

Отчет о качестве расшифровки структуры белка 1VNE методом рентгеноструктурного анализа

Фоменко Елена Алексеевна, студентка 4 курса ФББ МГУ им. М.В. Ломоносова

Москва, 2014

Аннотация

В данном отчете подробно исследуется качество белковой структуры 1VNE, рассматриваются различные ее параметры и индикаторы качества, а также проводится сравнение с оптимизированной моделью из PDB_redo.

Введение

YSDC_BACSU – белок семейства M42 (*glutamyl aminopeptidase family, MEROPS*; EC=3.4.11), предполагаемая аминопептидаза бактерии *B. subtilis*, длиной 373 а.к. Ген *YsdC* консервативен для некоторых архей, термофилов и патогенных видов бактерий.

Структура белка имеет PDB-код 1VNE. Она была расшифрована в рамках проекта Structural GenomiX (SGX) и опубликована в PDB в 2003 году. Проект шел с 2000 по 2002 год; его задачей была расшифровка структур бактериальных белков, для которых в то время не было гомологов по последовательности среди структур PDB. При этом среди них не должно было быть гомологичных по последовательности друг другу. Эти белки высоко консервативны для многих бактерий, а для некоторых предполагалась важная роль в бактериальном росте. Экспрессию белков проводили в *E.coli*. Работа описана в статье Badger J и коллег [1].

Результаты и обсуждение

Для структуры измерено 38362 рефлекса, и все они были использованы. Из них 97,9% (37558) превышают стандартное отклонение более, чем в 3 раза (*вычисления в Excel*). Максимальное разрешение 1.9 Å, минимальное – 45.72 Å. Уточнение модели проводилось программой REFMAC 4.

Фазовая проблема для 1VHE решалась популярным в настоящее время методом многоволнового аномального рассеяния, с помощью модификации метионинов в селен-метионины (MSE). Метод молекулярного замещения не использовался еще и по причине отсутствия гомологов. Структура 1VHE и есть результат модификаций. Положения селена определены с помощью программы SnB [1, 3].

В ассиметрической единице – одна молекула, но белок формирует димер с симметричной молекулой, своим меньшим субдоменом [1]. В качестве биологической единицы авторами принят мономер. Однако PISA и PQS предсказали существование белка в форме тетраэдра из 12 мономеров. Такая структура была показана и для некоторых других белков семейства M42 [4]. Некристаллографических симметрий в ассиметрической ячейке нет.

В активном сайте, определенном остатками H68, D182, E214, E215, D237, и H325, обнаружен связанный катион металла, но для белковой активности необходимо, скорее всего, два катиона. Предполагают, что обнаруженный катион – это Zn^{2+} , но у экспериментаторов есть сомнения, по причине неоднозначности результатов [1]. Помимо белка и катиона металла, также определены 362 молекулы воды (атомы кислорода).

Анализ качества структуры

По данным PDB, значение R -фактора составляет 0.176; R_{free} – 0.193. $R_{free}-R=0.017$, что свидетельствует о хорошем качестве модели; подозрений на переоценку нет. 8 остатков определить не удалось, у 5 остатков есть отсутствующие атомы.

Среднее значение B -фактора (теплового фактора) = 10.95 \AA^2 .

Для последующего определения индикаторов качества структуры использовались протокол проверки качества WHAT_CHECK [5], сервис MolProbity [2], сервер EDS [6].

По картам Рамачандрана, полученным в Molprobity (**рис. 1**) видно, что все остатки лежат в разрешенных областях, при этом 97% остатков лежит в предпочитаемых областях (идеальный показатель – 98%, но отклонение незначительно); по данным картам маргинальные остатки не обнаружены.

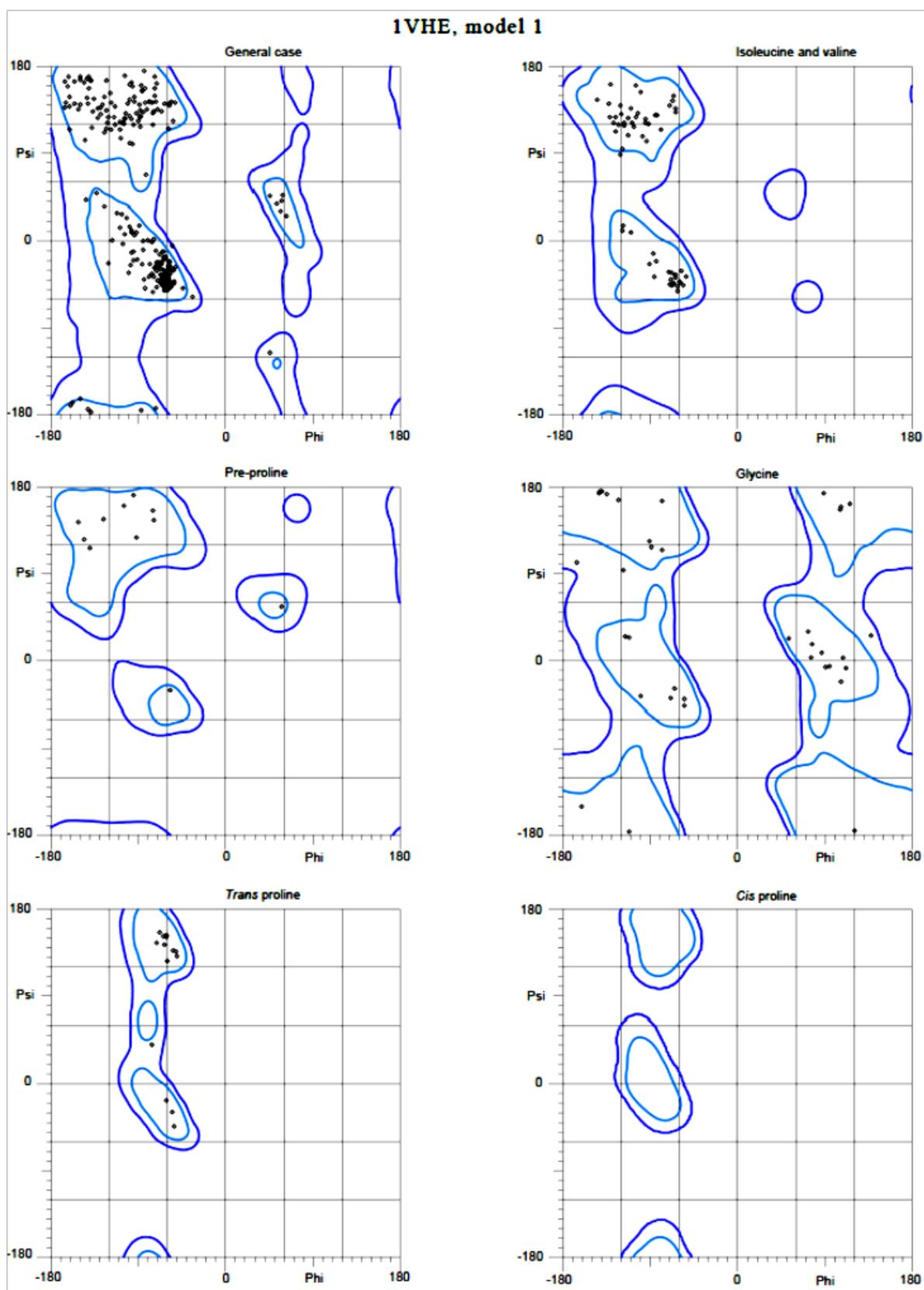


Рис. 1 Карты Рамачандрана, полученные в Molprobity, для остатков 1VHE. Разрешенные области ограничены синими линиями, предпочитаемые – голубыми линиями.

Molprobity score (общая оценка) составил 1.39 (98-ая перцентиль), что указывает на очень хорошее качество. Clashscore (число недопустимых наложений атомов на 1000) также хороший – 4.31 (98-ая перцентиль).

RMS Z-score для длин связей слишком низкий: 0.267 (нормальное значение для этого показателя равно 1, а RMS Z-score < 0.667 может говорить о слишком строгих ограничениях при оптимизации модели). На слишком строгие ограничения указывает и низкое стандартное отклонение для значений углов ω : 1.524 (как правило, этот показатель равен около 5.5).

RMS Z-score для углов N-C α -C равен 1.534 (в идеале – 1).

Найдено три молекулы воды (атома кислорода), не образующих водородные связи: 554, 600, 640. А также обнаружено 11 остатков, не выходящих на поверхность белка, некоторые атомы которых должны были бы участвовать в водородных связях, но, по данным структуры, не участвуют.

Для Gln213 подозревается необходимость инверсии боковой цепи.

По данным статьи [1], между Asp182 и Asn183 обнаружена цис-пептидная связь, которая может быть важна для активного сайта. В WHAT_CHECK для Asp182 указан нетипичный угол N-C α -C (а также еще для семи остатков).

SER31 находится в форме нетипичного ротamera (на основе статистики по структурам PDB).

Более чем для 20 остатков отмечены слишком короткие межатомные расстояния. Но 0.29 Å – наиболее высокая разница из полученных – не слишком большая ошибка.

Среднее значение RSR-фактора, по данным EDS, равно 0.094 со стандартным отклонением 0.044, что говорит о хорошем соответствии координат атомов экспериментальной электронной плотности.

Обнаружено 10 остатков с высоким RSR-фактором. Наибольшие значения у Pro126 (0.322), His367 (0.318), Ser125 (0.311). График RSR для всех остатков структуры изображен на **рис. 2**.

Некоторые потенциально маргинальные остатки приведены в **таблице 1**.

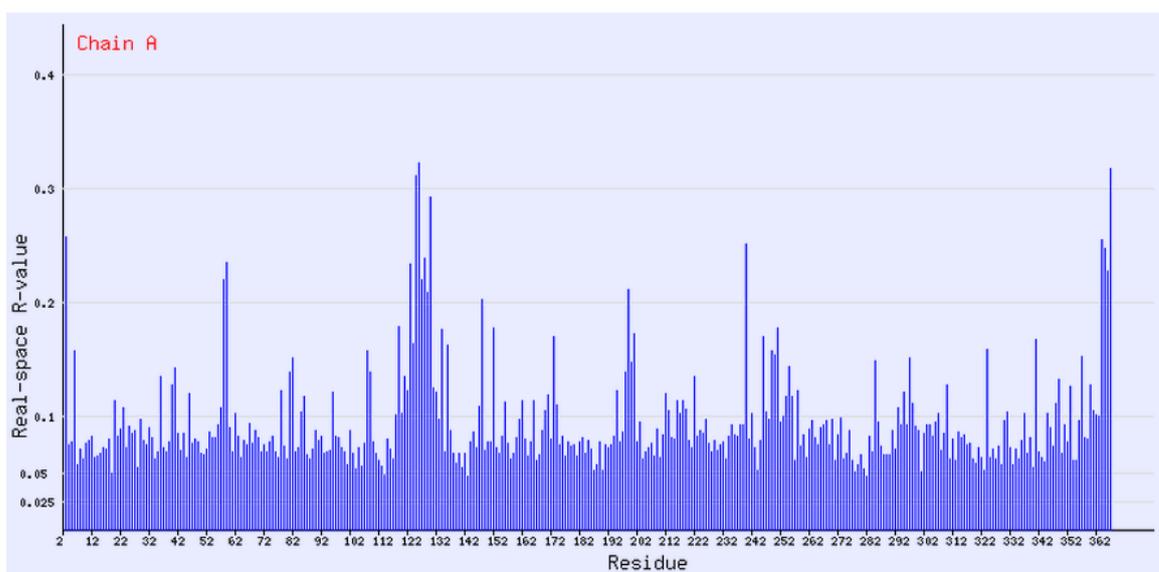


Рис. 2 Зависимость RSR от номера остатка структуры, полученная в EDS.

Таблица 1

Некоторые потенциально маргинальные остатки структуры 1VNE

Маргинальные остатки/гетеромолекулы	Показатель
O554, O600, O640	Не образуют водородных связей
Asp182 и Asn183	Цис-конформация связи
Gly67, Asn172, Asp182	Нетипичный угол N-C α -C (отклонение RMS на 4.75 σ , 4.82 σ , 5.84 σ , соответственно)
Gln213	Инверсия боковой цепи
Ser31	Нетипичный ротамер
Leu4	N не участвует в водородной связи
Lys341	Слишком короткое расстояние от HOH684: NZ \leftrightarrow O=2,41 Å
Pro126, His367, Ser125	Высокие RSR: 0.322, 0.318, 0.311, соответственно

Далее рассматриваются 5 наиболее интересных, на мой взгляд, случаев из **таблицы 1**.

Цис-конформация связи Asp182 - Asn183

В статье авторов структуры [1] подчеркивается данная особенность. Остаток 182 входит в активный сайт белка, и цис-конформация связи Asp182 - Asn183 только подчеркивает его роль (**рис.3**).

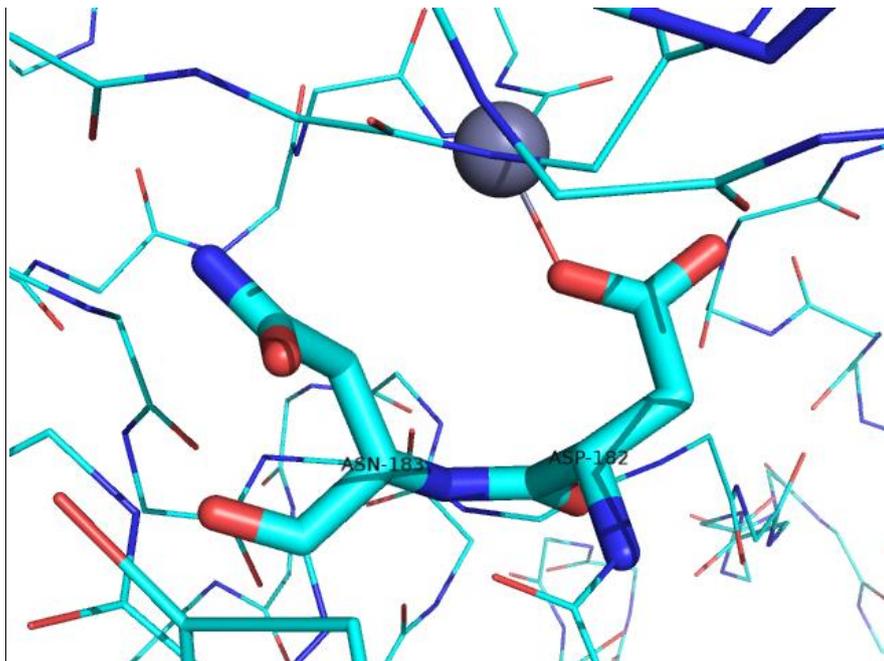


Рис. 3 Цис-пептидная связь между Asp182 и Asn183

Чтобы убедиться, что данное расположение атомов оправданно, посмотрим на распределение электронной плотности на этом участке (**рис. 4**).

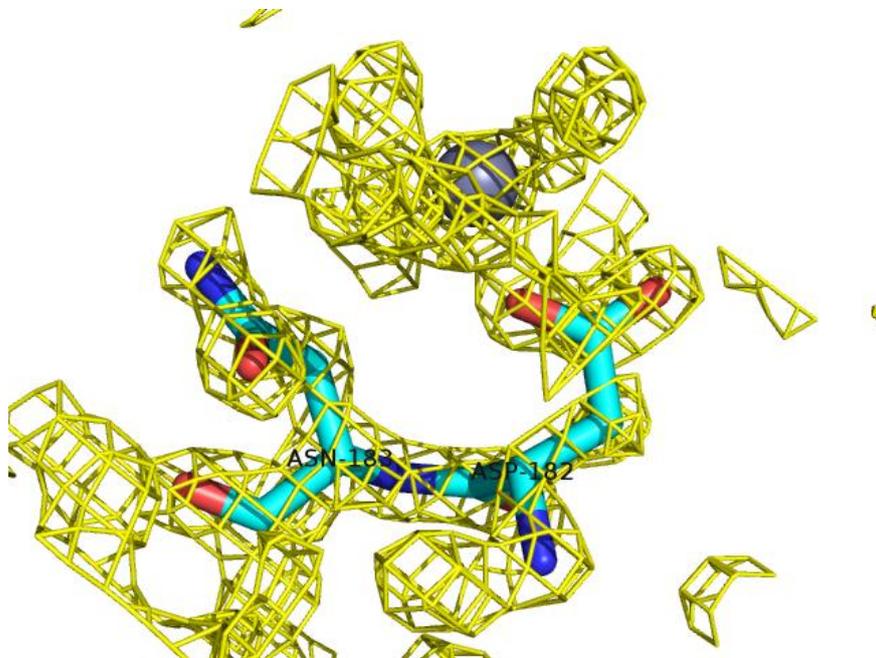


Рис. 4 Распределение электронной плотности вокруг Asp182 и Asn183. Атом цинка показан в виде серого шара. Уровень подрезки 2.5.

На **рис. 4** видим, что вписано все совершенно верно. Значит, ошибки здесь нет: такова особенность структуры.

Нетипичный угол N-C α -C в Asp182

Asp182 входит в активный сайт белка. Возможно, нетипичный угол тоже можно объяснить этим фактом? Соответствующий RMS отклоняется на 5.84 σ .

Нужно учесть, что RSR для этого остатка очень низкий: 0,078. На **рис.5** изображено распределение электронной плотности вокруг остатка и показан «нетипичный» угол N-C α -C. Видно, что остаток вписан совершенно верно, и небольшое отклонение угла от нормального значения, скорее всего, – не ошибка.

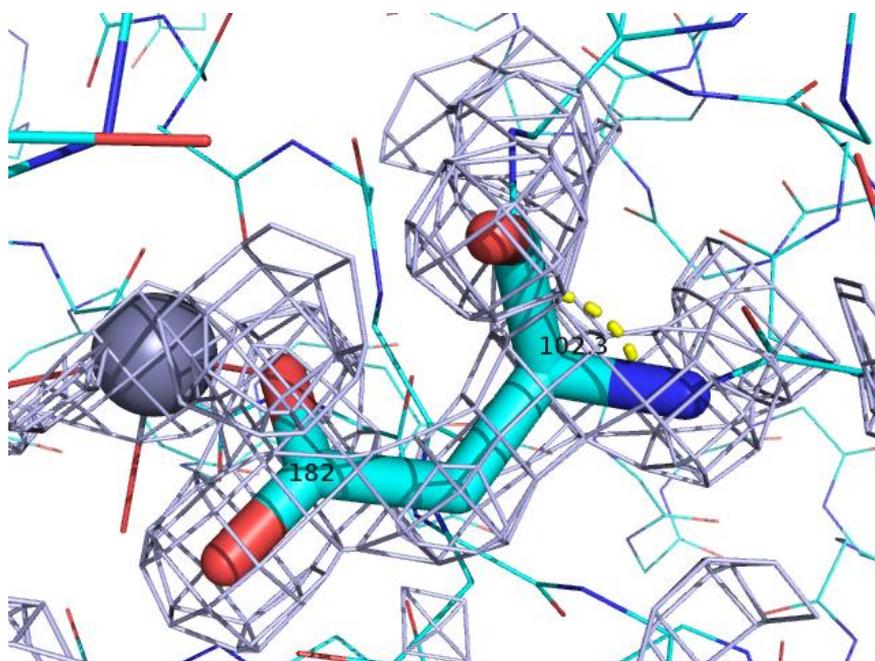


Рис.5 Распределение электронной плотности вокруг Asp182; уровень подрезки 2. Показан угол N-C α -C.

Pro126 с высоким RSR

RSR остатка пролина-126 довольно высокий: 0.322. Такое значение может говорить о том, что остаток, возможно, плохо вписан в электронную плотность. На самом деле, Pro126 более-менее вписывается в электронную плотность только при уровне подрезки 0.5, при этом однозначность вписывания не очевидна (**рис. 7**).

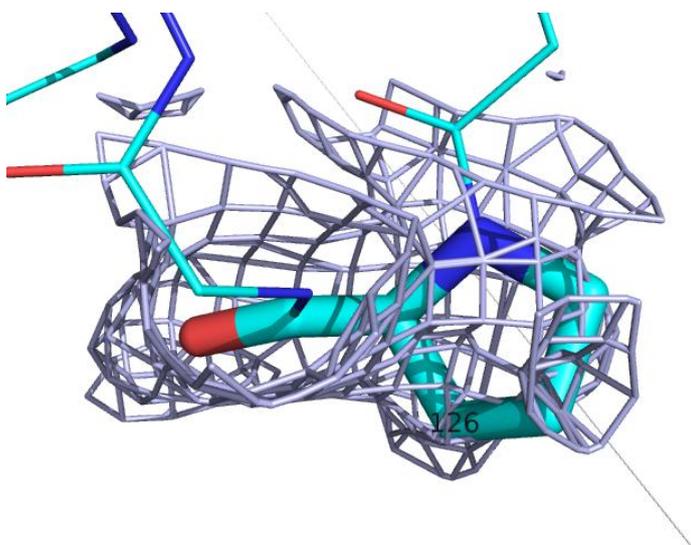


Рис. 7 Распределение электронной плотности вокруг Pro126. Уровень подрезки 0.5.

Плохое соответствие связано, скорее, не с ошибками авторов, а с плохими экспериментальными данными на этом участке. Для этого остатка больше не обнаружено каких-либо странностей в других показателях. Несмотря на плохой RSR, ошибки здесь, кажется, нет.

Маленькое расстояние от атома NZ остатка Lys341 до HON684

Расстояние от азота боковой цепи лизина-341 до кислорода воды-684 равно всего 2.41 Å (**рис. 8**).

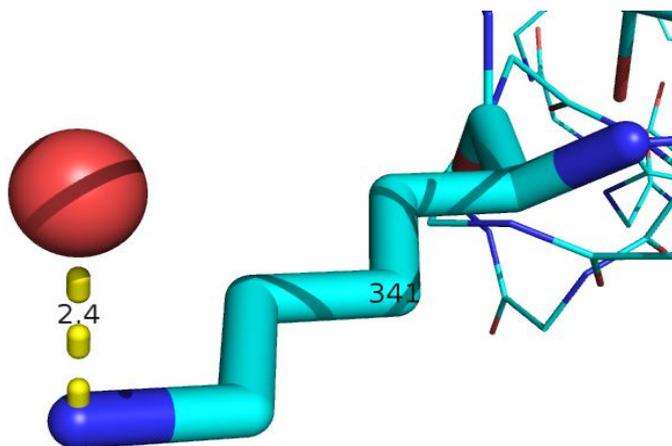


Рис. 8 Расстояние от атома NZ остатка Lys341 до O воды-684.

Для Lys341 RSR равен 0.167 – не самое маленькое значение, но и не слишком высокое. Стоит взглянуть на электронную плотность (**рис. 9**).

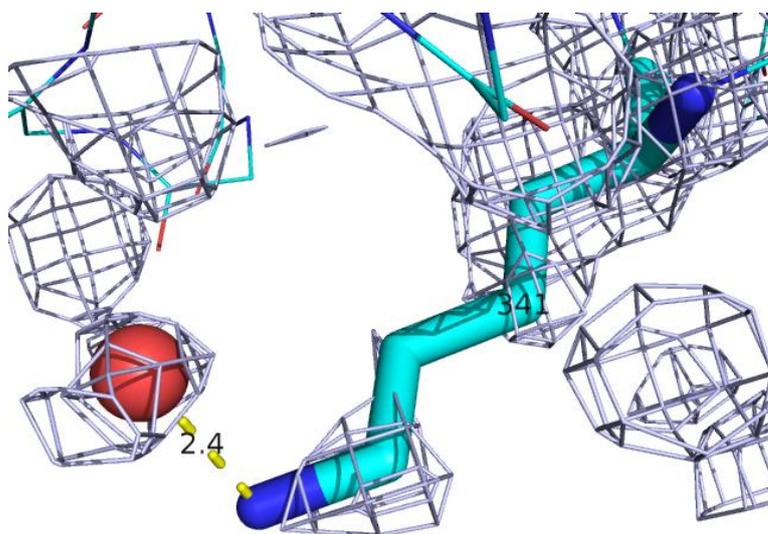


Рис. 9 Распределение электронной плотности вокруг Lys341. Уровень подрезки 1.

Видим, что ошибки в определении координат кислорода воды нет, они однозначно определяются. Проблема может быть в расположении боковой цепи лизина. Возможно, небольшая неточность в определении координат для этого остатка, в связи с недостаточным качеством экспериментальных данных на этом участке, привела к маленькому расстоянию между его NZ и кислородом воды. Других аномалий для этого остатка не обнаружено.

Инверсия боковой цепи в Gln213

Информация о необходимости инверсии для Gln213 найдена в протоколе WHAT_CHECK. Необходимость связана с тем, что текущее положение атомов кажется энергетически невыгодным. Данный остаток в исходной и исправленной с помощью Molprobity структурах представлен на **рис. 10**.

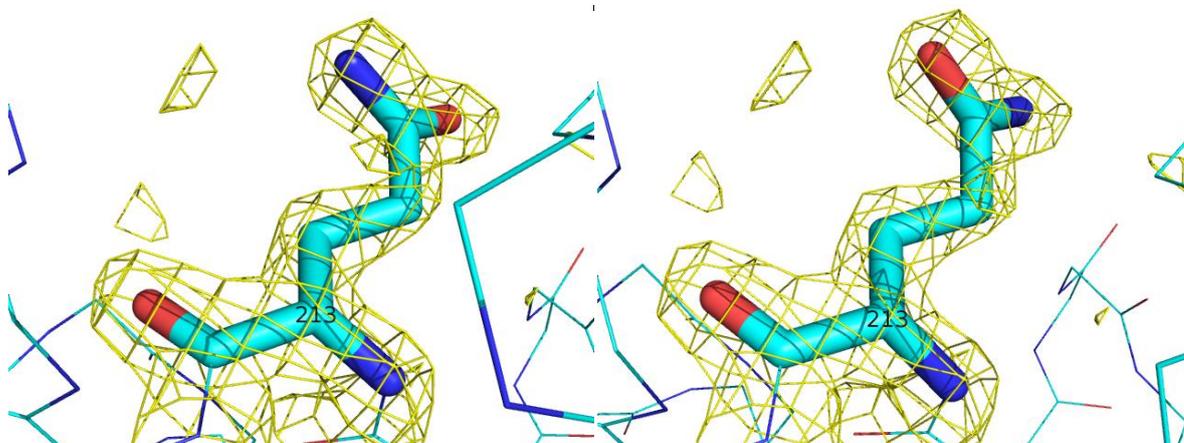


Рис. 10 Остаток Gln213 в исходной (слева) и исправленной (справа) структурах с распределением электронной плотности. Уровень подрезки 2.

Инверсия связана, по-видимому, с ошибкой авторов при расшифровке.

Сравнение структуры с моделью из PDB_redo

На **рис. 11** представлено совмещение структуры 1VHE и модели из PDB_redo. Визуально не обнаружено участков, сильно изменивших свое положение после оптимизации, что указывает на хорошее качество исходной структуры.

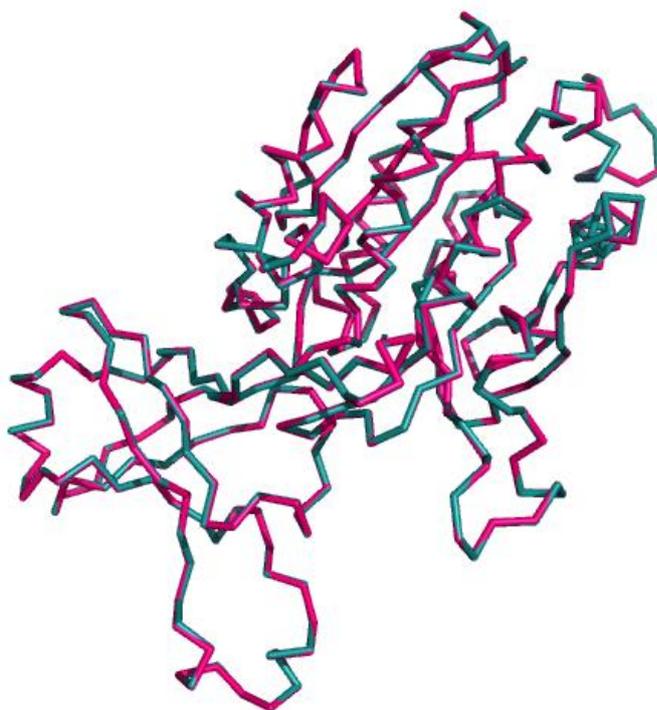


Рис. 11 Совмещение исходной структуры (окрашена розовым цветом) и модели из PDB_redo (окрашена зеленым цветом) в виде *ribbons*.

Проведем сравнение некоторых параметров исходной и оптимизированной моделей.

Molprobitu score (общая оценка) улучшился с 1.39 до 0.81 (100-ая перцентиль), clashscore с 4.31 улучшился до 1.07 (100-ая перцентиль). Ниже приведена таблица сравнения R-факторов.

Сравнение R-факторов исходной и оптимизированной моделей

Показатель	Исходная структура	Модель PDB_redo
<i>R</i>	0.176	0.13
<i>R-free</i>	0.193	0.155
<i>R-free - R</i>	0.017	0.025

Разность *R-free - R* несколько увеличилась, что говорит не в пользу оптимизированной модели. Также не решилась проблема с нетипичным ротамером: исходный Ser31 поправлен, но появился новый – Pe85. Инверсия боковой цепи, однако, исправлена.

Заключение

1VNE – структура высокого качества с хорошими показателями, без грубых ошибок. Тем не менее, есть недостающие остатки и атомы; обнаружен остаток с инверсией боковой цепи; найдено 11 остатков, не выходящих на поверхность белка, не образующих водородные связи; некоторые показатели указывают на слишком строгие ограничения при оптимизации. Сравнение с моделью из PDB_redo показало, что структура может быть оптимизирована; кардинальных изменений при оптимизации не произошло, что указывает на хорошее качество исходной структуры.

Список литературы

- 1) Badger J, Sauder JM, Adams JM, et al. Structural analysis of a set of proteins resulting from a bacterial genomics project. *Proteins* 2005 Sep 1; 60(4): 787-96
- 2) Сервер Molprobity: <http://molprobity.biochem.duke.edu/index.php>
- 3) Weeks CM, Miller R. Optimizing Shake-and-Bake for proteins. *Acta Crystallogr D Biol Crystallogr* 1999; 55: 492–500.
- 4) Franzetti,B., Schoehn,G., Hernandez,J.F., Jaquinod,M., Ruigrok,R.W. & Zaccai,G. Tetrahedral aminopeptidase: a novel large protease complex from archaea. *EMBO J* 2002; 21: 2132-2138
- 5) WHAT_CHECK: <http://swift.cmbi.ru.nl/gv/whatcheck/>
- 6) Electron Density Server: <http://eds.bmc.uu.se/eds/>
- 7) Protein Data Bank: <http://www.rcsb.org/pdb/home/home.do>