

МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
им. М.В. ЛОМОНОСОВА

ФАКУЛЬТЕТ БИОИНЖЕНЕРИИ И БИОИНФОРМАТИКИ

**Отчёт по качеству расшифровки структуры с
PDB-кодом 3D30, полученной методом
рентгеноструктурного анализа**

Работа

студентки 4 курса

Александры Анисимовой

Москва 2015

Оглавление

| | |
|--|----|
| Список сокращений..... | 3 |
| Аннотация | 3 |
| Введение..... | 3 |
| Результаты и обсуждение | 6 |
| Общая информация о модели | 6 |
| Значения индикаторов качества модели в целом | 7 |
| R-фактор | 7 |
| R-free | 8 |
| Пространственный R-фактор..... | 8 |
| Z-score | 9 |
| Температурный фактор | 10 |
| Карты Рамачандрана | 11 |
| Clashscore..... | 13 |
| Остатки с маргинальными по отклонению от ротаметров боковыми цепями..... | 13 |
| Другие параметры | 13 |
| Проверка структуры с помощью пакета WHAT_IF и протокола WHAT_CHECK..... | 13 |
| Аминокислотные остатки с подозрительным окружением..... | 13 |
| Повороты пептидной цепи..... | 14 |
| Альтернативные положения атомов | 14 |
| Ошибки в номенклатуре | 14 |
| Вода без водородных связей..... | 14 |
| Поиск и анализ маргинальных остатков в структуре..... | 14 |
| Met1..... | 15 |
| Tyr73 | 16 |
| НОН 214..... | 18 |
| Lys188 | 19 |
| Asn100..... | 20 |
| Lys196 | 21 |
| Сравнение модели из PDB с моделью из PDB_redo..... | 22 |
| Заключение..... | 24 |
| Ссылки на серверы и базы данных | 25 |
| Список литературы..... | 25 |

Список сокращений

EDS, Electron Density Server

PDB, Protein Data Bank

RSR, пространственный R-фактор (Real Space R-factor)

VAST, Vector Alignment Search Tool

Аннотация

В данной работе представлена информация о структуре с PDB-кодом 3D30. Целью данной работы является поиск и систематизация информации по качеству расшифровки структуры 3D30. Особое внимание уделяется различным индикаторам качества структуры и анализу маргинальных остатков. По результатам работы сделан вывод по качеству расшифровки структуры 3D30.

Введение

Структура PDB-кодом 3D30 получена в 2008 году для бактериального экспансина YoaJ (EXLX1) в рамках работы по изучению его функции группой под руководством F. Kerffa. [1]

Экпансины широко распространены у растений. Они ослабляют клеточные стенки, чтобы растительные клетки могли расти. Это небольшие белки (около 26кДа), состоящие из двух доменов: D1 и D2 [2]. В геноме некоторых групп бактерий были найдены гены, имеющие сходство с генами, кодирующими растительные экспансины. Белки, кодируемые ими, имеют около 15% идентичности с растительными.[3]

Экпансин YoaJ (EXLX1), выбранный для исследования, это секреторный белок Грам-положительной почвенной бактерии сенной палочки (*Bacillus Subtilis*), для которой известно, что она может заселять слой почвы практически неотделимый от корней растений (ризосферу)[4].

EXLX1 имеет массу 23 кДа. Модель 3D30 позволяет заключить, что EXLX1 состоит из двух доменов, характерных для экспансинов. Белок имеет форму эллипсоида длиной в 60 Å и шириной в 37 Å. Поиск близких структур (использовался алгоритм VAST [5]), показал, что наиболее близкой структурой обладают белки EXPB1 из пыльцы кукурузы (PDB-код 2HCZ) и тимофеевки луговой (PDB-код 1N10). Известно, что эти белки приводят к удлинению клеточной стенки растений [6], [7].

EXLX1, как и другие экспансины, состоит из доменов D1 и D2, связанных коротким линкером. В статье «Structure-Function Analysis of the Bacterial Expansin EXLX1» Georgelis et al., которая вышла на три года позже обсуждаемого исследования, показано что связь между этими доменами важна для функции белка, а также, что основной функцией домена D1 является закрепление экспансина на клеточной стенке растения, а домен D2, нужен как для закрепления, так и для активности белка [8]. В рамках этой работы также были получены две структуры экспансина EXLX1 в комплексе с полисахаридами (PDB-коды 4FG4 и 4FER) [2]. Однако их разрешение хуже, чем у структуры 3D30.

Домен D1, образованный остатками 1-108, состоит из трех альфа-спиралей и дубли-пси бета-бочонок, состоящий из шести бета-тяжей. С помощью структурного выравнивания было показано, что этот домен, возможно, участвует в узнавании полисахаридов. Также с помощью выравнивания последовательностей белков с близкими структурами, были определены консервативные остатки. Наиболее консервативными остатками являются Thr12 и Asp82. Они образуют консервативную водородную связь. Все другие консервативные остатки (Gly21, Ala22, Ala36, Gly53, Thr70, Asp71 и Leu83), за исключением Gly53, находятся в области желобка, которая, возможно, выступает в роли сайта связывания полисахарида [1].

Домен D2 образован остатками 113–208. Он образован двумя бета-листами, состоящими из четырех антипараллельных тяжей. Интересно, что три ароматических остатка (Trp125, Trp126, Tyr157) образуют плоскую поверхность на этом домене. Такая платформа характерна для участка связывания углеводов типа А (type A carbohydrate-binding module) [9]. Позже было показано, что три этих остатка необходимы для функции белка [8].

Структура PDB-кодом 3D30 с выделенными доменами, элементами вторичной структуры, консервативными остатками показана на рисунке 1А, а возможный сайт связывания полисахарида показан на рисунке 1В.

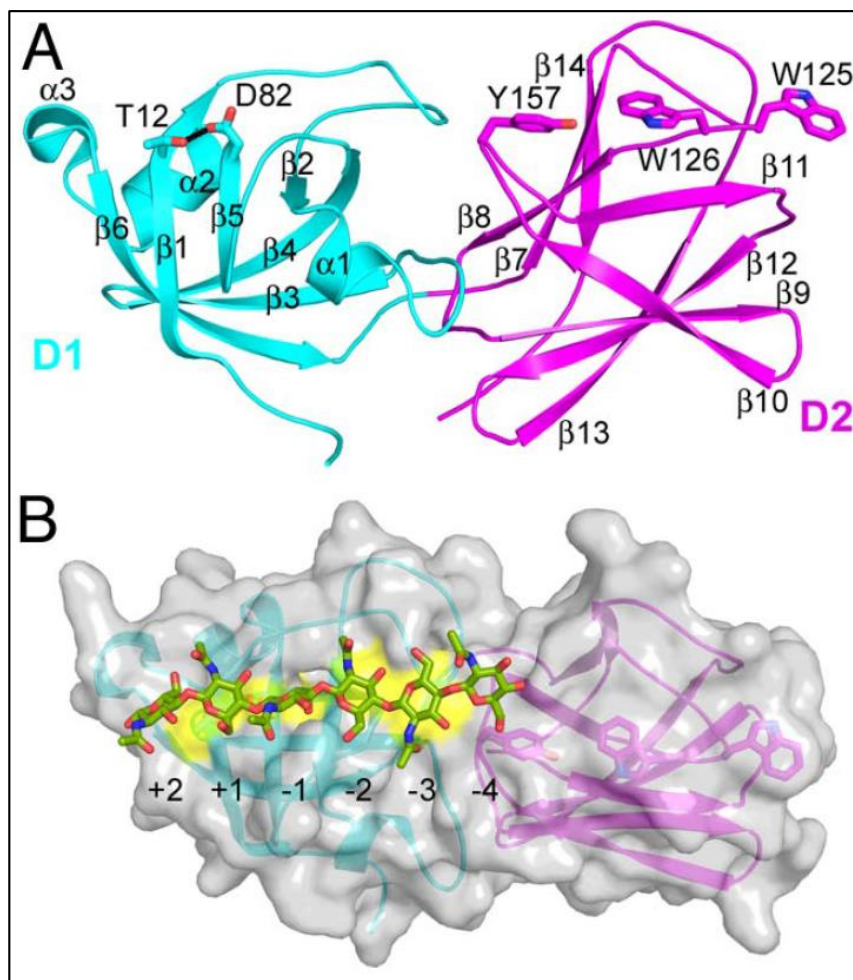


Рис.1. А. Структура PDB-кодом 3D30. Домен D1 выделен голубым, а D2 розовым. Обозначены элементы вторичной структуры. Выделены остатки Thr12 и Asp82, а также Trp125, Trp126, Tyr157. В. Возможный сайт связывания полисахарида (хитогексоза) на структуре 3D30.

Так как EXLX1 имеет структурное сходство с экспансинами растений, решено было проверить похож ли EXLX1 на них функционально. Стандартный тест на активность экспансина, который состоит в определении способности белка вызывать удлинение клеточных стенок растительных клеток [2], показал, что EXLX1 обладает небольшой (примерно в 10 раз менее активен, чем растительные экспансины), но значимой активностью. Также было показано, что он может присоединяться к клеточной стенке растений и к пептидогликану. Однако авторам не удалось продемонстрировать литическую активность EXLX1.

Кроме того, был исследован мутантный штамм *B.subtilis* с нокаутом по гену, кодирующему EXLX1. Он обладает значительно меньшей способностью к колонизации ризосферы растений, чем штамм дикого типа. Таким образом был сделан вывод, что бактериальный экспансин EXLX1 необходим бактериям для колонизации поверхностного слоя корней растений [1].

Результаты и обсуждение

Общая информация о модели

Модель 3D30 была получена в 2008 году в результате обработки данных, полученных методом рентгеноструктурного анализа, для кристалла экспансина YoaJ (EXLX1) из сенной палочки (*Bacillus subtilis*). Кристаллизация и решение структуры были выполнены коллективом бельгийских и французских ученых Frederic Kerffa, Ana Amoroso, Raphael Hermana, Eric Sauvagea, Stephanie Petrellab, Patrice Fileea, Paulette Charliera, Bernard Jorisa, Akira Tabuchic, Nikolas Nikolaidisc, Daniel J. Cosgrovec. [1]

Кристалл был получен методом hanging-drop vapor diffusion [1], при котором препарат белка в висячей капле в разбавленном буфере помещается над концентрированным буферным раствором, содержащим осадители. В процессе установления равновесия между каплей и раствором концентрация белка и осадителя в капле увеличиваются. В результате могут создаться благоприятные для кристаллизации белка условия, и он образует кристалл. [10] В данной работе был использован буфер с pH 4.4, содержащий 4 М формиат натрия и 100 мМ ацетат натрия. Полученные кристаллы помещали в 50% глицерин и замораживали N₂ ж. [1]

Условия получения кристалла объясняют состав лигандов в комплексе. Комплекс образован молекулой экспансина YoaJ (EXLX1), а лигандами являются муравьиная кислота и глицерин.

Кроме кристалла нативного белка, был также получен такой же (по структуре) кристалл его SeMet производного. Он использовался для решения фазовой проблемы методом молекулярного замещения (Molecular replacement). Всего в полученной структуре было 7 атомов селена, так как в последовательности экспансина есть 7 остатков метионина (исключая первый N-концевой метионин), можно сказать, что они все были успешно заменены на селенометионин. Структура производного была решена методом одноволнового аномального рассеяния (single-wavelength anomalous diffraction method). Для неё указано разрешение 2.5 Å. Её можно найти в PDB по коду 2BH0. С использованием этой структуры была решена фазовая проблема и рассчитано положение полипептидной цепи для 174 остатков и боковых цепей для 29 остатков.[1]

Число измеренных рефлексов для структуры составляет 38922 (строка `_reflns.number_obs` в файле структурных факторов), при полноте набора структурных факторов 99.9%. Диапазон разрешений структурных факторов, указанный на сервере EDS, составляет от 1.90 Å до 41.51 Å. Эти рефлексы были разделены на две группы: одна для

уточнения структуры 2ВН0 (раздел в PDB-файле DATA USED IN REFINEMENT) – 36971 рефлексов, а другая для валидации полученной модели (для определения R-free) – 1951 рефлексов.

Разрешение структуры составляет 1.90 Å при полноте 99.7%. Удивительно, но в PDB-файле указан диапазон рефлексов от 1.90 Å до 45.40 Å. Длина волны для рефлекса с самым низким разрешением указана больше, чем та, что указана в файле структурных факторов.

В записи PDB 3D30 указаны следующие кристаллографические характеристики (поле CRYST1):

```
CRYST1 143.780 143.780 143.780 90.00 90.00 90.00 I 2 3 24
```

Первые три числа - это длины направляющих векторов кристалла. Они все одинаковые и составляют 143.780 Å. Следующие три числа - это углы между направляющими векторами кристалла. Все эти углы прямые. Кристаллографическая группа I 2 3. Это кубическая объемноцентрированная ячейка (body-centred). Всего в ней 24 молекулы. В PDB всего 0.41% структур имеет такую элементарную ячейку.

Экспансин YoaJ (EXLX1) состоит из одной полипептидной цепи, некристаллографических симметрий в ячейке нет.

Значения индикаторов качества модели в целом

R-фактор

Стандартный фактор достоверности (R-фактор) – это величина, которая характеризует модель после оптимизации. Это мера различия оптимизированной модели и экспериментальных данных. Для вычисления R-фактора высчитываются значения структурных факторов по модели (F_{hkl}^{calc}) и сравниваются с экспериментальными значениями структурных факторов (F_{hkl}^{obs}) при тех же h, k, l. Формула для R-фактора представлена ниже:

$$R = \frac{\sum_{hkl} |F_{hkl}^{calc} - F_{hkl}^{obs}|}{\sum_{hkl} F_{hkl}^{obs}} * 100\%$$

Стоит отметить, что в процессе оптимизации модели используется не эта формула, а сумма квадратов разностей структурных факторов. Также учитываются длины химических связей и углы между атомами.

Хорошими считаются значения R-фактора меньше 25%. Однако значения меньше 10% говорят о переоптимизации модели.

Для модели 3D30 значение R-фактора, указанное в PDB, составляет 16.5 % (16.8% на сервере EDS). Это значение попадает примерно в середину интервала оптимальных значений.

R-free

Вычисление значения R-free помогает выяснить, была ли модель переоптимизирована. R-free вычисляется по той же формуле, что и R-фактор, только по специально оставленной части рефлексов, по которым не проводилась оптимизация. Обычно их число составляет около 10% от числа всех измеренных рефлексов (1951 из 38922 рефлексов для модели 3D30; это 5%). Для хорошей модели R-free должен быть немного больше, чем R-фактор. Если разность значений R-free и R-фактора превышает 8%, то имеет место переоптимизация модели.

R-free для модели 3D30 составляет 19.1%. Таким образом, разность значений R-free и R-фактора равна 2.6%. Это очень хорошее значение.

Пространственный R-фактор

Пространственный R-фактор (Real Space R-factor, RSR) характеризует соответствие «экспериментальной» электронной плотности и электронной плотности, построенной по модели. Хорошими считаются значения от 10% до 20%. RSR используется не для оценки качества всей структуры, как R-free и R-фактор, а для оценки локального качества. То есть, для отдельных аминокислотных остатков. На сервере EDS можно получить диаграмму со значениями RSR для каждого остатка. Пространственный R-фактор для всех остатков структуры 3D30 показан на рисунке 2.

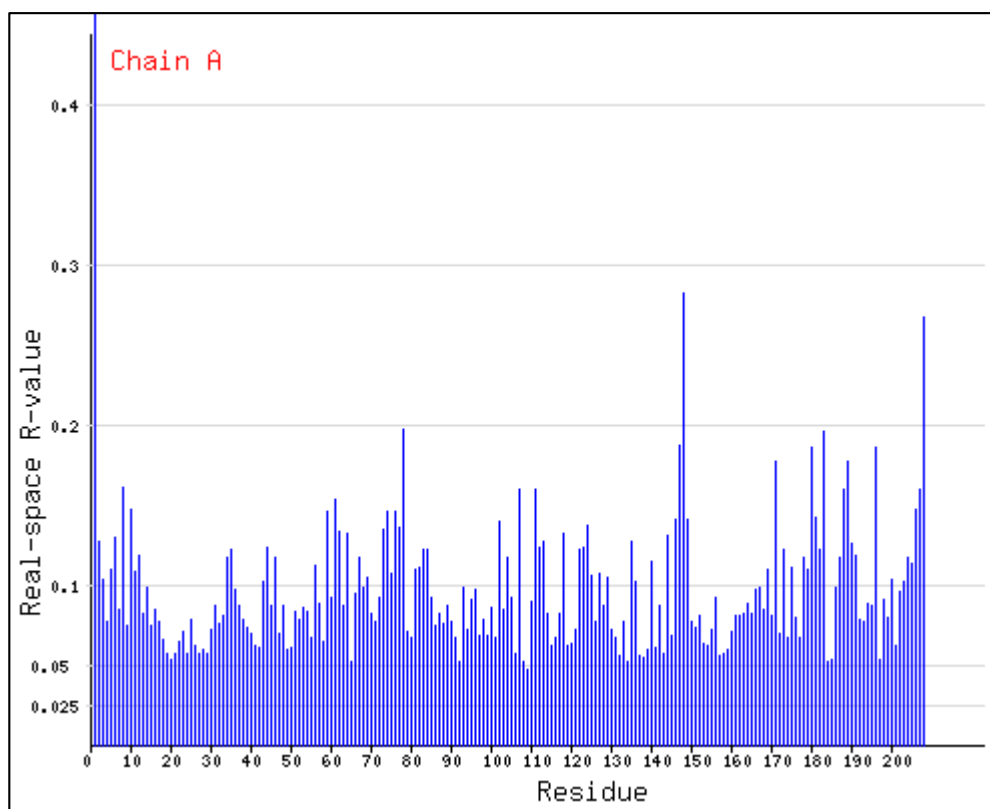


Рис. 2. Пространственный R-фактор для всех аминокислотных остатков структуры 3D30.

На рисунке 2 можно отметить аминокислотные остатки, значения RSR для которых превышают 20%. Это первый и последний остатки (Met1, Glu208), а также Lys148. Еще два остатка имеют значение RSR близкое к 20%. Это Arg78 (RSR=19.8%) и Lys183 (RSR=19.6%). Можно отметить, что эти остатки либо находятся на «краю» полипептидной цепи (первый и последний), либо имеют длинные радикалы.

Среднее значение RSR по всем аминокислотным остаткам для структуры 3D30 указанное на сервере EDS составляет 10% со стандартным отклонением 4.8%. Это хорошие значения.

Z-score

Для более точного определения остатков, плохо вписанных в электронную плотность используется показатель Z-score RSR. Это оценка RSR относительно других структур. Он вычисляется для каждого остатка путём сравнения его RSR со средним RSR для такого же типа остатков из структур с близкими значениями разрешения. Для остатков с плохим RSR, но хорошим Z-score RSR можно заключить, что координаты атомов расшифрованы плохо, но не хуже, чем в других подобных структурах. Если для остатка Z-score больше 2, значит, что он маргинальный.

Z-score для всех остатков структуры 3D30 показан на рисунке 3.

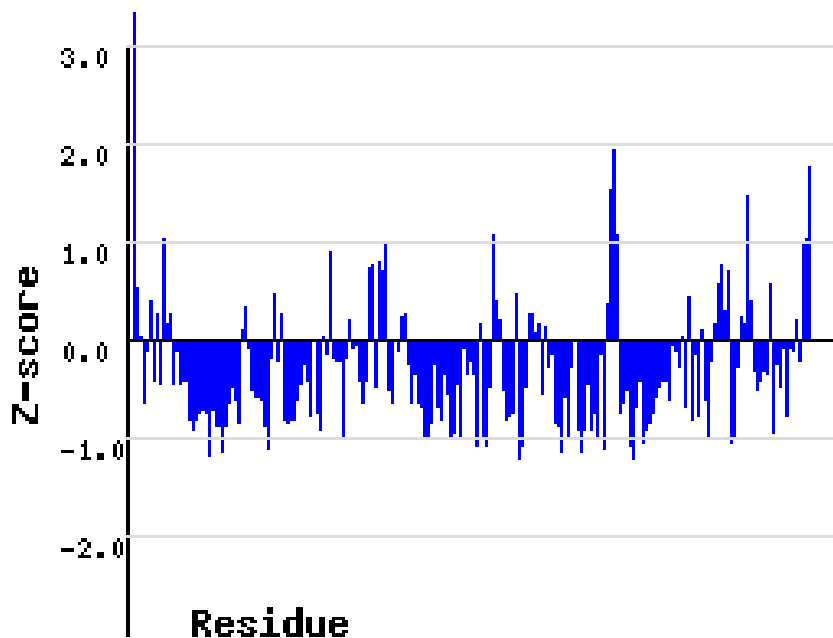


Рис. 3. Z-score для всех аминокислотных остатков структуры 3D30.

Значение Z-score больше 2 только для Met1. Оно составляет 6.9. Следующее плохое значение Z-score составляет 1.96 и соответствует Lys148.

Температурный фактор

Температурный фактор (B-фактор) характеризует меру динамической неупорядоченности кристалла, обусловленную тепловым движением, а также небольшими отклонениями в расположении молекул белка в кристалле. Так как с повышением температуры увеличиваются тепловые колебания атомов, этот фактор увеличивается с увеличением температуры. На рисунке 4 показаны значения B-фактора для усредненные для каждого остатка структуры 3D30. Эти значения достаточно высокие для всех остатков (около 35), однако они распределены достаточно равномерно по всей структуре. Наибольшие значения B-фактора характерны для остатков с длинными боковыми цепями, например, для лизина и аргинина.

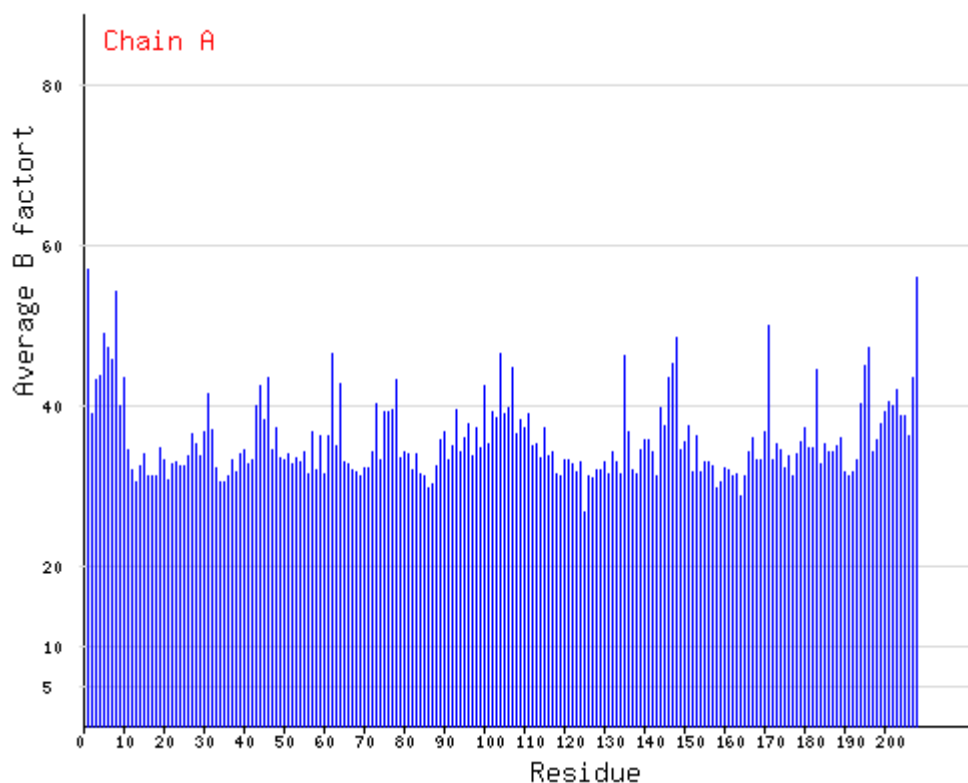


Рис.4. Значения B-фактора для усредненные для каждого остатка структуры 3D30.

Далее был проведен анализ структуры с помощью сервиса MolProbity. Название модели было подано на вход программе с указанием добавить к модели атомы водорода. В результате выдачи программы были получены результаты в виде карт Рамачандрана, параметра Clashscore, отклонений боковых цепей остатков от ротамеров и другие. Они представлены ниже.

Карты Рамачандрана

Карты торсионных углов (карты Рамачандрана) служат хорошей оценкой моделей, независимой от оптимизации. Они помогают найти маргинальные остатки по конформации остова молекулы. Однако на данный момент этот параметр оценки модели является очень известным, и авторы часто обращают на него внимание при оптимизации модели.

Карты Рамачандрана для модели 3D30 были построены с помощью сервиса MolProbity. Они представлены на рисунке 5. 98% аминокислотных остатков находятся в предпочитаемой области. А 100% в допустимой. Получается, что в модели 3D30 нет остатков, находящихся в запрещенной области. То есть нет остатков с конформацией, сильно отклоняющейся от принятой.

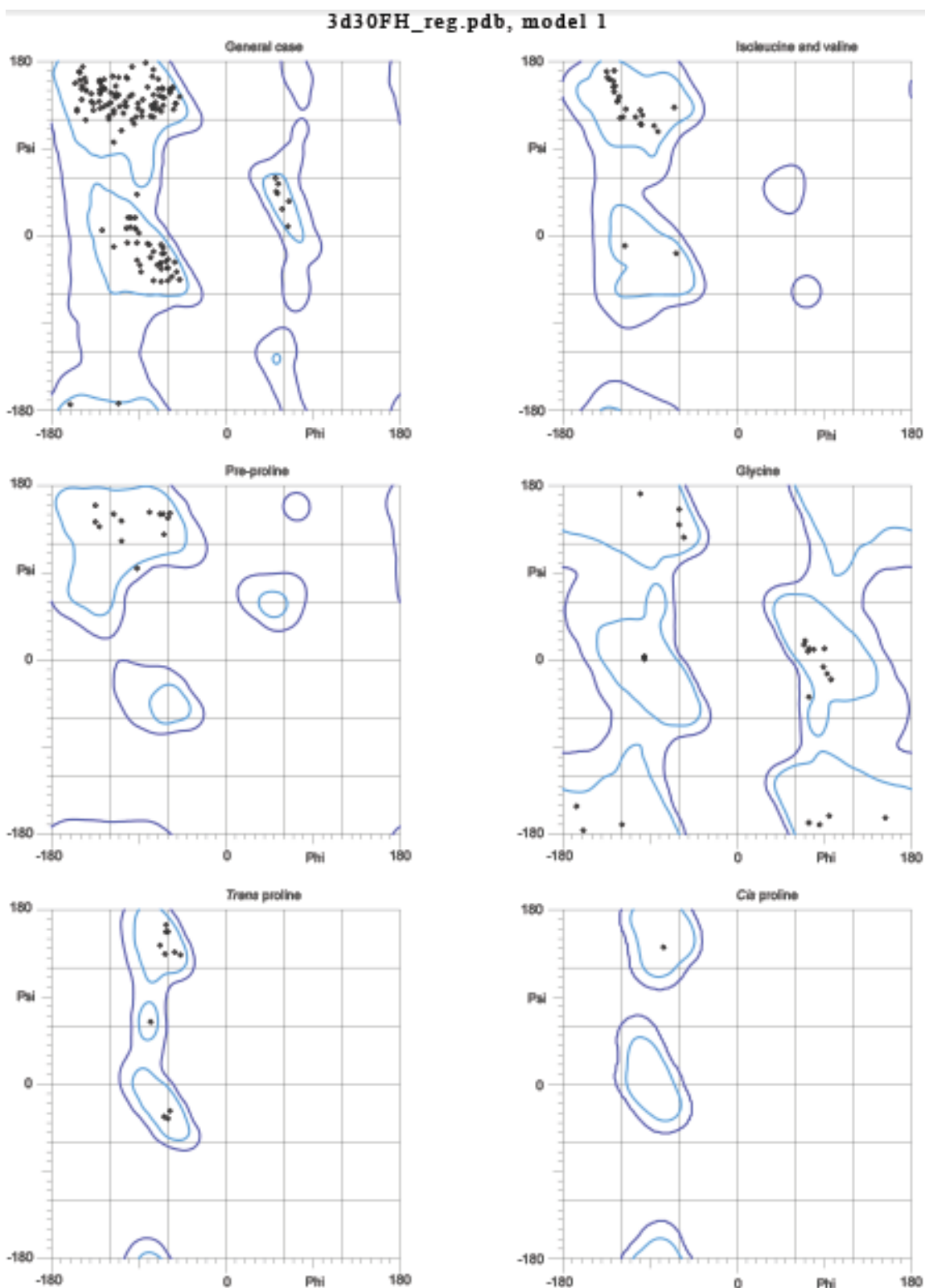


Рис. 5. Карты Рамачандрана для структуры 3D30, построенные с помощью сервиса MolProbity.

Так как все остатки, не попадающие в предпочитаемую область на карте Рамачандрана, следует считать маргинальными, то нужно обратить внимание на те 2% остатков, которые в неё попали. Это остатки Asn43, Tyr73, Gly97.

Clashscore

Параметр Clashscore – это число недопустимых наложений атомов на 1000. Для данной структуры он составляет 3.24. Перцентиль данной структуры по отношению к структурам с разрешением $1.90\text{\AA} \pm 0.25\text{\AA}$ составляет 98%. Значит, 98% структур с близким разрешением имеют Clashscore хуже, чем структура 3D30.

Остатки с маргинальными по отклонению от ротамеров боковыми цепями

Ротамеры - это боковые цепи в типичных для данного типа остатка конформациях. Параметр «Poor rotamers» - это число остатков с маргинальными по отклонению от ротамеров боковыми цепями. Этот параметр позволяет найти маргинальные остатки по боковым цепям. Кроме того, он служит хорошей оценкой качества модели, потому что не он не зависит от оптимизации.

Для структуры 3D30 есть три остатка, которые определяются как маргинальные по этому параметру. Это остатки Lys95, Lys171 и Lys188. Три остатка - достаточно большое число. Кроме того, указано также число остатков хорошо соответствующих возможным ротамерам – 168 остатков. Эти показатели говорят о том, что много боковых цепей в структуре не соответствуют типичным конформациям.

Другие параметры

Для одного C β атома структуры 3D30 зафиксировано отклонение больше, чем на 0.25\AA от ожидаемого положения.

Согласно данным сервиса MolProbity структура 3D30 имеет высокое качество. 99% структур с разрешением $1.90\text{\AA} \pm 0.25\text{\AA}$ оценены этим сервисом хуже, чем структура 3D30.

Проверка структуры с помощью пакета WHAT_IF и протокола WHAT_CHECK

Структура 3D30 без проблем прошла проверку по большинству выбранных вручную параметров протокола WHAT_CHECK. Наиболее интересные случаи, когда были обнаружены подозрительные участки структуры представлены ниже. В базе данных PDBREPORT было выявлено много различных несоответствий стандартным параметрам, в основном касающихся длин связей и углов боковых цепей в структуре. Некоторые случаи также рассмотрены ниже.

Аминокислотные остатки с подозрительным окружением

Остатки, окружение которых вызывает подозрение, обнаруженные с помощью протокола WHAT_CHECK: Met1, Lys62, Lys95, Glu208, Thr14, Asp26, Tyr73, Asn164. Возможно, что какие-то из этих остатков образуют симметричные контакты, образуют

контакты с ионами или лигандами, находятся в активном центре или же их координаты определены неверно. Однако, в базе данных PDBREPORT показаны другие остатки с подозрительным окружением: Lys188, Lys62, Arg78, Glu191.

Для остатков Asn100, Asn113 и Asn133 есть подозрение, что поворот их боковой цепи на 180° (инверсия) мог бы привести к более успешному формированию водородных связей.

Повороты пептидной цепи

Возможно, что для указанных остатков имел место поворот участка пептидной цепи на 180°. Это остатки Asp4, Gly15, Ala49, Glu75.

Альтернативные положения атомов

В структуре 3D30 есть атомы, для которых указаны альтернативные положения. Для остатков, указанных ниже, возможно лучше было бы выбрать в качестве основного другое альтернативное положение. Это остатки Ser16 и Gly17.

Ошибки в номенклатуре

Для Arg173 и Arg178 в базе данных PDBREPORT определена возможная инверсия в названиях концевых групп боковой цепи (N-H-1 вместо N-H-2, и наоборот).

Вода без водородных связей

Была обнаружена молекула воды (номер 214), не образующая водородные связи. PDBREPORT предложил её подвинуть.

Поиск и анализ маргинальных остатков в структуре

Наиболее интересные маргинальные остатки, найденные с помощью проведенного анализа структуры представлены в таблице 1. Стоит обратить внимание на то, что ни один из остатков, которые считаются необходимыми для функции белка, не был отмечен как маргинальный ни по одному из рассмотренных параметров.

Таблица 1. Маргинальные остатки, найденные в структуре 3D30.

| Маргинальный остаток | Критерии, по которым отобраны остатки |
|----------------------|---|
| Met1 | 1. RSR > 20% 2. Z-score > 2 (=6.9) 3. Подозрительное окружение (WHAT_CHECK) |
| Glu208 | 1. RSR > 20% 2. Подозрительное окружение (WHAT_CHECK) |
| Lys148 | 1. RSR > 20% 2. Z-score = 1.96 |

| | |
|---------|---|
| Arg78 | 1. RSR=19.8% |
| Lys183 | 1. RSR=19.6% |
| Asn43 | 1. Находится вне предпочитаемой зоны на карте Рамачандрана (в допустимой) |
| Tyr73 | 1. Находится вне предпочитаемой зоны на карте Рамачандрана (в допустимой) 2. Подозрительное окружение (WHAT_CHECK) |
| Gly97 | 1. Находится вне предпочитаемой зоны на карте Рамачандрана (в допустимой) |
| Lys95 | 1. Несоответствие конформации ротаметрам 2. Подозрительное окружение (WHAT_CHECK) |
| Lys188 | 1. Несоответствие конформации ротаметрам 2. Подозрительное окружение (PDBREPORT) |
| Gly15 | 1. Возможно, нужен поворот соответствующего участка пептидной цепи на 180° |
| Asn100 | 1. Подозрение на необходимость инверсии боковой цепи |
| Asn113 | 1. Подозрение на необходимость инверсии боковой цепи |
| НОН 214 | 1. Молекула воды, которая не образует водородных связей |

Met1

Met1 это первый остаток в последовательности белка. Его RSR намного больше 20% (= 46.2%), а B-фактор = 56.9. Получается, что этот координаты этого остатка совсем не соответствуют «экспериментальной» электронной плотности, и что его положение в различных молекулах кристалла сильно различается (или же остаток совершает тепловые колебания). Так как это первый атом цепи, логично было бы предположить, что его положение очень сложно определить из-за высокой подвижности для любой структуры с близким разрешением. В этом случае не очень корректно обращаться к Z-score, который для этого остатка составляет 6.9, потому что сравнение будет производиться также и с метионинами «в центре» полипептидной цепи. Хорошо было бы сравнить его именно с первыми остатками метионина других структур. Обычно первый метионин не важен для структуры и функции белка, поэтому неточность его определения не критична.

Остаток Met1 структуры 3D30 показан на рисунке 6. Электронная плотность показана на уровне подрезки 0.5. Если взять уровень подрезки выше, то визуализировать её не удастся. Получается, что значения функции электронной плотности в области Met1 низкие. На рисунке 6 видно, что есть области, в которых указано наличие атомов, но нет электронной плотности и наоборот. Концевая –CH₃ группа метионина находится далеко от участка, на котором есть электронная плотность, и на котором она вероятнее всего должна находиться. Это небольшое облачко электронной плотности вряд ли соответствует молекуле воды, так как оно очень близко к остатку и молекулы воды в этой области не найдены.

Возможно, авторам структуры стоило бы обратить внимание на этот остаток и определить координаты его атомов лучше. Однако, так как этот остаток не важен для работы белка и поддержания структуры, возможно, действия авторов оправданы.

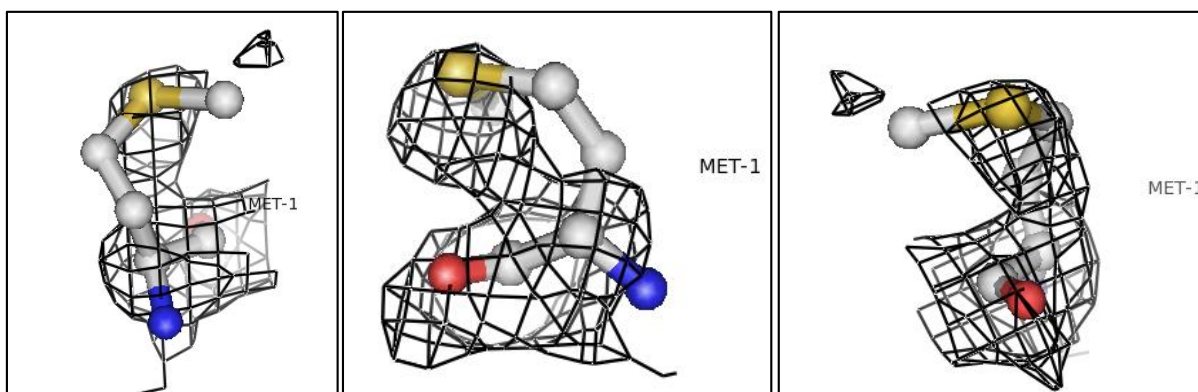


Рис.6. Визуализация функции электронной плотности в области Met1 на уровне подрезки 0.5 σ .

Tyr73

Остаток Tyr73 вызывает подозрения, так как он находится вне предпочитаемой зоны на карте Рамачандрана, и пакет WHAT_CHECK установил «некомфортное окружение» этого остатка. Визуализация электронной плотности вокруг остатка Tyr73 на уровне подрезки 2 σ представлена на рисунке 7. Видно, что остаток хорошо вписан в электронную плотность.

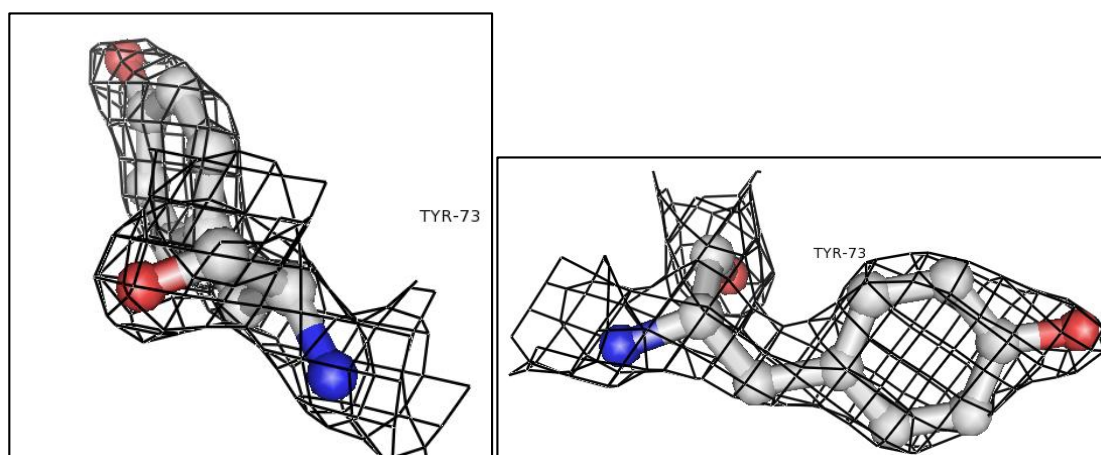


Рис.7. Визуализация функции электронной плотности в области Tyr73 на уровне подрезки 2 σ .

Посмотрим внимательнее на его окружение. Он расположен на поверхности молекулы так, что от –ОН его боковой цепи до ближайшего остатка, с которым он может образовать водородную связь (Thr12) составляет 3.4 Å. Однако, про этот остаток известно, что он образует консервативную водородную связь с Asp82, которая к тому же короче (2.7 Å). Тогда Tyr73 остается образовать водородную связь с Tyr10, но расстояние между

атомами этих остатков, способными образовать водородную связь, слишком большое (3.9 Å). Эти остатки и расстояния показаны на рисунке 8. Расположение остатков относительно молекулы белка показано на рисунке 9.

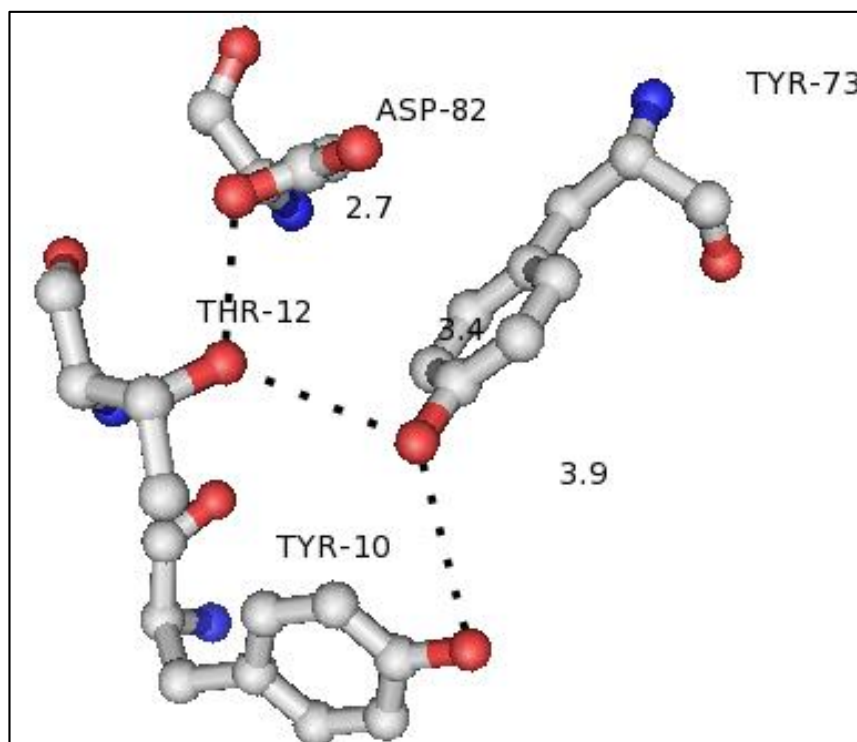


Рис.8. Остатки Asp82, Tyr73, Thr12, Tyr10 структуры 3D30. Показаны расстояния между некоторыми атомами. Расстояние между атомами представлено в ангстремах.

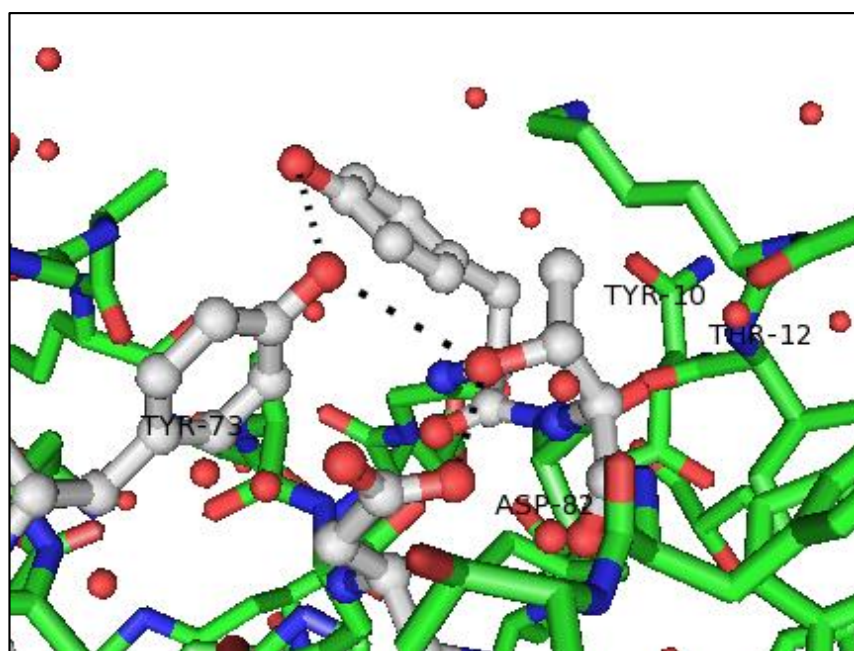


Рис.9. Остатки Asp82, Tyr73, Thr12, Tyr10 структуры 3D30 (атомы углерода показаны серым), а также участок структуры 3D30 (атомы углерода показаны зеленым). Молекулы воды представлены красными сферами.

Расположение остатка Tyr73 наводит на мысль о возможном контакте с соседней молекулой в кристалле. Симметричная единица была восстановлена. На рисунке 10 показано, что –ОН группа остатка Tyr73, вероятно, образует водородную связь с молекулой воды, которая становится видна при восстановлении симметричной единицы.

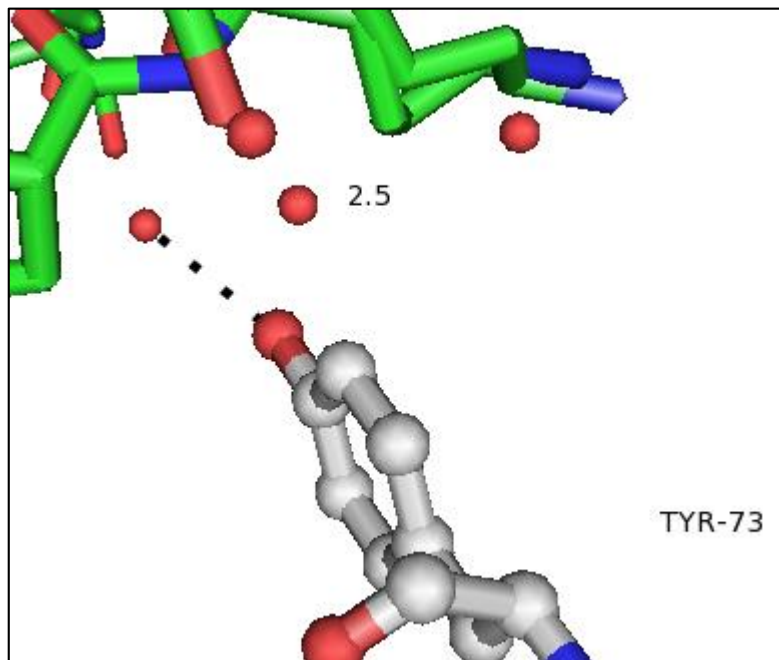


Рис.10. Остаток Tyr73 структуры 3D30 (атомы углерода показаны серым), а также участок симметричной единицы (атомы углерода показаны зеленым). Молекулы воды представлены красными сферами. Расстояние между атомами представлено в ангстремах.

НОН 214

Для молекулы воды 214 в базе данных PDBREPORT указано, что она не образует водородных связей. Молекула воды 214 показана на рисунке 11 синим в окружении других молекул воды (показаны красным) и аминокислотных остатков структуры 3D30. Однако, как видно на рисунке 11, она может образовывать водородные связи с тремя молекулами воды. Возможно, для этих молекул воды могут быть реализованы другие, более подходящие водородные связи, и на связь с водой 214 уже не остаётся места. Тем не менее, подозрение на то что молекула воды 214 не образует водородных связей выглядит неоправданным.

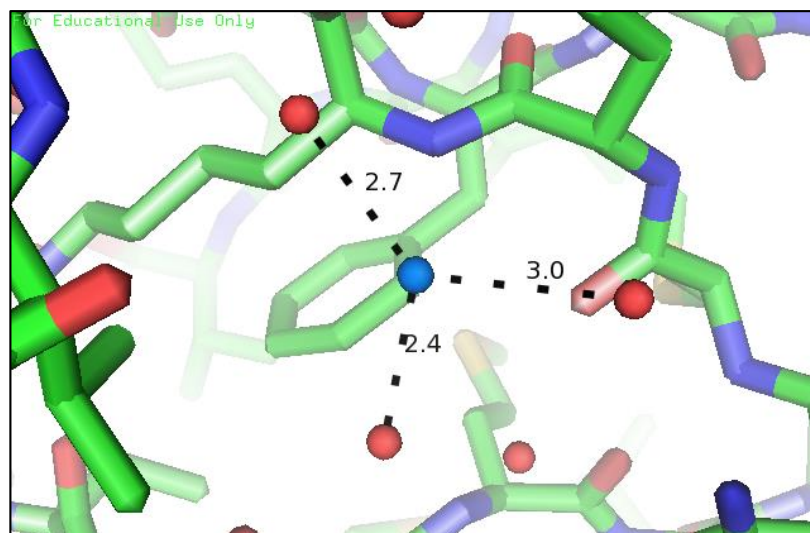


Рис.11. Молекула воды 214 структуры 3D30 (показана синей сферой), а также участок структуры 3D30 (атомы углерода показаны зеленым). Молекулы воды представлены красными сферами. Расстояние между атомами представлено в ангстремах.

Lys188

Для остатка Lys188 было выявлено несоответствие ротамерам, а также подозрительное окружение. В PDB-файле указаны две возможные конформации Lys188 с коэффициентами заполнения по 0.5 для каждого. Рассмотрим электронную плотность для остатка Lys188 на уровне подрезки 2σ , она показана на рисунке 12. Для углеродов боковой цепи не видно функции электронной плотности. Возможно, поэтому авторы предложили две возможных конформации боковой цепи. Однако удивительно, что какой-то из них был выбран так, что он не соответствует ротамерам. Здесь наблюдается некий произвол со стороны авторов.

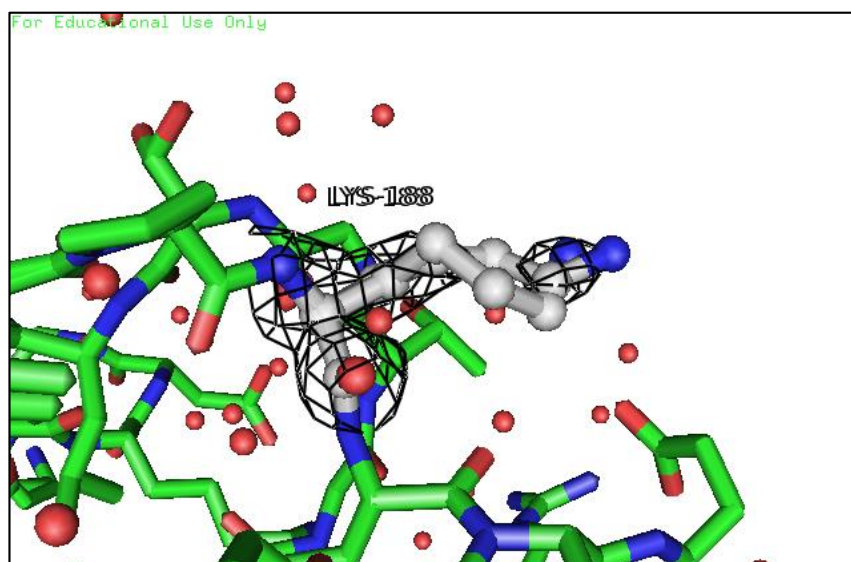


Рис.12. Визуализация функции электронной плотности на уровне подрезки 2σ в области Lys188 структуры 3D30 (атомы углерода показаны серым), а также участок структуры 3D30 (атомы углерода показаны зеленым). Молекулы воды представлены красными сферами.

Так как боковая цепь остатка Lys188 направлена вне глобулы, и не похоже, что она образует какие-либо контакты с элементами этой асимметрической единицы было решено проверить, образует ли она контакты с соседней молекулой в кристалле. На рисунке 13 показаны расстояния (до 3 \AA) от концевых групп Lys188 до элементов соседней молекулы белка в кристалле, а также отнесенных к ней молекул воды. Возможно, что какие-то из этих расстояний реализуются как водородные связи, и положение боковой цепи Lys188 стабилизируется.

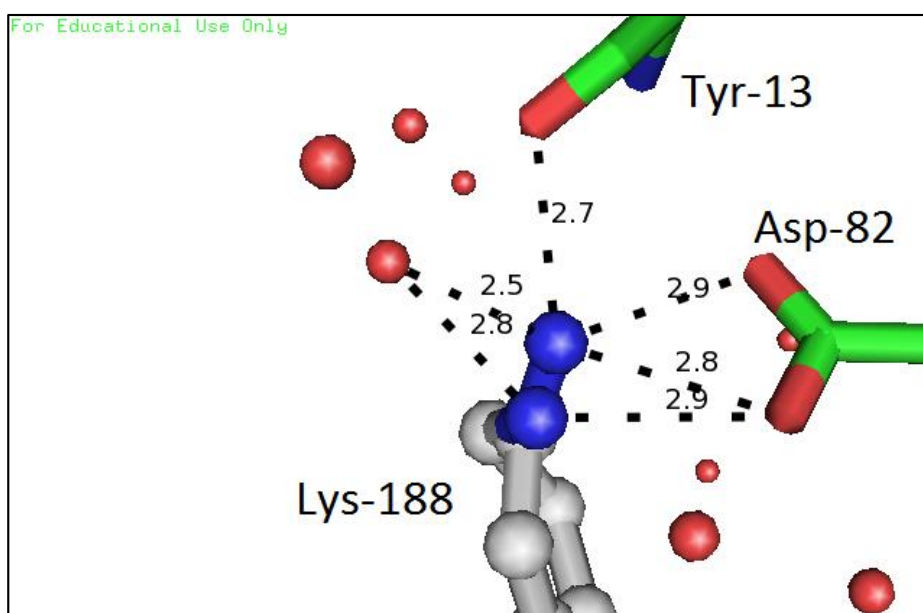


Рис. 13. Остаток Lys188 структуры 3D30 (атомы углерода показаны серым), а также участок соседней молекулы (атомы углерода показаны зеленым). Показаны расстояния (до 3 \AA) от концевых групп Lys188 до элементов соседней молекулы белка в кристалле, а также отнесенных к ней молекул воды. Молекулы воды представлены красными сферами.

Asn100

Для остатка Asn100 есть подозрение на необходимость инверсии боковой цепи. На рисунке 14 видно, что в указанной конформации Asn100 образует водородную связь с молекулой воды. Оснований для инверсии боковой цепи не отмечено. На рисунке 14 также показана электронная плотность в области остатка Asn100 на уровне подрезки 2σ . При инверсии цепи, она перестала бы вписываться в электронную плотность, а значит координаты атомов отделились бы от правильных значений.

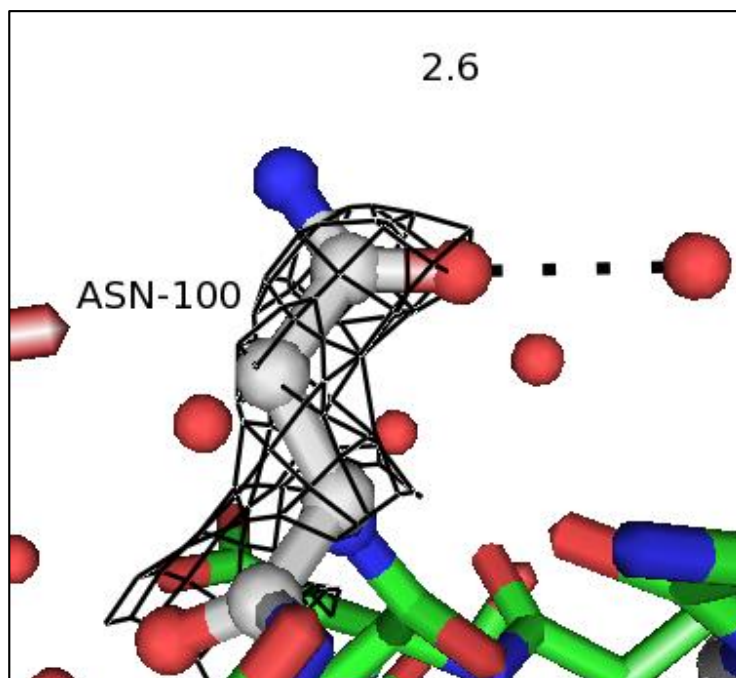


Рис.14. Визуализация функции электронной плотности на уровне подрезки 2σ в области Asn100 структуры 3D30 (атомы углерода показаны серым), а также участок структуры 3D30 (атомы углерода показаны зеленым). Молекулы воды представлены красными сферами. Расстояние от остатка Asn100 до молекулы воды (вероятная водородная связь) представлено в ангстремах.

Lys196

Остаток Lys196 не был отмечен как маргинальный ни одним из параметров. Однако, в процессе работы со структурой 3D30 на него было обращено внимание. Визуализация его электронной плотности на уровне подрезки 0.5σ показана на рисунке 15. Как видно на рисунке координаты атомов боковой цепи остатка указаны абсолютно произвольно. Очень странно, что этот остаток прошел проверку качества по такому большому числу параметров. Тем более, сто боковая цепь обращена вне глобулы. Попытаться объяснить это удивительное наблюдение не удалось даже восстановив соседние молекулы (Lys196 находится далеко от них и не образует контактов).

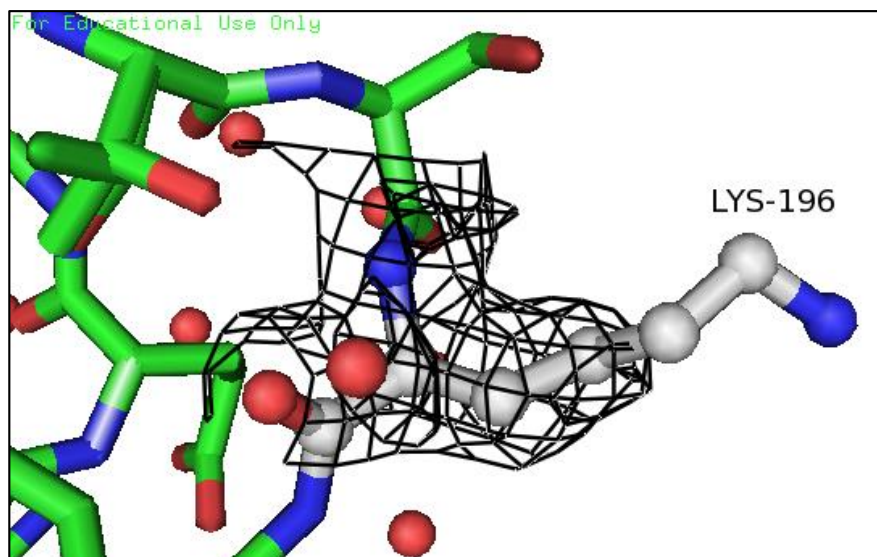


Рис.15. Визуализация функции электронной плотности на уровне подрезки 0.5σ в области Lys196 структуры 3D30 (атомы углерода показаны серым), а также участок структуры 3D30 (атомы углерода показаны зеленым).

Сравнение модели из PDB с моделью из PDB_redo.

Сервис PDB_redo оптимизирует структуру стандартной проверенной программой, исходя из представленных авторами PDB файла экспериментальных данных. В таблице 2 представлено сравнение некоторых параметров структур 3D30 оригинальной и оптимизированной.

Таблица 2. Сравнение оригинальной структуры 3D30 и оптимизированной сервисом PDB_redo.

| Параметр | Из PDB файла | Посчитанный по данным модели 3D30 | После оптимизации |
|---|--------------|-----------------------------------|-------------------|
| R | 0.1630 | 0.1662 | 0.1656 |
| R-free | 0.1910 | 0.1896 | 0.1904 |
| Соответствие ротамерам | - | -2.045 | -1.463 |
| Число доноров/акцепторов водородной связи, которые её не образуют | - | 22 | 11 |

Как видно из таблицы 2 улучшились в основном «биологические» показатели качества модели. Те показатели, по которым возможна оптимизация, изменились слабо.

Стоит также отметить, что показатели R и R-free из PDB-файла близки к посчитанным сервисом по тем же данным. Они не сильно изменились после оптимизации.

Наложение двух структур (3D30 и 3D30, оптимизированной сервисом PDB_redo) друг на друга показано на рисунках 16 и 17. Как видно из рисунка 16 вторичная структура моделей полностью совпадает, на рисунке 17 видно, что изменена конформация некоторых боковых цепей аминокислотных остатков. Розовым показана исходная структура, а голубым оптимизированная. На рисунке 17 подписаны некоторые остатки, для которых видно изменение конформации. Все они, кроме Lys196, были отмечены как маргинальные по различным параметрам, а Lys196 был отмечен вручную.

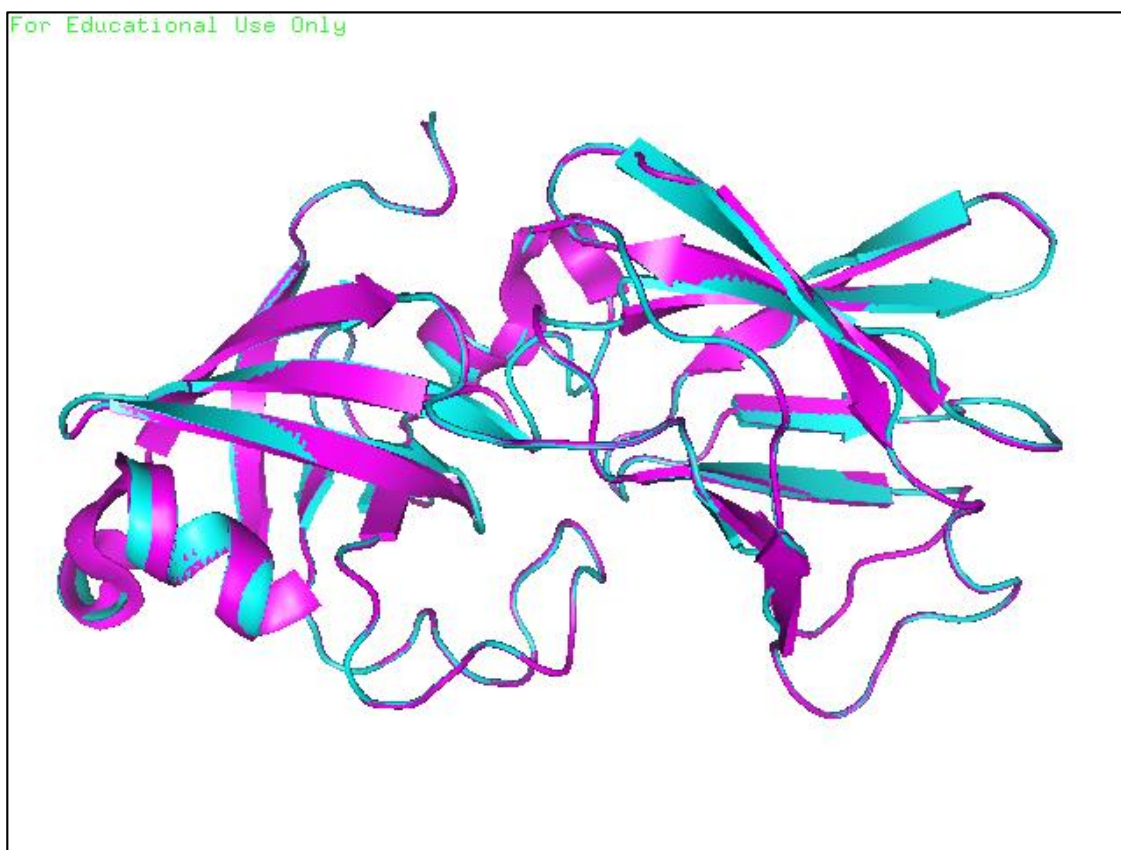


Рис.16. Сравнение вторичной структуры моделей 3D30 и 3D30, оптимизированной сервисом PDB_redo. Розовым показана исходная структура, а голубым оптимизированная.

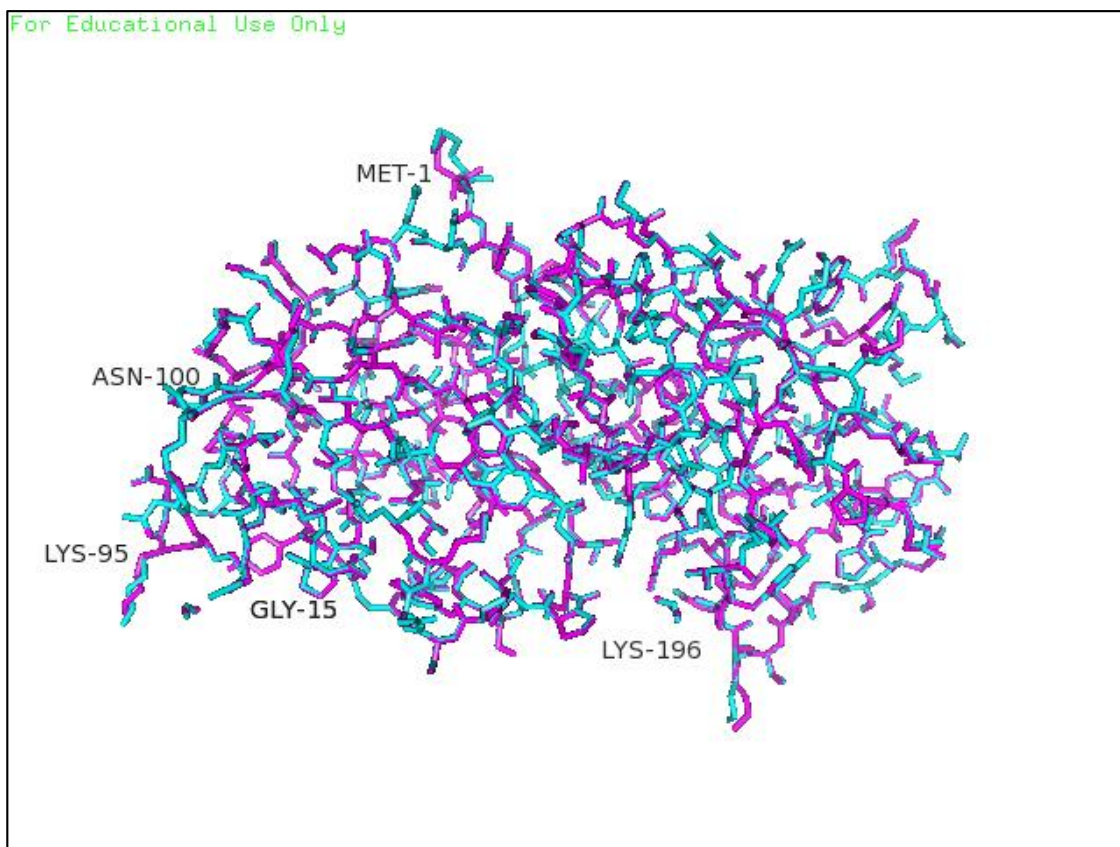


Рис.17. Сравнение структуры моделей 3D30 и 3D30, оптимизированной сервисом PDB_redo. Розовым показана исходная структура, а голубым оптимизированная. Подписаны некоторые остатки, для которых видно изменение конформации.

Заключение

Модель 3D30 кажется сделана очень хорошо. Набор рефлексов обладает хорошим разрешением и большой полнотой данных. Показатели качества структуры в целом также очень хорошие. Также отлично выглядят карты Рамачандрана, построенные для этой модели. Все параметры, которые возможно оптимизировать, оптимизированы.

Однако для некоторых аминокислотных остатков наблюдается несоответствие ротамерам, есть большое количество доноров/акцепторов водородной связи, которые её не образуют. Подробный анализ остатков, определённых как маргинальные, выявил, что некоторые остатки могут быть не маргинальными, просто они могут взаимодействовать с соседними молекулами ячейки, которые не попадают в данную асимметрическую единицу. Также стоит отметить, что ни один из остатков, отмеченных как важные для функции, не попал в разряд маргинальных.

Также хочется отметить на примере остатка Lys196, что параметры, определяющие маргинальные остатки, могут некоторые из них пропустить.

Ссылки на серверы и базы данных

Структура 3D30 на сервере PDB

<http://www.rcsb.org/pdb/explore/explore.do?structureId=3d30>

Структура 3D30 на сервере EDS

<http://eds.bmc.uu.se/cgi-bin/eds/uusfs?pdbCode=3d30>

Сервис MolProbity

<http://molprobity.biochem.duke.edu/index.php>

Протокол WHAT_CHECK

<http://swift.cmbi.ru.nl/servers/html/index.html>

Структура 3D30 на сервисе PDB_redo

http://www.cmbi.ru.nl/pdb_redo/d3/3d30/index.html

База данных PDBREPORT

<http://swift.cmbi.ru.nl/gv/pdbreport/>

Список литературы

1. Kerff F., Amoroso A., Herman R., Sauvage E., Petrella S., Filée P., Charlier P., Joris B., Tabuchi A., Nikolaidis N., and Cosgrove D. J. (2008) Crystal structure and activity of *Bacillus subtilis* YoaJ (EXLX1), a bacterial expansin that promotes root colonization., *Proc. Natl. Acad. Sci. U. S. A.*, **105**, 16876–81.
2. Sampedro J. and Cosgrove D. J. (2005) The expansin superfamily., *Genome Biol.*, **6**, 242.
3. Li Y., Darley C. P., Ongaro V., Fleming A., Schipper O., Baldauf S. L., and McQueen-Mason S. J. (2002) Plant expansins are a complex multigene family with an ancient evolutionary origin., *Plant Physiol.*, **128**, 854–64.

4. Bais H. P., Fall R., and Vivanco J. M. (2004) Biocontrol of *Bacillus subtilis* against infection of *Arabidopsis* roots by *Pseudomonas syringae* is facilitated by biofilm formation and surfactin production., *Plant Physiol.*, **134**, 307–19.
5. Madej T., Gibrat J. F., and Bryant S. H. (1995) Threading a database of protein cores., *Proteins*, **23**, 356–69.
6. Yennawar N. H., Li L.-C., Dudzinski D. M., Tabuchi A., and Cosgrove D. J. (2006) Crystal structure and activities of EXPB1 (*Zea m 1*), a beta-expansin and group-1 pollen allergen from maize., *Proc. Natl. Acad. Sci. U. S. A.*, **103**, 14664–71.
7. Li L. C. and Cosgrove D. J. (2001) Grass group I pollen allergens (beta-expansins) lack proteinase activity and do not cause wall loosening via proteolysis., *Eur. J. Biochem.*, **268**, 4217–26.
8. Georgelis N., Tabuchi A., Nikolaidis N., and Cosgrove D. J. (2011) Structure-function analysis of the bacterial expansin EXLX1., *J. Biol. Chem.*, **286**, 16814–23.
9. Boraston A. B., Bolam D. N., Gilbert H. J., and Davies G. J. (2004) Carbohydrate-binding modules: fine-tuning polysaccharide recognition., *Biochem. J.*, **382**, 769–81.
10. Dessau M. A. and Modis Y. (2011) Protein Crystallization for X-ray Crystallography, *J. Vis. Exp.*