

МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ИМЕНИ М.В.
ЛОМОНОСОВА
ФАКУЛЬТЕТ БИОИНЖЕНЕРИИ И БИОИНФОРМАТИКИ

Отчёт по качеству расшифровки структуры
белка **TenI** (PDB код **3QH2**)
методом рентгеноструктурного анализа

Выполнила:
студентка 4-го курса
Малеева Александра

Москва
2015

Аннотация

В данном отчете проведен поиск и анализ информации по качеству расшифровки структуры белка TenI (PDB-код - 3QH2 [2]) из *Bacillus subtilis*. Рассмотрены некоторые индикаторы локального качества структуры для определенных аминокислотных остатков. В ходе работы удалось познакомиться с различными серверами, которые предоставляют возможность получения параметров для оценки качества белков.

Введение

Тиамин является важным кофактором в организмах прокариот и низших эукариот. Его биосинтез – сложный и многоступенчатый процесс, который достаточно хорошо изучен в микроорганизмах [7].

В *Bacillus subtilis* гены *tenI* кластеризованы с генами, которые вовлечены в биосинтез тиамин [8]. Белок TenI – продукт гена *tenI*. Ранее было показано, что TenI участвует в регуляции внеклеточных ферментов (например, щелочной протеазы) [9]. Однако эту функцию сильно репрессирует тиамин, поэтому данная регуляторная роль не давала никакого понимания истинной функции этого белка [8]. Впоследствии TenI был описан как тиаминаз II, которая является защитным белком от деградированного тиамин [10].

TenI из *Bacillus subtilis* (Рис.1) - димер с молекулярным весом 22 783 Да. Состоит из 221 аминокислотного остатка [2]. Он очень схож по последовательности с тиаминфосфат синтазой ThiE [8]. Было выяснено, что TenI – неизвестное недостающее звено, тиазольная таутомераза, в биосинтезе тиамин [8].

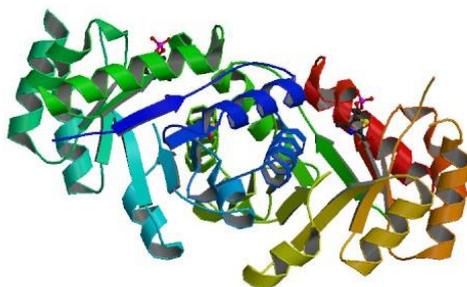


Рис. 1. Белок TenI из Bacillus subtilis (из PDB)

Исследуемый белок относится к семейству тиамин-монофосфат синтаз (TMP). Интересно, что он является единственным белком в семействе, который обладает регуляторной функцией.

В статье, в которой была описана структура TenI в комплексе с продуктом реакции, авторы также приводят предполагаемый механизм катализируемой им реакции (Рис. 2). Ранее этими же авторами была представлена его кристаллическая структура [11].

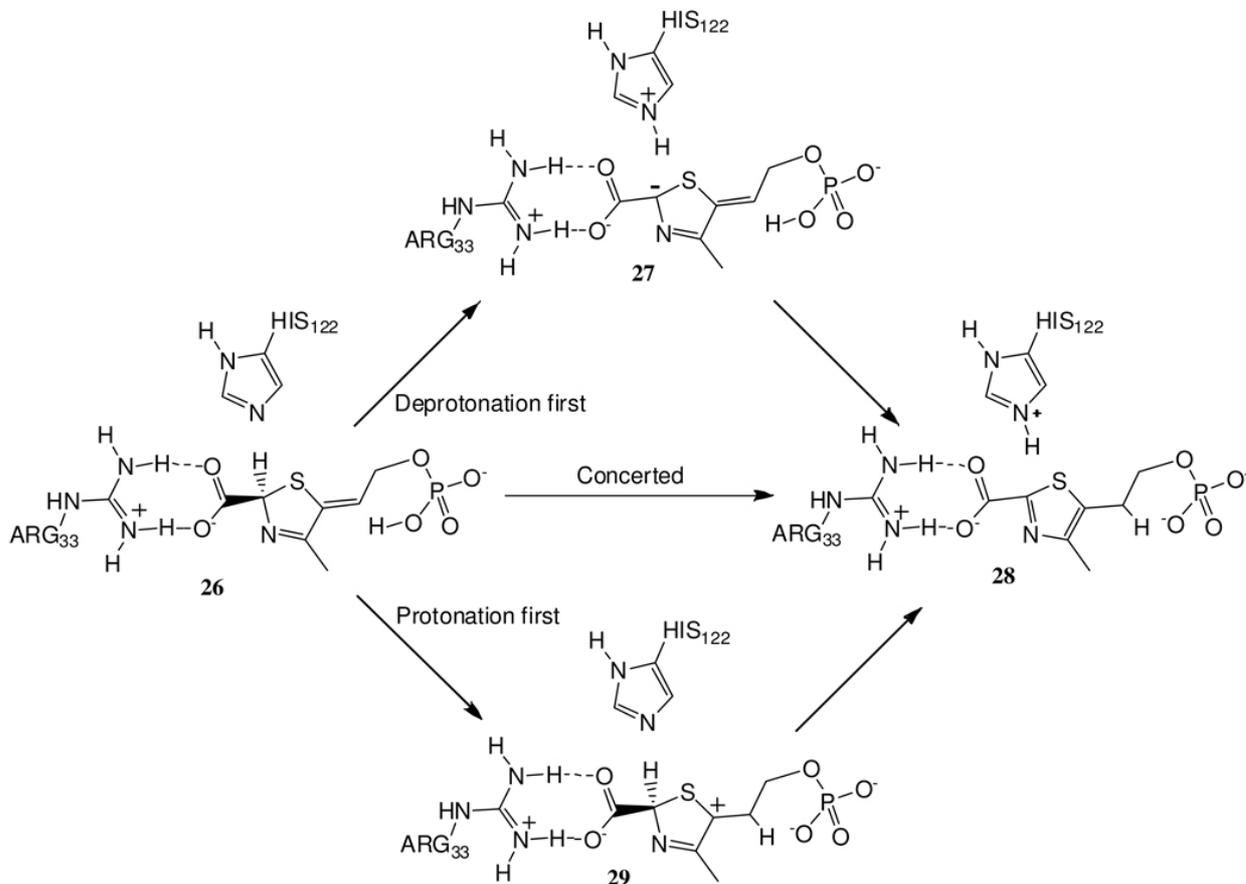


Рис. 2. Предполагаемый авторами механизм для катализируемой TenI реакции ароматизации тиазола

Результаты и обсуждение

Общая информация о модели

Комплекс 3QH2 состоит из четырех цепей А, В, С, D. Содержит лиганды 3NM (цепи А, В, С, D) и сульфат-ион (цепь В).

Комплекс был расшифрован в 2011 году группой ученых Nazra AB, Nan Y, Chatterjee A, Zhang Y, Lai RY, Ealick SE, Begley TP. Авторами было измерено 55590 рефлексов. Из них 55267 использовалось для оптимизации модели. Тестирование проводилось на 2801 рефлексах, что составляет 5,3%. Фазовая проблема решалась методом молекулярного замещения с использованием ранее описанной этими же авторами структуры 1YAD из *Bacillus subtilis*.

- Разрешение: 2,23 Å
- Диапазон расширения: 42.75 - 2.23 Å

- Полнота набора структурных факторов: 99,4 %

Кристаллографическая ячейка относится к $C 2 2 2 1$ (ромбическая). Ее параметры приведены в Таблице 1.

Таблица 1

Параметры кристаллографической ячейки

Длина (Å)	Угол (°)
a = 98.05	$\alpha = 90.00$
b = 105.42	$\beta = 90.00$
c = 219.26	$\gamma = 90.00$

Индикаторы качества модели

Представленная ниже информация была получены из [1] и [3].

- R-фактор: 0.173
- R-free: 0.223

Хорошими значения принято считать R-фактор < 0.25 и R-free < 0.20 . Соответственно, R-фактор говорит о хорошем качестве модели, а R-free говорит о среднем качестве модели.

Разница между значениями R-фактора и R-free составляет 5% (меньше, чем 5,3%), что соответствует хорошему качеству модели.

RSR (real-space R-фактор) показывает, насколько построенная по модели электронная плотность сочетается с «экспериментальной». Его значение составляет 0.111 при стандартном отклонении 0.043. Хорошими значениями принято считать $RSR < 10\%$, плохими – $RSR > 20\%$. Значение RSR для исследуемой модели говорит о ее среднем качестве (с натяжкой можно сказать, что структура имеет хорошее качество).

По выше представленным оценкам рассматриваемая модель находится на грани хорошего качества.

Построение карты производилось с помощью MolProbity [4]. Сервис имеет достаточно широкий спектр операций. В частности, получая на вход файл в формате `pdb`, он выдает карту Рамачандрана, которая предназначена для отображения торсионных углов (ψ и ϕ) полипептидной цепи (для удобства разбито на 6 групп). Карта для исследуемой структуры представлена на Рис. 3.

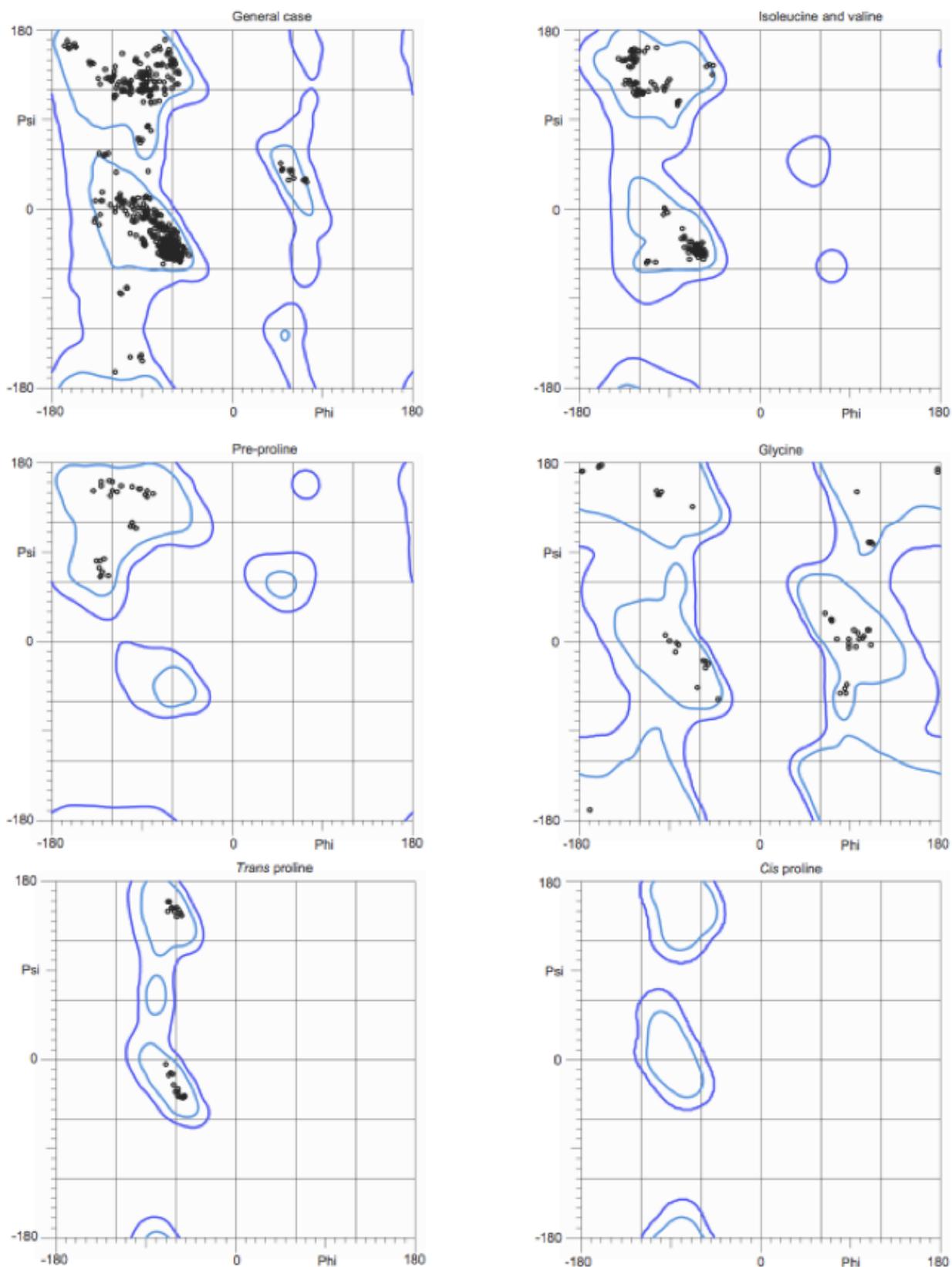


Рис. 3. Карта Рамачандрана для белка TenI

98% аминокислотных остатков находятся в предпочитаемой области, а 100% – в допустимой. Таким образом в рассматриваемой модели нет остатков, которые находятся в запрещенной области и, соответственно, имеют сильно отклоняющиеся конформации.

Кроме того MolProbity представляет рассчитанные критерии качества модели (Таблица 2).

Таблица 2

Критерии качества модели (с сервиса MolProbity)

Protein Geometry	Poor rotamers	14	2.19%	Goal: <0.3%
	Favored rotamers	596	93.12%	Goal: >98%
	Ramachandran outliers	0	0.00%	Goal: <0.05%
	Ramachandran favored	752	97.28%	Goal: >98%
	MolProbity score [^]	1.44		99 th percentile* (N=10112, 2.23Å ± 0.25Å)
	Cβ deviations >0.25Å	0	0.00%	Goal: 0
	Bad bonds:	0 / 6174	0.00%	Goal: 0%
	Bad angles:	0 / 8326	0.00%	Goal: <0.1%
Peptide Omegas	Cis Prolines:	0 / 28	0.00%	Expected: ≤1 per chain, or ≤5%
	Cis nonProlines:	5 / 747	0.67%	Goal: <0.05%

Согласно данным, представленным в Таблице 2, в модели нет Ramachandran outliers – абсолютно маргинальных остатков по карте Рамачандрана, которые лежат вне допустимой области (что мы и видим по Рис.3). А вот Poor rotamers (число остатков с маргинальными по отклонению от ротамеров боковыми цепями) составляет 2.19%, что значительно превышает допустимое значение 0.3%.

Большинство связанных с геометрией белка параметров лежат в области хороших значений.

В пользу модели также говорит параметр MolProbity score – интегральная оценка структуры по данным этого сервиса, которая объединяет ClashScore, оценку ротамеров и карты Рамачандрана, нормализуя их на одинаковое разрешение. Он равен 1.44, что является очень хорошим результатом. Еще один параметр ClashScore отображает число недопустимых наложений атомов на 1000 (т.е. 5 соответствует 5%), перцентиль рассматриваемой структуры по отношению к другим структурам с примерно таким же разрешением. Согласно ClashScore исследуемая модель является одной из лучших с разрешением 2.23Å ± 0.25Å.

Z-score и температурный фактор

Z-score RSR - это показатель, который оценивает RSR относительно других структур. Он вычисляется сравнением RSR остатка со средним RSR остатков такого же типа из других структур с примерно тем же разрешением. Z-score для каждой цепи изучаемой модели (Рис. 4) получен с помощью сервиса EDS [3].

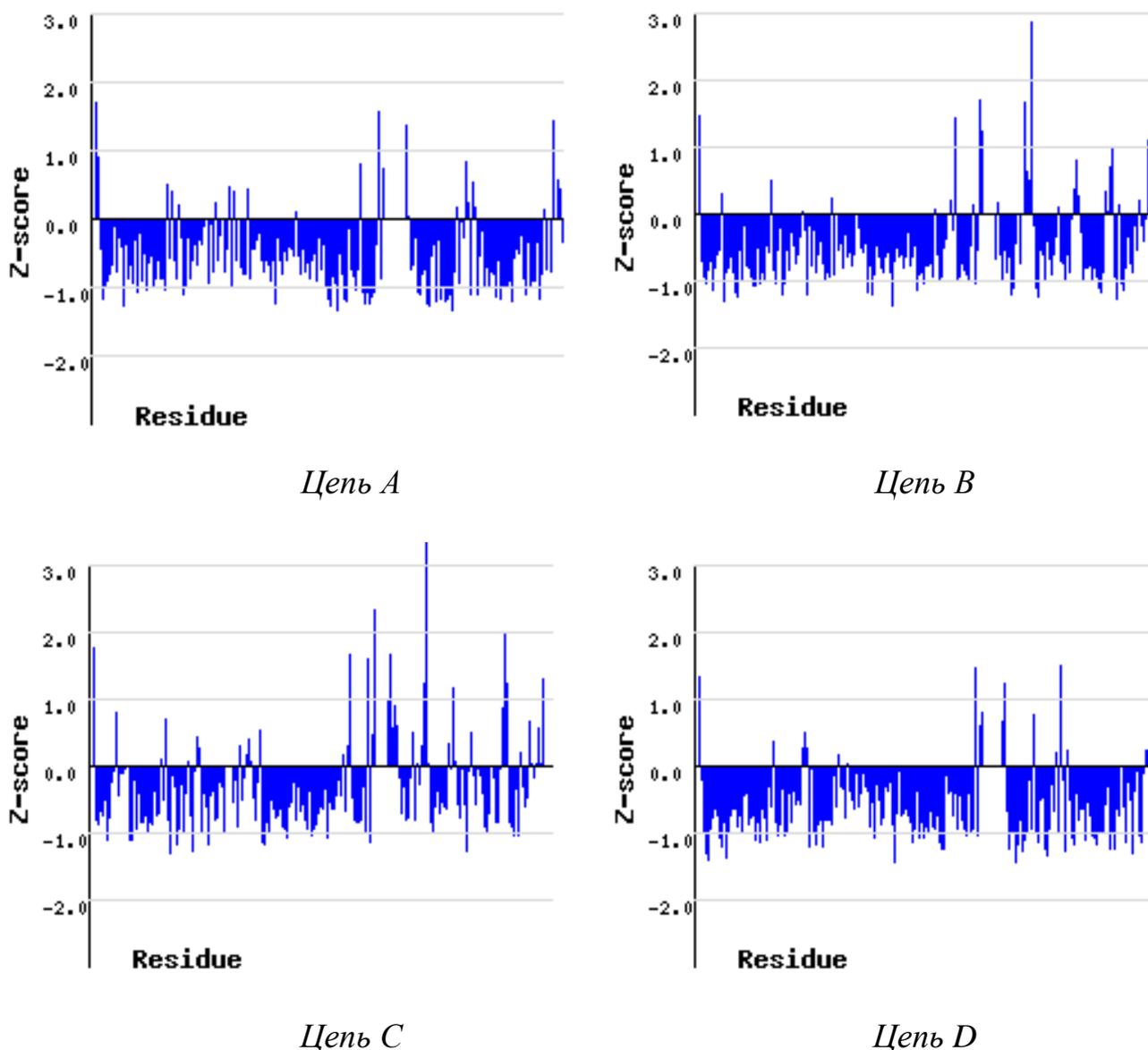


Рис. 4. Z-score аминокислотных остатков для структуры 3QH2

Из Рис. 4 видно, что маргинальные остатки по Z-score присутствуют в цепи B (для Ser148 Z-score = 2.86) и цепи C (для Ser148 Z-score = 3.68, для Glu125 Z-score = 2.33). В цепи C для Ala183 Z-score = 1.95, что является достаточно плохим значением.

В цепи A Z-score аминокислотных остатков не превышает значения 1.6 (для His122 Z-score = 1.56).

В цепи D Z-score аминокислотных остатков не превышает значения 1.5 (для Asp160 Z-score = 1.49).

Температурный фактор (B-фактор) – характеристика меры динамической неупорядоченности кристалла, которая может быть обусловлена различными факторами (тепловое движение, отличия в расположении молекул белка в кристалле).

Сервис позволяет получить значения B-фактора (например, Рис. 5).

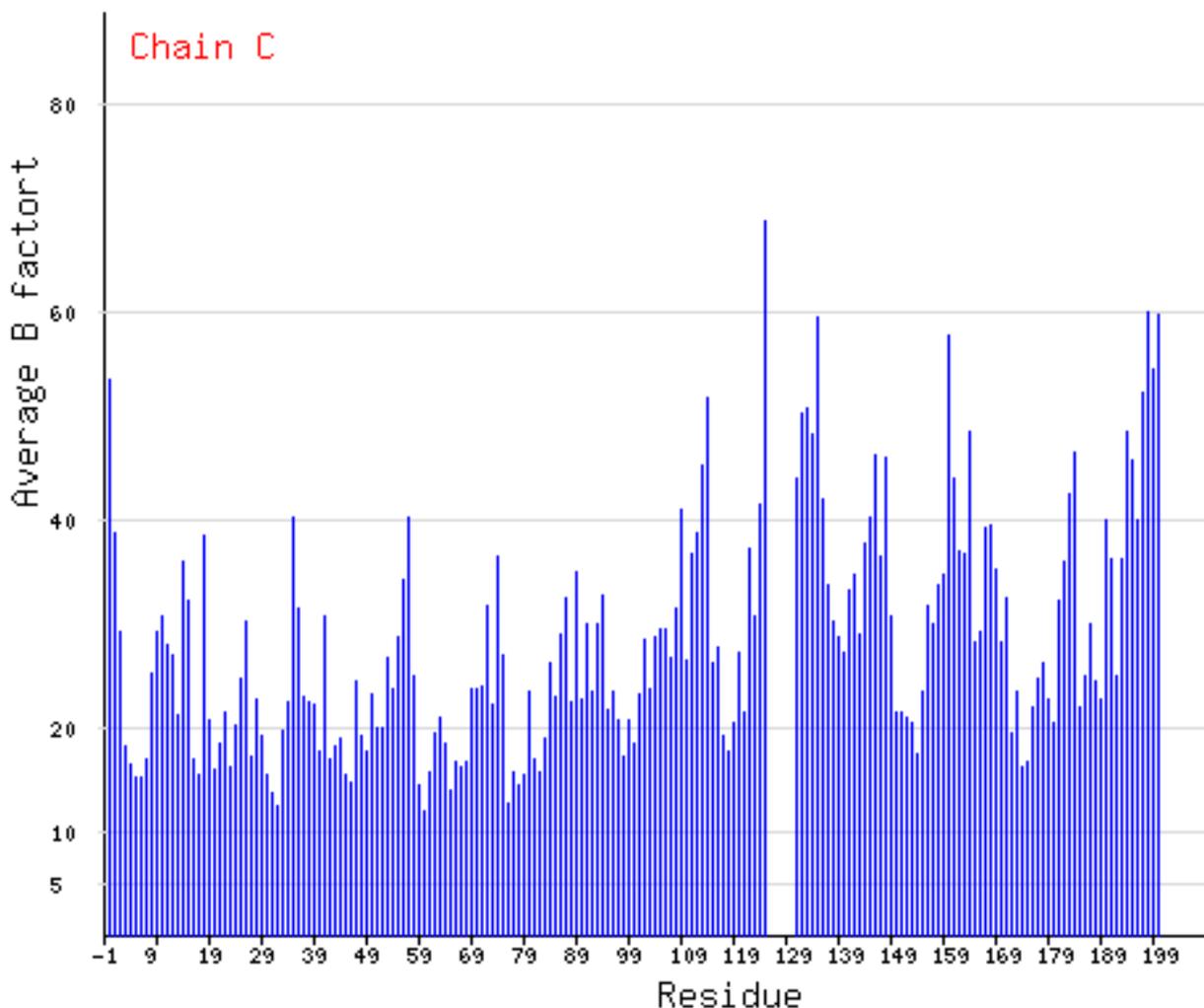


Рис. 5. B-фактор для цепи C структуры 3QH2

Наибольшее значение B-фактора в цепи А было получено для Glu202 (B-фактор = 61.4). Для других остатков значения не превышали 60.

Наибольшее значение B-фактора в цепи В было получено для Agr200 (B-фактор = 67.4). Высокие значения B-фактор имели Glu198 и Met199 (60.8 и 61.3 соответственно).

Наибольшее значение B-фактора в цепи С было получено для Glu125 (B-фактор = 68.6). Высокие значения также имели Arg135 (B-фактор = 59.4), Asp160 (B-фактор = 57.8), Glu198 (B-фактор = 60.1), Arg200 (B-фактор = 59.7).

Маргинальные остатки и их анализ

Для анализа и детального изучения маргинальных остатков использовались [3], [5].

Наибольшее подозрение вызывают маргинальные остатки по RSR и Z-score.

Таблица 3

Маргинальные остатки в структуре ЗОН2, выбранные для анализа

Маргинальный остаток	Критерий
Ser148	Z-score = 3.68 RSR > 20% (30.3%) Странное окружение
Glu125	Z-score = 2.33 RSR > 20% (29.7%)
Glu198	RSR > 20% (27.8%) Один из самых высоких температурных факторов
Arg200	RSR > 20% (24.0%) Странное окружение
Ala183	Z-score = 1.95 RSR > 20% (23.6%)

Все изображения ЭП выбранных для анализа маргинальных остатков (Рис. 6 – Рис. 9) выполнены при 1.5σ .

Для изображенного на Рис. 6 Ser148 завышены показатели. Однако видно, что он хорошо вписывается в ЭП. Маргинальность не является результатом недостаточных экспериментальных данных.

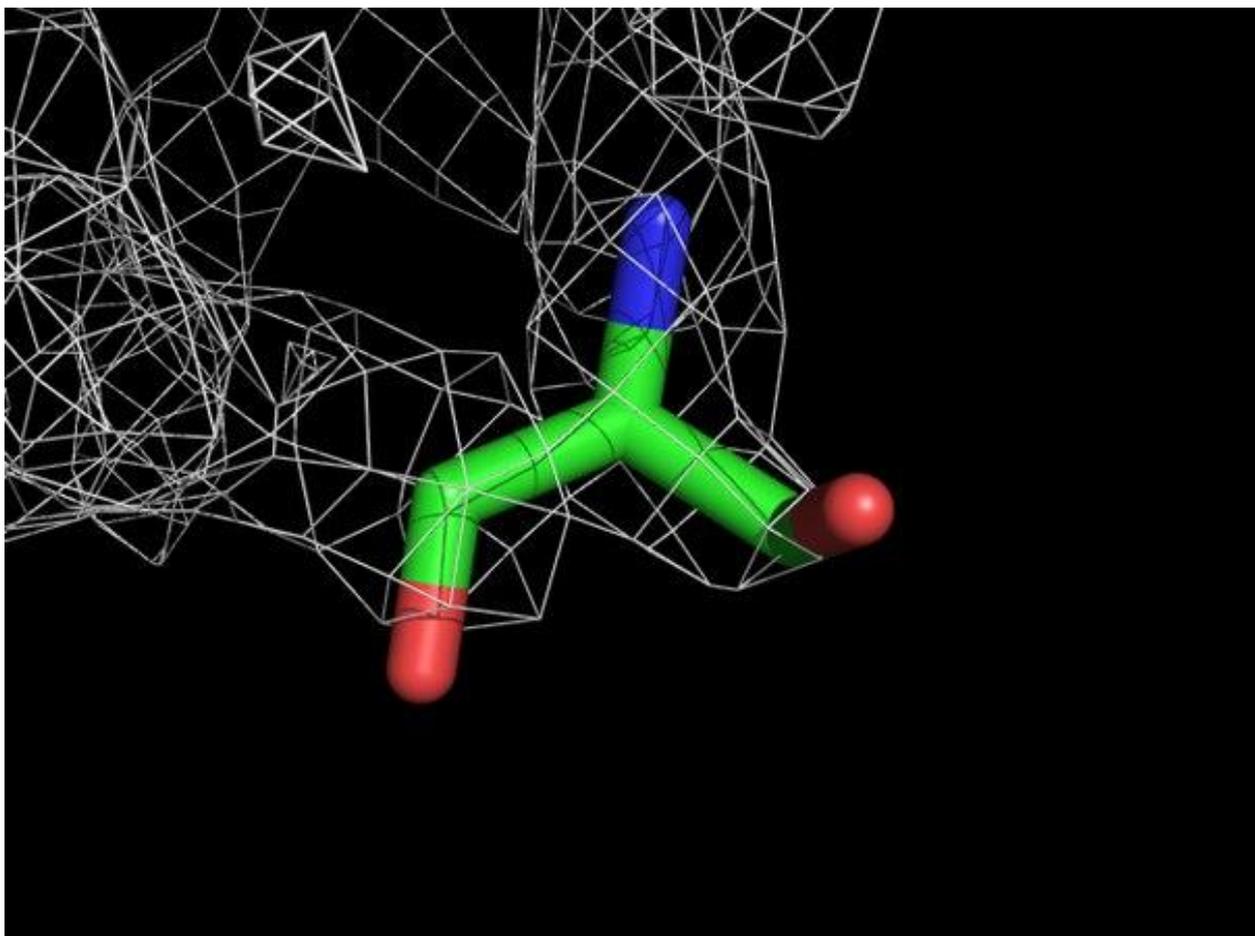


Рис. 6. Изображение ЭП для Ser148

Для Glu125 (Рис.7) характерен высокий температурный фактор (68.8). немного завышен . А также видно, что остаток не очень хорошо вписывается в ЭП. Скорее всего маргинальность этого остатка – результат ошибки расшифровки.

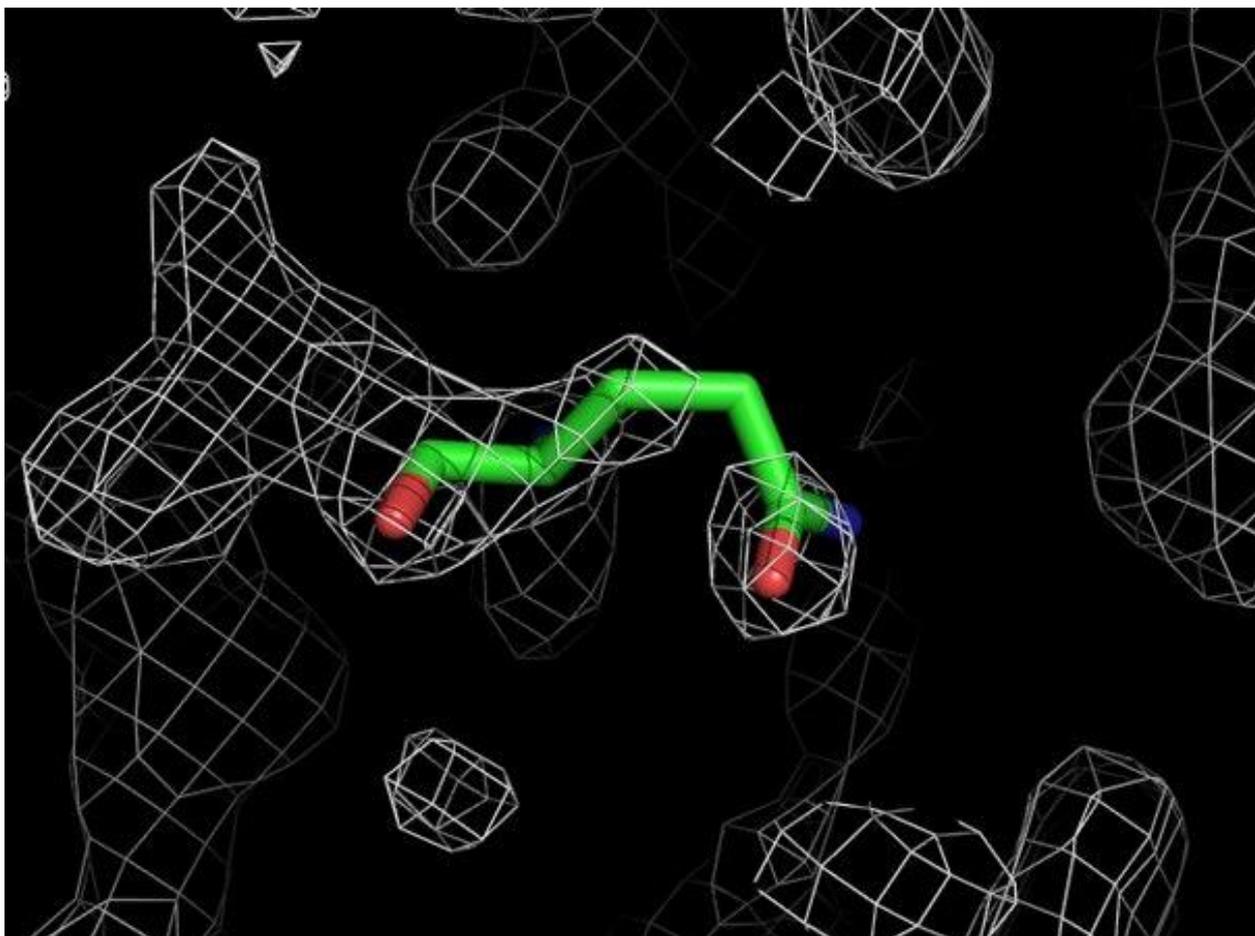


Рис. 7. Изображение ЭП для Glu125

RSR для Glu198 равен 27,8%. На Рис. 9 видно, что он плохо вписывается в ЭП. Также он имеет достаточно высокий температурный фактор. Однако показатель Z-score у него в норме. Маргинальность остатка является результатом недостаточной точности расшифровки координат атомов.

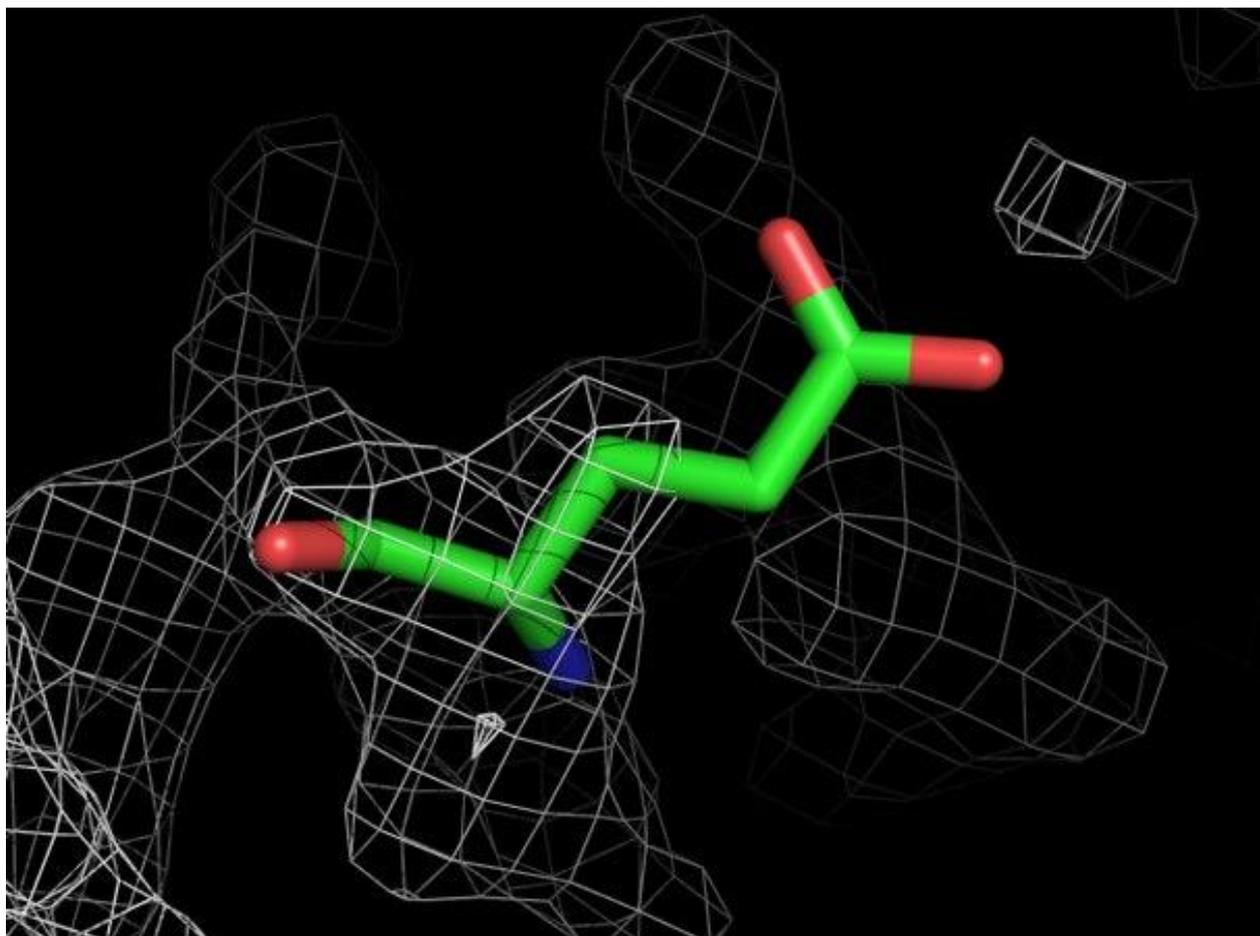


Рис. 8. Изображение ЭП для Glu198

Для Arg200 (Рис. 9) завышен RSR, но Z-score в норме. Температурный фактор имеет не высокое значение. Кроме того, в PDBREPORT для Agr200 была обнаружена возможная инверсия в названиях концевых групп боковой цепи (N-H-1 вместо N-H-2). Маргинальность остатка является результатом недостаточной точности расшифровки координат атомов.

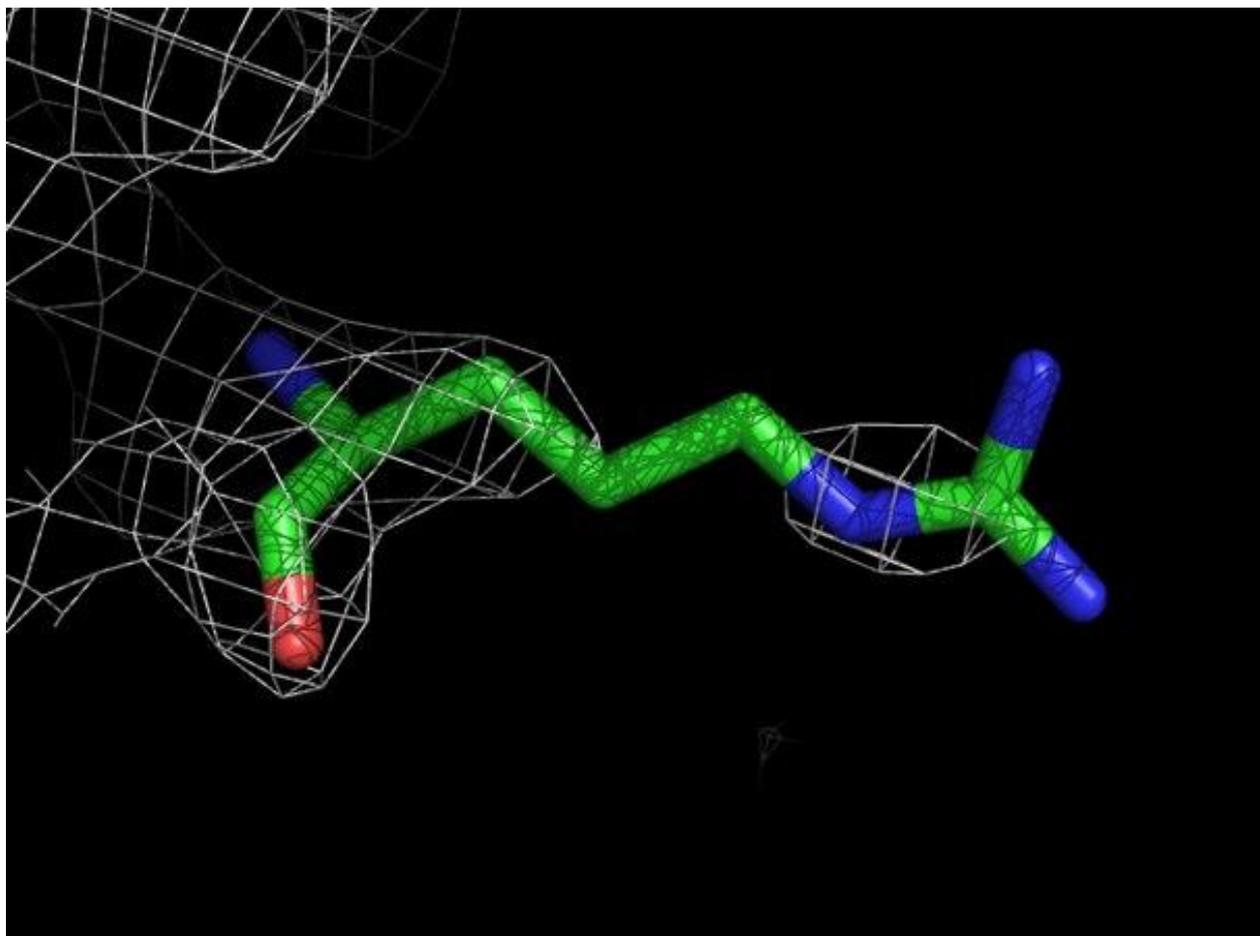


Рис. 9. Изображение ЭП для Arg200

Сравнение моделей из PDB и PDB_redo

PDB_redo [6] – база данных, которая содержит более полную и обновленную информацию о моделях структур, полученных с помощью PCA. Оптимизация структуры основана на экспериментальных данных, которые представлены в PDB-файле.

Было проведено сравнение 3QH2 с оптимизированной моделью из PDB_redo (Таблица 4).

Таблица 4

Сравнение некоторых показателей качества структуры 3QH2

показатель	Для данных из файла pdb	Для построенной по данным модели	Для модели после полной оптимизации
R	0.1730	0.1748	0.1763
R-free	0.2230	0.2193	0.1922
Соответствие ротамерам	-	-1.863	-0.749

Очевидно, что все показатели приняли лучшее значение (все параметры, не только представленные в Таблице 2). Это говорит о большем соответствии модели экспериментальным данным.

Заключение

Рассмотренная модель 3QH2 имеет достаточно хорошее качество разрешения. Общепринятые критерии качества в диапазон хороших значений (хотя некоторые находятся на грани между хорошими и средними значениями). На карте Рамачандрана нет остатков вне предпочитаемой зоны. При оптимизации с помощью PDB_redo показатели улучшаются.

Ссылки на БД и серверы

1. PDB - <http://www.rcsb.org/pdb/home/home.do> (для 3QH2 - <http://www.rcsb.org/pdb/explore/explore.do?structureId=3QH2>).
2. PubMed - <http://www.ncbi.nlm.nih.gov/pubmed>.
3. EDS - <http://eds.bmc.uu.se/eds/index.html> (для 3QH2 - <http://eds.bmc.uu.se/cgi-bin/eds/uusfs?pdbCode=3qh2>).
4. MolProbity - <http://molprobity.biochem.duke.edu/>
5. PDBREPORT - <http://swift.cmbi.ru.nl/gv/pdbreport/>
6. PDB_redo - http://www.cmbi.ru.nl/pdb_redo/ (для 3QH2 - http://www.cmbi.ru.nl/pdb_redo/qh/3qh2/index.html)

Список литературы

7. Jurgenson CT, Begley TP, Ealick SE, «*The structural and biochemical foundations of thiamin biosynthesis*», Annu Rev Biochem. 2009;78:569-603. doi: 10.1146/annurev.biochem.78.072407.102340.
8. Hazra AB, Han Y, Chatterjee A, Zhang Y, Lai RY, Ealick SE, Begley TP, «*A missing enzyme in thiamin thiazole biosynthesis: identification of TenI as a thiazole tautomerase*», J Am Chem Soc. 2011 June 22; 133(24): 9311–9319. doi:10.1021/ja1110514.
9. Pang AS, Nathoo S, Wong SL, «*Cloning and characterization of a pair of novel genes that regulate production of extracellular enzymes in Bacillus subtilis*», J Bacteriol. 1991 Jan;173(1):46-54.
10. Toms AV, Haas AL, Park JH, Begley TP, Ealick SE, «*Structural characterization of the regulatory proteins TenA and TenI from Bacillus subtilis and identification of TenA as a thiaminase II*», Biochemistry. 2005 Feb 22;44(7):2319-29.
11. Toms AV, Haas AL, Park JH, Begley TP, Ealick SE, «*Structural characterization of the regulatory proteins TenA and TenI from Bacillus subtilis and identification of TenA as a thiaminase II*», Biochemistry. 2005 Feb 22;44(7):2319-29.

