, плохо покрытого электронной плотностью тем более, что водород вообще добавлен автоматически).



Рисунки 6 и 7. Слева - остатки 450 Mse и 451 Gln и их электронная плотность. Справа - гидрофобное окружение остатка 450 Mse.

Можно предположить, что селенометионин (также, как и метионин в природном белке на его месте) участвует в гидрофобных взаимодействиях, ведь атом селена достаточно крупный, а его электроотрицательность близка по значениям к электроотрицательности углерода. Вероятно, в гидрофобном взаимодействии с ним участвуют остатки валин-21, метионин-446, лейцин-447. Нет ощущения, что для участия в таких взаимодействиях селенометионину нужно как-то изогнуться относительно своей стандартной ротамерной формы.

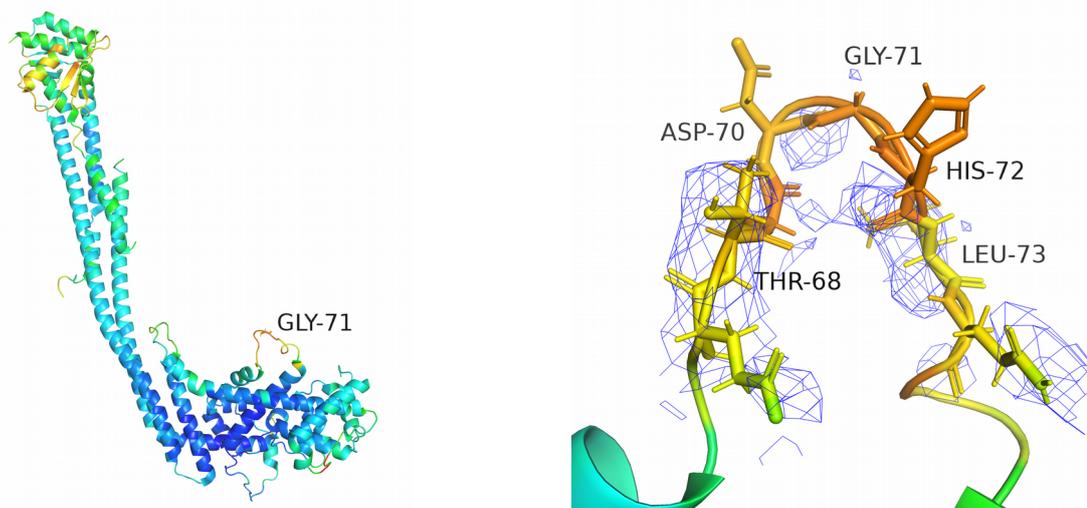
В целом, можно сказать, что расшифровка структуры в этом месте не очень хорошая, но это связано в основном с подвижностью этого участка и наличием нестандартной аминокислоты: вероятно, эта а.о. 450 отличается по углам (в остове и в ротамере) именно из-за того, что это аминокислота с другими свойствами относительно тех, по которым «правильные» параметры были составлены.

Маргинал 2. chain B, 71 Gly

Главная причина подозрительности этого остатка — очень высокий RSRZ = 8.7. Кроме того, у него не очень хорошее положение на карте Рамачандрана и он распознаётся как аутлаер CaBLAM.

Оказалось, что с этим остатком всё довольно просто: он находится в середине неструктурированной петли, за счёт чего сильно подвижен, имеет высокие значения В-фактора (около 200 у разных атомов) и уже на уровне подрезки 0.6 (см. рисунок 9) даже атом кислорода остова не покрыт электронной плотностью. Подвижность этого участка объясняет плохое значение RSRZ. Однако, (мне) не совсем понятно, почему углы остова (положение на

карте Рамачандрана) не в предпочтительной области. Обычно так бывает, когда аминокислотный остаток участвует в каких-то взаимодействиях, и оказывается более выгодным для него отклониться от наиболее удобных углов, но всё же образовать связь. Но тут глицин не участвует во взаимодействиях с другими остатками: радикала нет, кислород остова смотрит в сторону гидрофобного радикала лейцина-73, азот тоже может взаимодействовать только если с растворителем. Вероятно, в действительности глицин имеет углы из «правильной» области, и проблема в неправильной интерпретации этого участка из-за подвижности.



Рисунки 8 и 9. Слева — положение глицина-71 в структуре (окраска по В-фактору), справа — электронная плотность петли, которой принадлежит глицин (уровень подрезки 0.6).

Маргиналы 3+4: chain A, 147-148.

Рассмотрим вместе остатки 147-His и 148-Pro. По данным сервиса, они имеют пересечение Ван-дер-Ваальсовых радиусов в 0.5 Å. Кроме того, гистидин имеет очень высокий RSRZ, а пролин — twisted.

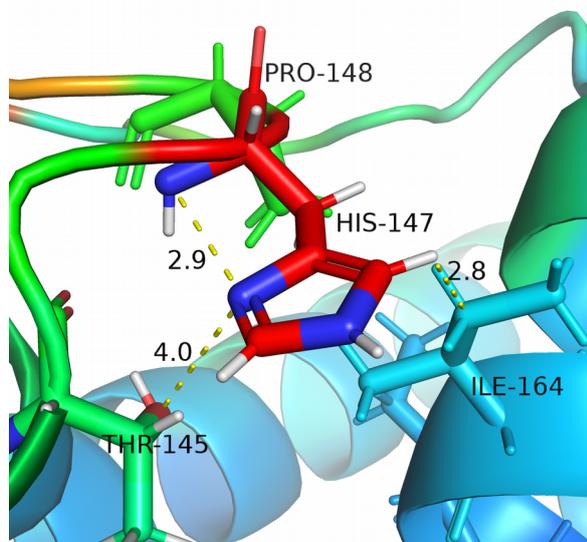


Рисунок 10. Окружение гистидина-147 и расстояния до соседних остатков.

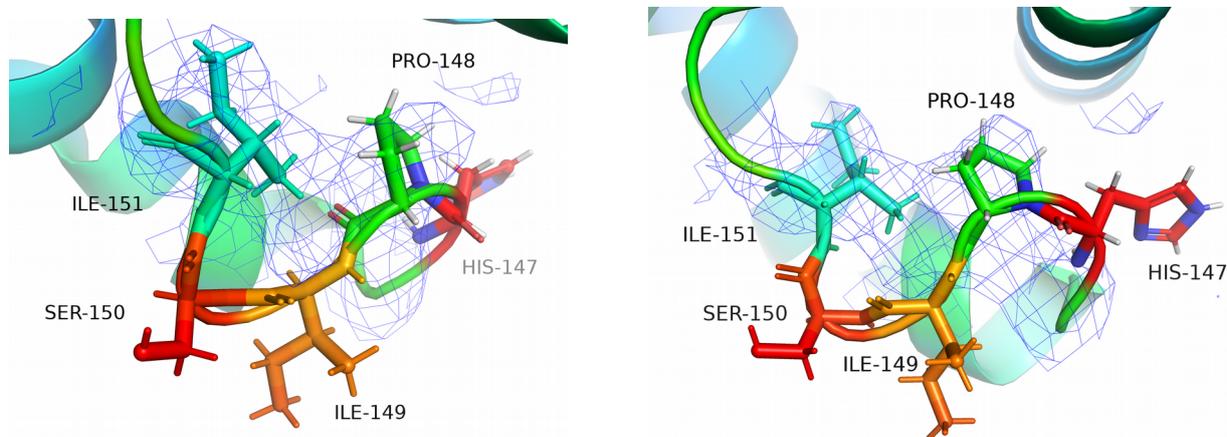
С гистидином примерно понятно: он, вероятно, имеет очень подвижный радикал, поэтому его В-фактор и RSRZ очень высокие. Можно также сказать, что он имеет

непредпочтительное положение на карте Рамачандрана, т. к. этому остатку не очень комфортно в своём окружении: нет других остатков, с которыми можно было бы образовать стекинг или водородную связь: рядом с радикалом гидрофобные изолейцин-164, а серин-145 и лизин-167 на расстоянии более 4х ангстрем. Поэтому гистидин изгибается, так чтобы образовалась водородная связь между азотом кольца и собственным остовным кислородом (длина 2.9, но угол не самый хороший) (см. рисунок 10).

Вероятно, перекрытие радиусов между CD-углеродом пролина и HB2-водородом пролина также можно объяснить плохим качеством электронной плотности гистидина, за счёт чего он вписан не совсем так, как он расположен в кристалле.

Пролин обозначен как *twisted*, значение угла $\omega = 147.2$. Насколько я понимаю, ω – это угол, определяющий транс- или цис- положение пептидной связи, и он должен быть близок к нулю или к 180 градусам. $\omega = 0$ соответствует цис-положению, что типично для пролина и нехарактерно для всех остальных аминокислот (это связано со стерическим отталкиванием), $\omega = 180$ соответствует транс-положению, что характерно для всех аминокислот, кроме пролина. У этого пролина $\omega = 147$, что странно по двум причинам: во-первых, почему оно не близко к 0 и 180 (т. е. пептидная связь не плоская), и, во-вторых, почему пролин в транс-положении?

Рассмотрим электронную плотность (см. рисунки 11-12), в которую этот пролин вписан. Интересно, что он находится в петле, и соседние остатки имеют совсем плохой В-фактор (около 200), а этот остаток имеет В-фактор тоже большой, но почти в два раза меньше (около 110).



Рисунки 11 и 12. Электронная плотность вокруг пролина-148 и соседних остатков. Уровень подрезки 1.5.

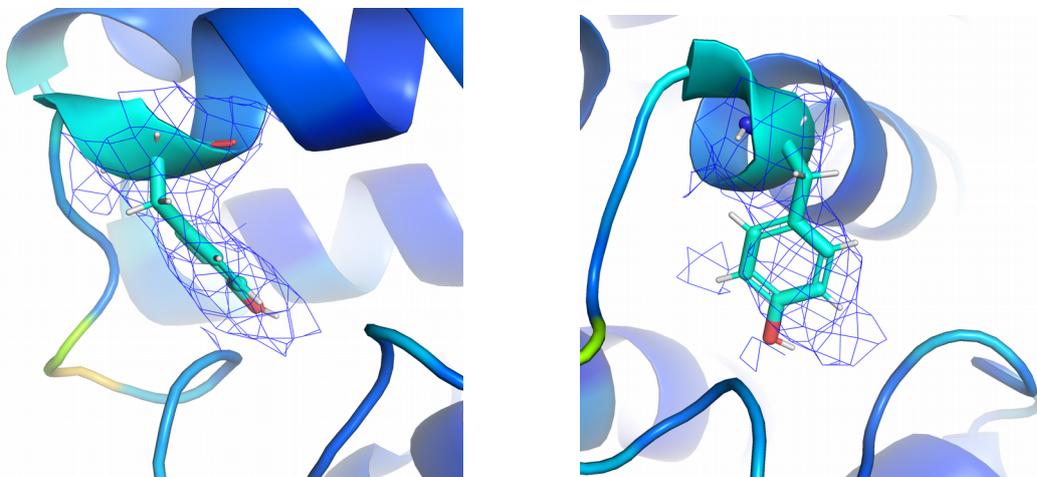
На рисунке 11 видно, что плотность вокруг пролина выглядит так, как будто пролин — это ароматическое кольцо: плотность вылезает вверх и вниз перпендикулярно плоскости кольца, при этом сами атомы углерода в кольце хуже покрываются плотностью, чем область над кольцом. Кажется, что если считать, что эта область соответствует пролину, пролиновое кольцо должно быть повёрнуто (если смотреть на рисунок, правую сторону кольца потянуть на себя, левую отдалить), но в таком случае угол ω станет ещё хуже.

Мне кажется, что показанная область электронной плотности в целом не соответствует вписанным в неё остаткам (147-151), и дело не в качестве ЭП (которая здесь не сильно хуже, чем в других участках), а в том, что пять остатков не влезают в эту электронную плотность, поэтому в модели два остатка (149 и 150) выпучиваются из плотности. Если посмотреть на этот участок электронной плотности (рисунок 12) и не обращать внимание на вписанные остатки, может показаться, что это волна из электронной плотности соответствует остову трёх идущих подряд остатков (зизагу из углеродов, а плотность на радикале 151, на остове

149 и радикале 148 — это смотрящие в разные стороны из цепи остова кислороды и азоты). Но тогда эта область соответствует меньшему числу остатков, и куда вписывать остальные — непонятно.

Маргинал 5. chain B, Tyr 40.

Этот остаток — не-ротамер и имеет клэш, хотя остальные параметры (в т.ч. углы в остове и В-фактор достаточно хорошие). Он находится на конце одной из альфа-спиралей.



Рисунки 13 и 14. Электронная плотность вокруг тирозина-40. Уровень подрезки 1.5.

На рисунках 13 и 14 показано, как радикал этого тирозина вписывается в отведённую ему электронную плотность. Видно (рисунок 13, вид «с торца»), что этот участок ЭП точно соответствует ароматическому кольцу, но если смотреть на кольцо сверху, кажется, что кольцо тирозина как будто сдвинуто относительно кольца электронной плотности, кажется, что его можно подвинуть (по рисунку 14 - вправо) так, чтобы кислород (как многоэлектронный атом) был покрыт «выростом» плотности.

С помощью wizard > mutagenesis я попробовала заменить этот тирозин на один из 4х ротамеров тирозина, но они вписываются в электронную плотность значительно хуже, чем вариант, предложенный авторами.

Мне кажется, что в кристалле этот тирозин немного повёрнут (относительно положения, указанного авторами) и сильно повёрнут относительно ротамера, для того, чтобы образовать водородную связь между кислородом тирозина и остовным кислородом лейцина-27. На рисунке 15 видно, что положение, предложенное авторами, не выгодно, т. к. и не соответствует ротамеру, и находится слишком далеко от соседних остатков, чтобы стабилизироваться за счёт каких-нибудь взаимодействий. Повёрнутое же положение (ближе к ЭП) позволит образовать водородную связь (расстояние 2.7 померено от водорода, на месте которого при повороте окажется кислород тирозина).

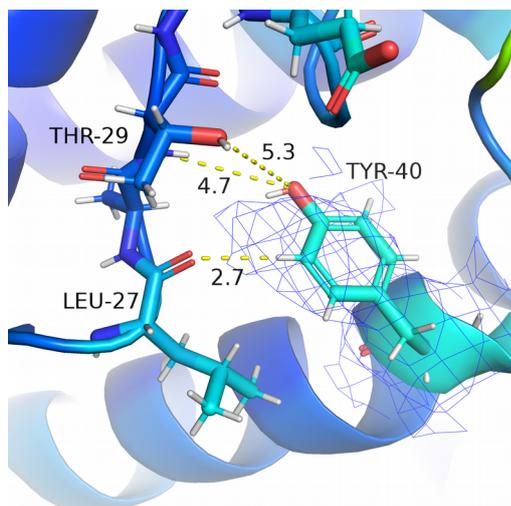


Рисунок 15. Тирозин-40, его электронная плотность (уровень подрезки 1.5) и окружение.

5. Выводы.

В целом, можно сделать вывод, что структура содержит довольно много маргинальных остатков, но большинство из них не связаны ни с особой биологической ролью этих остатков, ни с ошибками авторов при вписывании модели в электронную плотность. В частности, в модели очень много (10%) остатков с RSRZ > 2. В основном это не плохо вписанные остатки, а остатки, имеющие плохое качество электронной плотности из эксперимента, участки, куда остаток вписан, хотя соответствующего куска электронной плотности нет. Там, где качество электронной плотности хорошее, маргиналов (по сравнению с другими структурами похожего разрешения) немного: клэшей мало, запрещённых по Рамачандрану углов нет, кандидатов на флипы (по версии MolProbity) нет.

Ссылки на статьи и использованные сервисы.

[1] Conserved Tetramer Junction in the Kinetochore Ndc80 Complex. Valverde, R.; Harrison, S.C.
DOI: 10.1016 / j.celrep.2016.10.065

[2] <http://molprobity.biochem.duke.edu/>

Christopher J. Williams, Jeffrey J. Headd, Nigel W. Moriarty, Michael G. Prisant, Lizbeth L. Videau, Lindsay N. Deis, Vishal Verma, Daniel A. Keedy, Bradley J. Hintze, Vincent B. Chen, Swati Jain, Steven M. Lewis, Bryan W. Arendall 3rd, Jack Snoeyink, Paul D. Adams, Simon C. Lovell, Jane S. Richardson, and David C. Richardson (2018) MolProbity: More and better reference data for improved all-atom structure validation. Protein Science 27: 293-315.

[3] https://files.rcsb.org/pub/pdb/validation_reports/tc/5tcs/5tcs_full_validation.pdf

[4] <https://www.ebi.ac.uk/pdbe/entry/pdb/5tcs>