**МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ**

**имени М.В.ЛОМОНОСОВА**

##### ФАКУЛЬТЕТ БИОИНЖЕНЕРИИ И БИОИНФОРМАТИКИ

**Отчет о качестве расшифровки структуры белка 1MKI методом рентгеноструктурного анализа**

**Работу выполнил:**

**Новаковский Г.Э**

**Группа 402**

**Москва, 2014г.**

**Оглавление**

**1.Аннотация..................................................................................................................................3**

**2. Введение…….............................................................................................................................4**

**3. Результаты и обсуждение.......................................................................................................5**

**3.1. MolProbity...................................................................................................................5**

**3.2. PDB…………………...................................................................................................7**

**3.3. EDS…………………………………….......................................................................7**

**3.4 PDB\_redo…………………………………………………………………………....10**

**3.5 Таблица с маргиналами………………………………………………………….10**

**3.6 Важные остатки…………………………………………………………………...10**

**4. Выводы....................................................................................................................................12**

**5. Список использованных источников………....................................................................13**

1. **Аннотация**

В данном отчете представлен анализ структуры PDB-ID 1MKI. Была сделана попытка наиболее подробным образом рассмотреть все индикаторы качества модели на разного рода ресурсах (см. Список использованных источников). Были сделаны соответствующие выводы о качестве 1MKI.

Целью данного отчета являлось именно выявление ошибок и недочетов в данной структуре.

**2. Введение**

Расшифровка структуры фермента глутаминазы (PDB ID 1MKI), кодируемого геном YbgJ в организме Bacillus subtilis subsp. subtilis str. 168 (locus tag - BSU02430) приводится в статье [1]. Глутаминазы принадлежат к большому суперсемейству серин-зависимых бета-лактамаз и пенициллин-связывающих белков и катализируют гидролитическое дезаминирование L-глутамина, превращая его в L-глутамат. В данной статье [1] авторы помимо струтуры YbgJ из Bacillus subtilis также приводят данные расшифровки глутаминаз YlaM (из того же организма), YbaS и YneH (оба из E.coli). Авторам удалось показать (через направленный мутагенез) консервативные остатки, которые принимают непосредственное участие в катализе и присутствуют в активном центре фермента (главный из них – это Ser74; остальные Lys77, Lys268, Ser269). Помимо этого удалось предложить механизм реакции глутаминаз, измерить активность ферментов. Структурные формулы B.subtilis (YbgJ) и E.coli (YbaS) выявили ряд сходств с бета-лактамазами в плане укладки структур и ряда консервативных остатков. Также авторам удалось показать, что глутаминазы YbaS и YneH из E.coli способствуют кислотоучстойчивости данного организма, что говорит об их высокой биологической значимости.

Иными словами, 1MKI (глутаминаза 1 из Bacillus subtilis) не являлась единственной задачей авторов статьи, эта структура – один из результатов большой работы, проделанной с целью выявить закономерности в структуре глутаминаз в плане укладки и консервативных остатков, а также механизм превращения L-глутамина в L-глутамат этими ферментами.

**3. Результаты и обсуждение**

**3.1 MolProbity**

Если воспользоваться сервисом MolProbity, то можно получить следующую информативную таблицу:

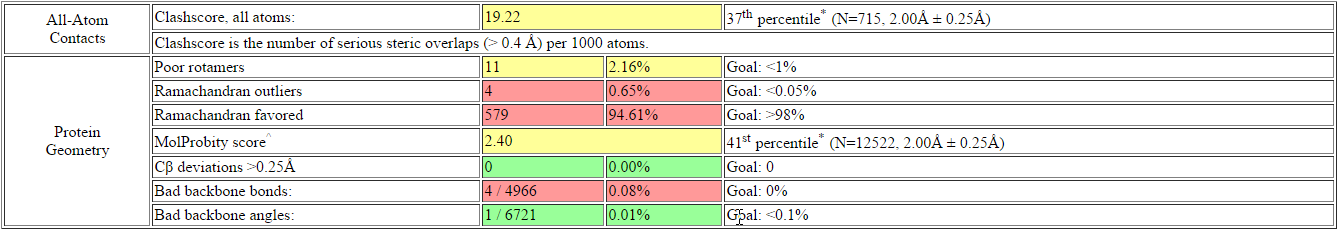


Рисунок 1. Данные сервиса MolProbity

В первой графе (Clashscore, all atoms) представлено число недопустимых наложений атомов на 1000. В данном случае – это 19.22, то есть 1,922 процента остатков. Перцентиль 37 процентов (данной структуры по отношению к структурам примерно такого же разрешения) говорит о том, у 37% структур Clashscore хуже данного, а хуже – значит больше. Если учесть, что лучшее значение перцентили – это 100%, то не трудно понять, что значение Clashscore для данной структуры (1mki) далеко не лучший.

Далее, если обращаться к следующим графам таблицы (рисунок 1), то можно извлечь следующую информацию по данной структуре:

1 - 11 остатков с маргинальными по отклонению от ротамеров боковыми цепями (2.16% от всех остатков, хотя в идеале должно быть <1%; это нельзя назвать плохим значением);

2 – 4 маргинала по карте Рамачандрана, лежат вне допустимой области (это 0.65% от всех остатков, тогда как при высоком разрешении таких остатков должно быть не больше 0.05%);

3 – 579 остатков в предпочитаемой области карты Рамачандрана, 94.61%; в идеале таких должно быть > 98%;

Менее информативные параметры указывают на то, что в структуре 1mki число С\_beta с неприемлемым отклонением от ожидаемого положения равно 0. Плохим значением является число ковалентных связей, существенно отклоняющихся от теории, равное 4 (в идеале оно должно быть равным 0). Число валентных углов, существенно отклоняющихся от теории, равно 1.

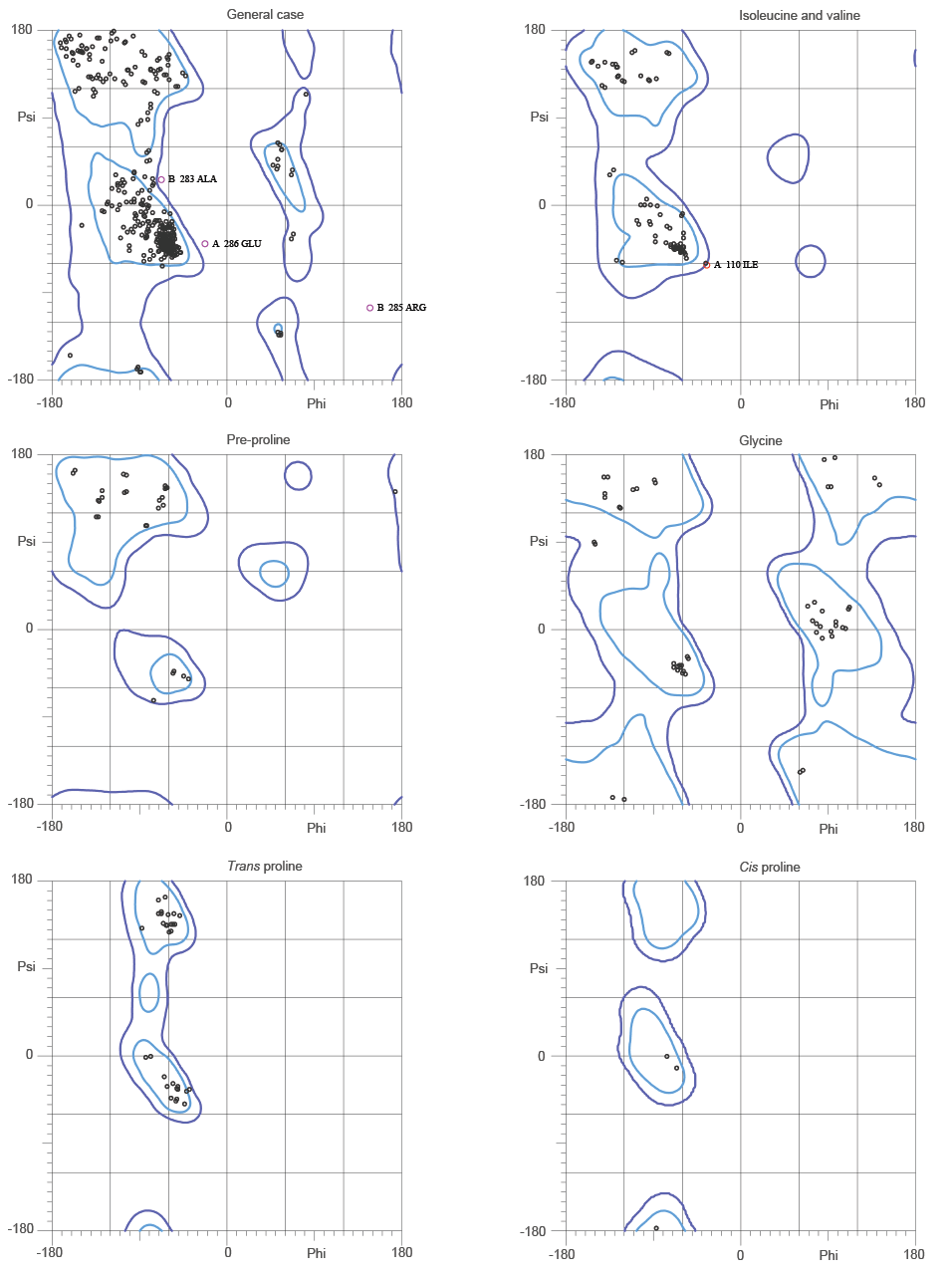


Рисунок 2. Карта Рамачандрана для структуры 1MKI

На рисунке 2 представлена карта Рамачандрана и соответствующие маргинальные остатки (лежащие вне разрешенной области): chain B Ala283, chain A Glu286, chain B Arg285 и chain A ILE110. На рисунке 3A представлено изображение одного из таких остатков в PyMol.

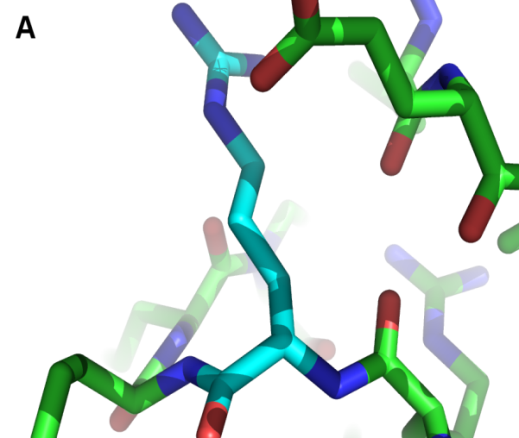
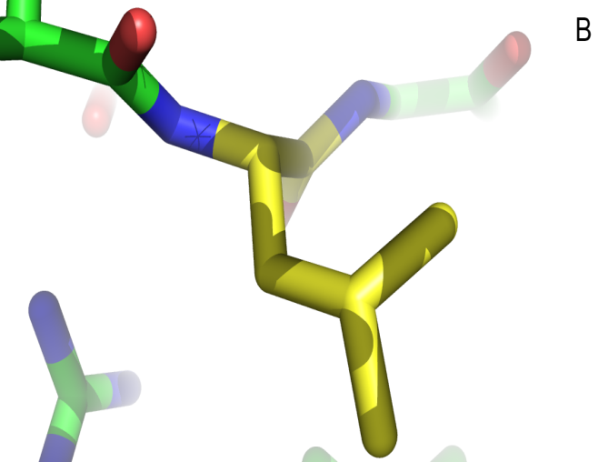
 

Рисунок 3. (A) Изображение маргинального остатка (по карте Рамачандрана) аргинин-285 цепи B, атомы углерода окрашены в cyan. (B) Изображение маргинального остатка (по ротамерам) лейцин-49 цепи А, атомы углерода окрашены в светло-желтый.

Также MolProbity выдал 11 маргиналов-ротамеров: chain A (Glu23, Leu49, Leu184, Arg284, Leu314), chain B (Asn45, Leu49, Ile51, Asn67, Leu309, Leu315). На примере одного остатка действительно видно, как неправильно повернута боковая группа (рисунок 3B).

Помимо этого MolProbity позволяет увидеть возможные инверсии боковых цепей Asn, Gln, His. Было выдано 11 clear evidence for flip: chain A (ASN45, HIS62, HIS227, ASN307), chain B (GLN17, HIS62, ASN67, ASN115, ASN122, ASN126, GLN233).

**3.2 PDB**

На сервере PDB лежит следующая информация по поводу данной структуры 1MKI:

1 – Разрешение 2.00 ангстрем (нормальное);

2 – R-value 0.216 (21.6%) – не очень хорошее значение;

3 – R-free 0.245 (24.5%). Разница R-free – R-value = 2.9% (переоптимизации нет);

4 – Группа симметрии P (примитивная ячейка) 21 21 2;

5 – Параметры элементарной ячейки кристалла: a=71.33, b=181.48, c=51.49 (в ангстремах), alpha=90, beta=90, gamma=90 (в градусах); в статье авторы говорят, что в асимметрической ячейке лежит две молекулы (две цепи А и В), тогда как сама молекула глутаминазы – димер димеров.

Число рефлексов (измеренных), лежащих в файле структурных факторов – 46691. Из этих измеренных рефлексов 34439 являются хорошими (значение модуля структурного фактора превышает соответствующее значение сигма в 3 раза – проверено с помощью питоновского скрипта), что составляет около 73.8% от всего набора рефлексов. То есть, по-честному авторы должны были выкинуть плохо померенные рефлексы и оставить только эти - хорошие. В файле PDB для данной структуры приведены пороги на отбор рефлексов для измерения: data cutoff sigma 0.000, data cutoff high abs(F) null, data cutoff low abs(F) null. Исходя из этих чисел, кажется, что авторы не особо трудились над отбором достоверных рефлексов, а брали практически все. Также благодаря PDB можно получить информацию, касающуюся определения структуры – то, что описано в статье [1]. Так, например, фазовая проблема при определении структуры была решена с помощью метода многоволнового аномального рассеяния (MAD). В этом методе использовались 3 разные длины волны - 0.97910, 0.97921, 0.954 (ангстрем). В качестве остатка с аномально рассеивающим атомом использовался селенометионин.

**3.3 EDS**

Если обратиться с серверу EDS, то можно найти информацию об RSR структуры 1MKI - 0.151 (это нельзя назвать ни хорошим, ни плохим значением). Полнота данных (Completeness of data) равна 95.0% (в файле PDB указана цифра 95.2). Минимальное и максимальное разрешение для использованных рефлексов равны 40.69 и 2.00 ангстрема соответственно. На этом же сервере можно найти информацию, касающуюся Z-score для отдельных остатков (среднее значение RSR и sigma считались для структур с разрешением в диапозоне 1.80-2.00Å). На рисунке 4 можно видеть таблицу с сервера EDS, показывающую информацию, касающуюся остатков для цепей А и B.

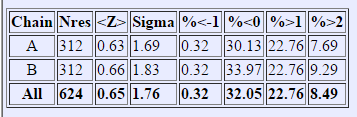
****

Рисунок 4. Таблица с данными по Z-score для структуры 1MKI

Видно, что 8.49% остатков для всей структуры имеет плохой z-score (>2). Значит, для этих остатков координаты атомов расшифрованы хуже, чем для остальных структур с подобным разрешением. Примеры таких остатков – chain A (ILE229 Z=4.02, ALA283 Z=17.25), chain B (ARG168 Z=3.46, ARG284 Z=12.91).

Далее, EDS позволяет увидеть участки (рисунок 5) с повышенным RSR фактором. Они особенно подозрительные на наличие неточностей в координатах атомов.

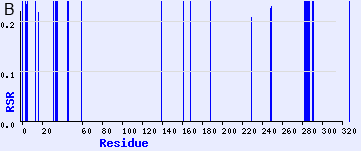
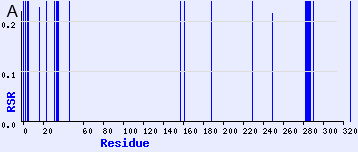
****

Рисунок 5. Участки с повышенным RSR фактором для цепей A (A) и B (B).

Эти изображения позволяют увидеть конкретные аминокислотные остатки, у которых плохой RSR-фактор (24 маргинала для цепи А и 29 – для цепи В). Видно, что у обеих цепей есть непрерывный отрезок из аминокислот (один и тот же) с повышенным RSR-фактором. Это остатки с 282 по 287 (Ser-Ala-Arg-Arg-Glu-Gln, у них же и плохой Z-score). Для примера изображен этот самый отрезок для цепочки А (рисунок 6). Видно, как плохо вписаны в электронную плотность соответствующие аминокислоты, если вообще можно сказать, что они вписаны.

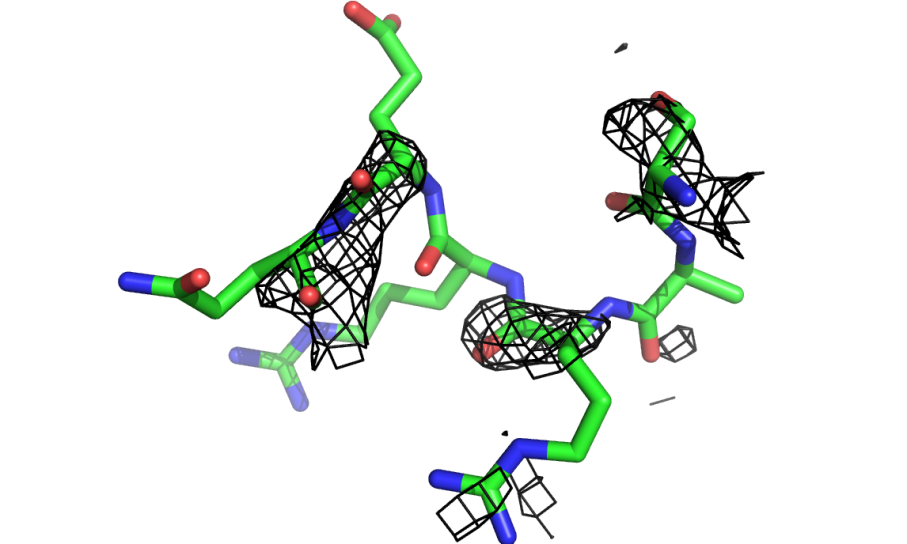


Рисунок 6. Отрезок из аминокислот 282-287 цепи А 1MKI, как пример остатков с плохим RSR. Уровень подрезки электронной плотности – 1.5 сигма.

Стоит отметить, что если посмотреть на эти остатки в PDB-файле, то (только для цепи А) у атомов аминокислот с 283 по 286 occupancy (коэффициент заселенности) равен 0.5 (причем, не показаны альтернативные координаты для этих атомов, в половине ячеек они есть, а в другой половине - нет). Также у этих остатков плохой температурный фактор (больше 50), что наряду с невысоким разрешением (2 ангстрема), может объяснить тот факт, что эти остатки плохо вписаны в электронную плотность. Если посмотреть на аналогичный отрез цепи B, то все соответствующие атомы имеют occupancy 1, однако температурный фактор у них очень высокий, у некоторых даже выше 100, что просто ужасно.

EDS позволяет посмотреть на распределение значений температурного фактора для разных остатков в цепях А и В структуры 1MKI (рисунок 7). На этом рисунке красным выделены пики, отвечающие тому самому отрезку из остатков с плохим RSR-фактором.

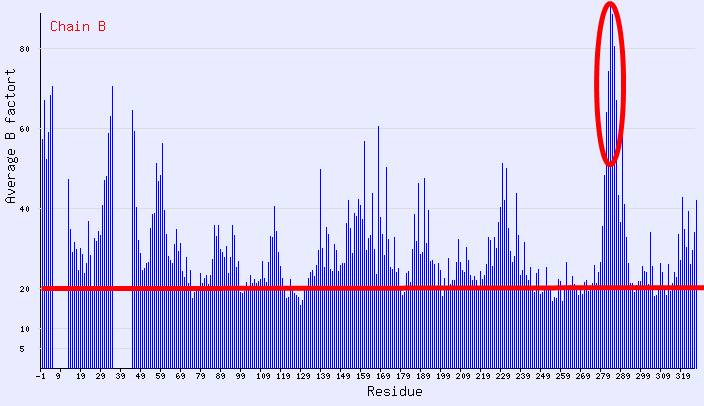
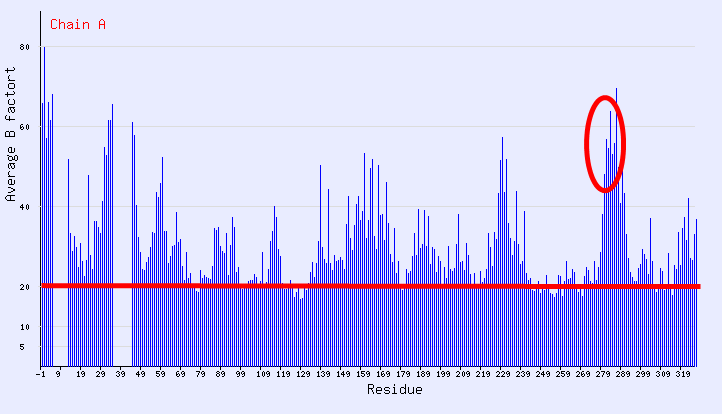
****

Рисунок 7. Распределение значений температурных факторов для цепей А и В структуры 1MKI. Красным овалом обведены значения для остатков 282-287 в обеих цепях. Красная линия означает границы нормального значения температурного фактора (не больше 20).

Видно, что значения температурных факторов для остатков из структуры 1MKI очень плохие. Разрешение структуры невысокое, значит для каждого атома пик электронной плотности очень размытый. Исходя из этого, трудно сказать что-то окончательное по поводу всей структуры (но вряд ли что-то принципиальное). Наверняка, при хорошем разрешении было бы возможно отличить разные пики от одного, что позволило бы говорить о разных конфигурациях остатков в разных ячейках и уточнить данные.

На рисунке 7 видны пропуски для остатков в обеих цепях с 6 по 12 и с 36 по 44 остатки. Этих остатков нет в PDB файле. Они просто пропущены в структуре. Авторы статьи [1] указывают на это и говорят, что это связано с разупорядоченностью соответствующих аминокислот и невозможностью вписать их в электронную плотность (рисунок 8).

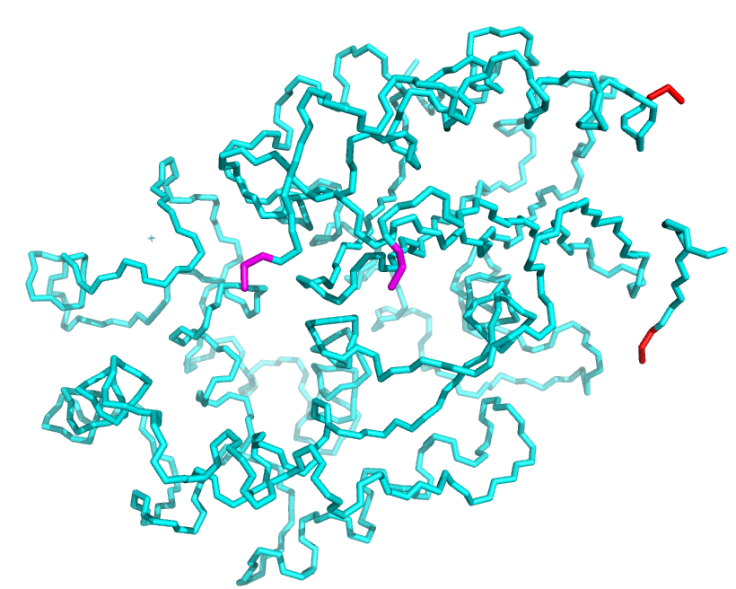


Рисунок 8. Изображена цепочка А структуры 1MKI (пептидный остов без боковых групп). Красным помечены остатки 13 и 5. Розовым – остатки 35 и 45. В каждой паре между остатками видны соответствующие пробелы.

**3.4. PDB\_redo**

PDB\_redo позволяет оптимизировать (уточнить) любую структуру с помощью ряда инструментов через полностью автоматизированный процесс, не требующий никакого вмешательства. Если сравнить данные для 1MKI из PDB и из PDB\_redo, то можно обнаружить некоторые улучшения в индикаторах качества модели. Так, например, R-factor и R-free после полной оптимизации приобрели значения 0.1689 и 0.2032 соответственно (в PDB их значения равны 0.2120 и 0.2450, см. выше). Помимо этого улучшились показатели для карты Рамачандрана, для ротамеров и для ряда других факторов.

**3.5. Таблица с маргиналами**

Исходя из всех тех индикаторов качества структуры, которые описаны выше, была построена сводная таблица маргиналов (не всех, а взятых для примера 10 аминокислот с наиболее выраженными отклонениями). В ней приводится номер остатка, цепь, которой он принадлежит и признак(-и), на основе которых этот остаток выделен, как маргинал (рисунок 9).

|  |  |  |
| --- | --- | --- |
| Маргинальные остатки | | |
| остаток | цепь | причина |
| Ala283 | A (B) | вне разрешенной области на карте Рамачандрана; плохой z-score и температурный фактор |
| Glu286 | A (B) | вне разрешенной области на карте Рамачандрана; плохой z-score и температурный фактор |
| Arg285 | A (B) | вне разрешенной области на карте Рамачандрана; плохой z-score и температурный фактор |
| Arg284 | A (B) | плохая конформация боковой группы; плохой z-score и температурный фактор |
| Ser282 | A (B) | плохой z-score и температурный фактор |
| Gln287 | A (B) | плохой z-score и температурный фактор |
| Asn67 | B | плохая конформация боковой группы; инверсия; температурный фактор |
| Ala35 | A (B) | плохой z-score и температурный фактор |
| Asn45 | B | плохая конформация боковой группы; плохой z-score и температурный фактор |
| Gln233 | B | плохой температурный фактор и инверсия |

Рисунок 9. Таблица с некоторыми маргинальными остатками.

Как пример маргинальных остатков более подробно были рассмотрены выше остатки из двух цепей Ser282-Ala283-Arg284-Arg285-Glu286-Gln287 (рисунок 5, 6, 7).

**3.6. Важные остатки**

Как указывают авторы статьи [1], в структуре 1MKI имеется ряд важных и консервативных остатков, формирующих активный центр и участвующих в катализе. Главный из них – это Ser74; остальные Lys77, Lys268, Ser269. В ходе проведенного анализа установлено, что ни один из данных остатков не является маргиналом по рассмотренным индикаторам (ротамеры, карта Рамачандрана, температурный фактор, Z-score). Не рассматривалось только благоприятное окружение, но в этом плане никаких проблем нет, как показано в статье [1]. Эти остатки и окружающая их электронная плотность с уровнем срезки 1.5 сигма представлены на рисунке 10.

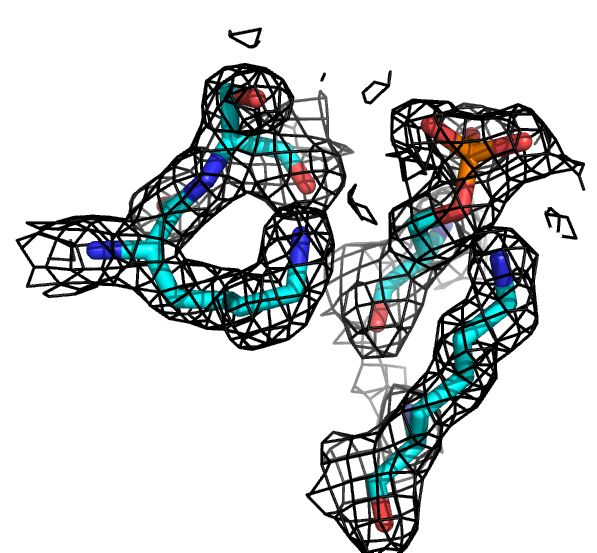


Рисунок 10. 4 важных аминокислотных остатка (из цепи А) структуры 1MKI, определяющие ее активный центр.

На рисунке 10 видно, как хорошо все атомы ложатся в «экспериментальную» электронную плотность. С гидроксильной группой серина связан фосфор (на рисунке 10 покрашен оранжевым), так что здесь представлен фосфосерин. Почему это так, неясно. Авторы про это ничего не пишут.

**4. Выводы**

По результатам проведенного анализа структуры 1MKI можно сказать, что качество последней является удовлетворительным. Структура померена с нормальным разрешением, относительно неплохо оптимизирована. Но есть и много недостатков.

1 – достаточно большое количество маргиналов на карте Рамачандрана;

2 – немалое число аминокислот с неправильной конформацией боковой цепи;

3 – разрешение подсчитано нечестно; необходимо было выкинуть больше рефлексов из экспериментальных данных;

4 – пропущены остатки 6-12 и 36-44 в обеих цепях;

5 – очень плохой средний температурный фактор;

6 – большое количество остатков с неудовлетворительным z-score;

**5. Список использованных источников**

1. – за основу бралась статья авторов Greg Brown, Alex Singer, Michael Proudfoot, Tatiana Skarina, Youngchang Kim, Changsoo Chang, Irina Dementieva, Ekaterina Kuznetsova, Claudio F. Gonzalez, Andrzej Joachimiak, Alexei Savchenko и Alexander F. Yakunin **Functional and Structural Characterization of Four Glutaminases from Escherichia coli and Bacillus subtilis** (2008);
2. – сервис MolProbity (<http://molprobity.biochem.duke.edu/>);
3. – сервис PDB (<http://molprobity.biochem.duke.edu/>);
4. – сервер EDS (<http://eds.bmc.uu.se/eds/index.html>);
5. – программа PyMol v-1.7.2.1 (<http://www.pymol.org/>);