

Отчёт по качеству PCA расшифровки структуры 4m9a

Автор: Шафиков Радик

Оглавление

Аннотация	2
Введение	2
Результаты и обсуждение	2
1. Общая информация о модели	2
2. Индикаторы качества модели	3
3. Примеры маргинальных остатков	6
4. Анализ маргинальных остатков и гетеромолекул	6
5. Сравнение модели из PDB с моделью из PDB_redo	8
Заключение	10
Приложение	11
Список литературы и сервисов	15

Аннотация

Данный отчет представляет собой анализ качества расшифровки структуры 4m9a на основе определенных показателей, таких как параметры кристаллической решетки, R-фактор, R_free, карта Рамачандрана, и др., которые могут быть получены с помощью различных сервисов.

Введение

Ацил-СоА дегидрогеназа [EC:1.3.8.7] – фермент класса оксидоредуктаз, катализирующий начальный шаг в каждом цикле бета-окисления жирных кислот в митохондриях клеток. Ее действие приводит к введению транс двойной связи между C2 и C3 атомами ac-CoA. Фермент является гомотетрамером. Пространственная структура фермента представлена на рис.1.

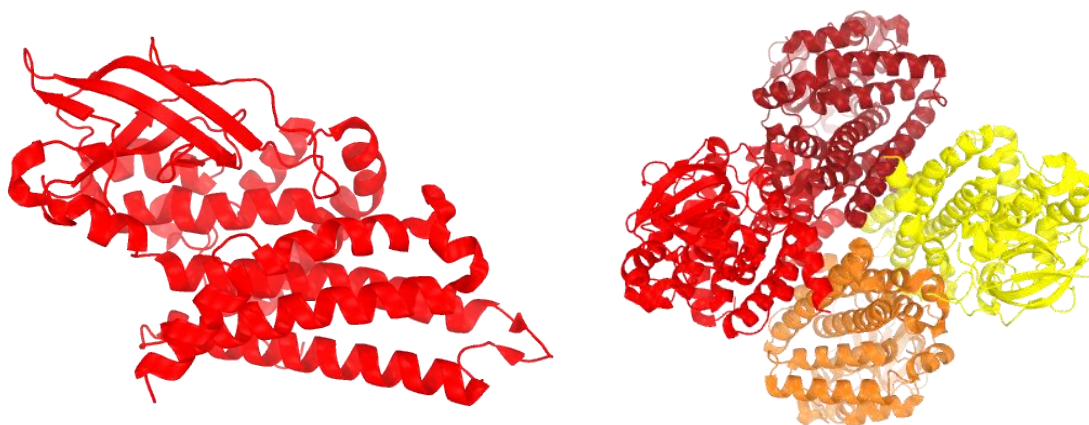


Рис.1. Структура моно- и тетрамера фермента

О том, зачем эта структура была получена, неизвестно, т.к. статья не опубликована, однако можно предположить, что это связано с инфекционными заболеваниями, которые могут быть вызваны *Burkholderia thailandensis* и ее «родственниками».

Результаты и обсуждение

1. Общая информация о модели

Фермент закристаллизован в виде гомотетрамера, связанного с FDA – аналогом FAD, служащий ингибитором; присутствуют молекулы воды. Авторы работы - Lukacs, С.М., Fairman, J.W., Edwards, Т.Е., Lorimer, D. из центра структурной геномики для инфекционных заболеваний, Сиэтл. Для решения фазовой проблемы был применен метод молекулярного замещения, используя данные из структуры 1UKW (информация из файла структурных факторов). Ниже рассмотрим некоторые характеристики модели в целом, которые можно узнать на сайте EC [1] и непосредственно из файла структурных факторов, который можно найти на сайте RSCB PDB [2].

Общее количество рефлексов – «пятен» разной интенсивности – 84655. Таково же количество измеренных структурных факторов. Число факторов, использованных для оптимизации модели, составляет 79684. Разрешение данных, по которым построена модель модели, достаточно хорошее - 2.20 Å (смоделированы и молекулы воды), при этом диапазон разрешений структурных факторов равен 46.83 - 2.20 Å.

Полнота набора структурных факторов составляет 99.2%. Самым распространенным типом недочетов является отсутствие измеренного модуля структурного фактора, однако также встречаются и пропуски на месте структурных факторов (см. пр.4) [3].

Теперь перейдем к описанию параметров кристаллографической ячейки. Файл PDB имеет запись о кристаллографических характеристиках в поле CRYST1, которое для структуры 4m9a выглядит следующим образом:

CRYST1 87.260 111.000 170.250 90.00 90.00 90.00 P 21 21 21 16

Здесь первые три значения являются длинами направляющих векторов кристалла, следующие три - углами между ними, далее идут название кристаллографической группы (P 21 21 21) и число молекул в ячейке (16). Разрешены следующие операции симметрии:

X,Y,Z	-X+1/2,-Y,Z+1/2	-X,Y+1/2,-Z+1/2	X+1/2,-Y+1/2,-Z
-------	-----------------	-----------------	-----------------

Асимметрической единицей является мономер; тетрамер может быть достроен применением операторов некристаллографической симметрии – он и находится внутри элементарной ячейки [4] и представлен на рис.2, где молекулы соседних ячеек восстановлены с помощью команды sutexr в PyMol [5].

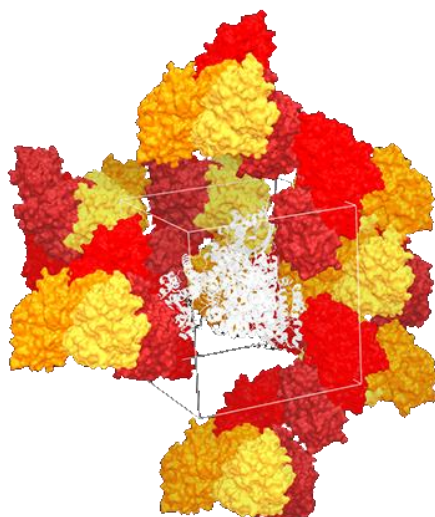


Рис.2. Расположение тетрамера внутри элементарной ячейки в окружении восстановленных командой sutexr молекул

2. Индикаторы качества модели

Теперь постараемся оценить модель в целом по некоторым индикаторам. Основные индикаторы приведены на сайте PDB в виде графа (см.рис.3), а более полный анализ структуры можно увидеть, кликнув Full report рядом с графом [6].

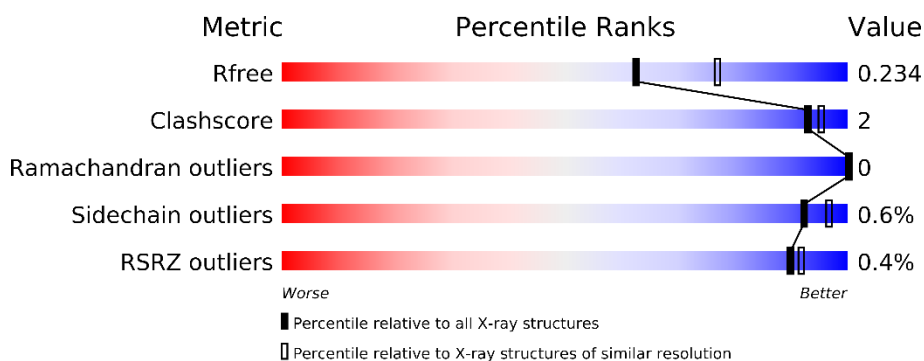


Рис.3. Граф с характеристиками структуры

Давайте же разберем, о чем все эти характеристики говорят. R-фактор характеризует соответствие модели экспериментальным данным – структурным факторам. Рассчитывается он следующим образом:

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

, где $F^{calc}(h,k,l)$ – рассчитанные модули структурных факторов, $F^{obs}(h,k,l)$ – измеренные в эксперименте. При оптимизации модели он минимизируется: чем меньше R-фактор, тем модель лучше. Для данной модели он равен 0.186. Однако более точно положение дел отражает R_free, который рассчитывается тем же образом, но только по контрольным рефлексам. Его значение составляет 0.234, что говорит о возможной переоптимизации модели. Граф из рис.3 показывает, что данное значение достаточно среднее как среди всех структур pdb, так и среди структур с похожим разрешением (чуть лучше, чем среднее).

С помощью сервиса MolProbity [7] отыщем иные характеристики.

Clashscore отображает число неблагоприятных стерических перекрываний атомов >0.4 Å на 1000 атомов и равняется 2. Всего таких перекрываний внутри асимметрической единицы 47 [6] [7], ни одно из них не связано с преобразованиями симметрии (включены также His, Asn, Gln с инверсиями боковой цепи). По крайней мере, нет ни одного атома, расположенного к оси симметрии ближе 0.77 Å [8]. Рассчитанный в MolProbity перцентиль равен 100th – один из лучших среди структур со сходным разрешением, что видно и на рис.3.

Далее, на рис.4 представлены карты Рамачандрана для всех аминокислотных остатков, а также отдельно для некоторых из них (рис.1 приложения). 97.0% (1432/1477, в идеале $>98\%$) расположены в благоприятных областях, 100.0% (1477/1477) – в разрешенных. Выбросов нет. Число остатков с маргинальными по отклонению от ротамеров боковыми цепями (poor rotamers) равно 9 (0.84%, в идеале $<0.3\%$); число и процент ротамеров в предпочитаемой области - 95.08% (в идеале $>98\%$).

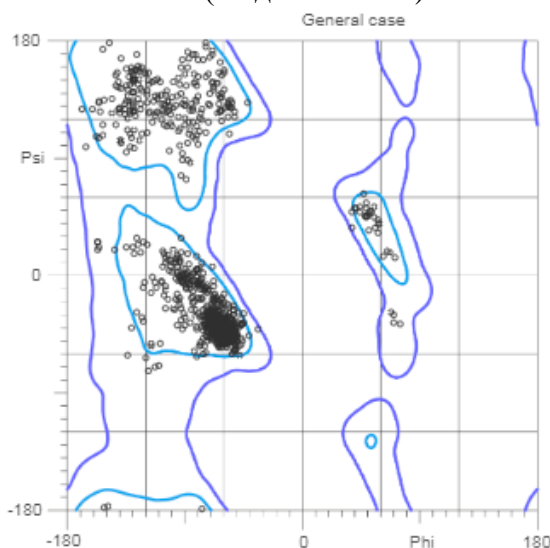


Рис.4. Общая карта Рамачандрана

C_{β} deviations > 0.25 Å - число C_{β} атомов с неприемлемым отклонением от ожидаемого положения – равняется 1 (0.07%, в идеале отсутствуют), число плохих валентных углов, существенно отклоняющихся от теории, равно 2 (0.01%), а число плохих валентных связей равняется 1 (0.01%). В отчете на сайте PDB представлено 2 отклоняющиеся валентные связи и 9 некорректных валентных углов.

Пространственный R-фактор позволяет оценить соответствие модели группы атомов “экспериментальной” электронной плотности. Рассчитывается следующим образом:

$$RSR = \frac{\sum_{A \in L} |\rho_{\text{эксп}} - \rho_{\text{модель}}|}{\sum_{A \in L} \rho_{\text{эксп}}} [\cdot 100\%]$$

Ниже на рис. 5, полученном в EDS [9], приведен RSR для остатков цепи A, а в приложении (рис.2) – для всех цепей.

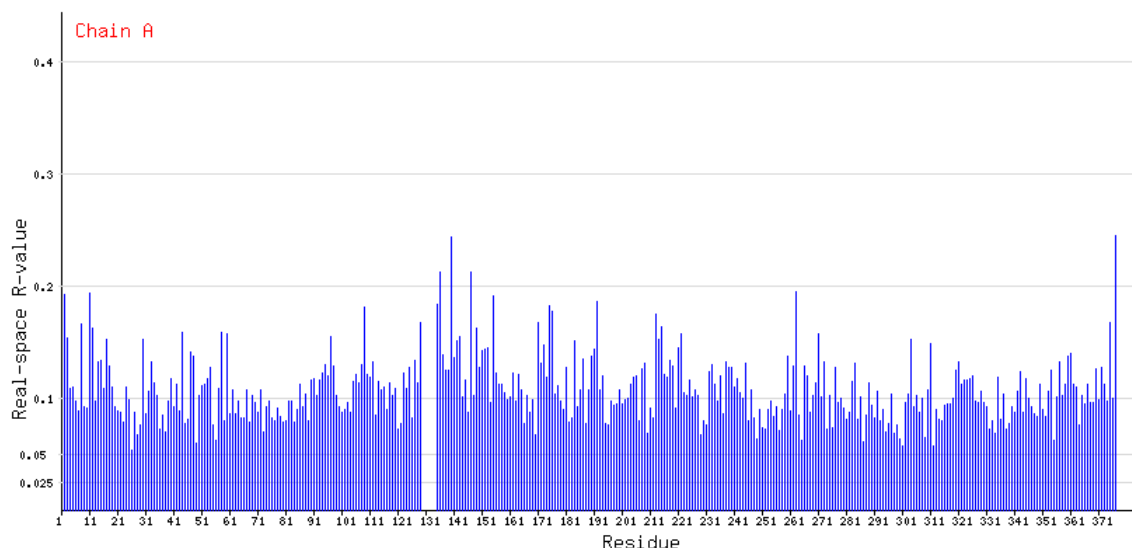


Рис.5. Значения RSR для остатков цепи A структуры 4т9а

Как видно из диаграммы, в модели немного остатков с RSR>20%; больше, чем у половины остатков, RSR меньше 10% - в зоне «хороших» значений. Остатки с большим RSR также можно выявить в составе significant regions, где отображаются наиболее «рискованные» остатки (см. рис.6+приложение, рис.3).

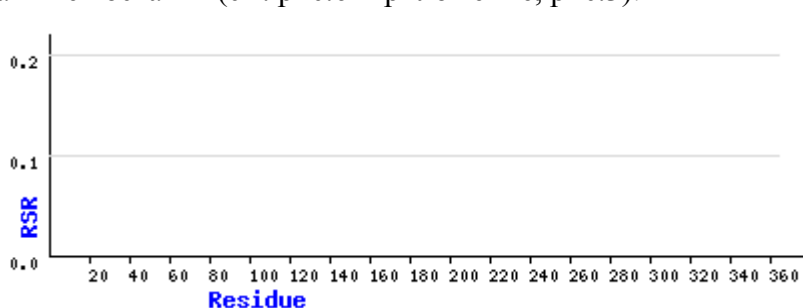


Рис.6. Significant regions цепи A

RSR-Z – это относительная оценка RSR для отдельных остатков. Для вычисления Z остатка (напр. Ala57) его RSR сравнивается со средним RSR для того же типа остатков (Ala) по выборке из PDB с примерно таким же разрешением (напр. 1.5-1.8 Å).

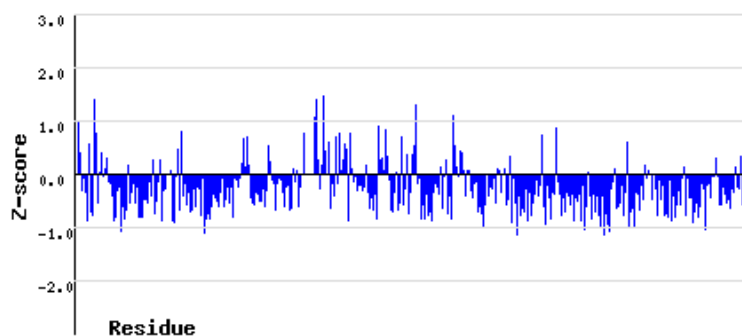


Рис.7. Значения RSR-Z для остатков цепи A (остальные цепи см. рис.4 приложения)

Высокие положительные значения $Z > 2$ у одного остатка цепи A свидетельствуют о том, что он плохо вписан в электронную плотность по сравнению с другими структурами с тем же разрешением, поэтому его можно рассматривать как маргинальный.

Затем с помощью сайта WHAT_CHECK [8] опцией Fine Packing Quality Control была оценена комфортность окружения для всех остатков. Не найдено ни одного остатка с параметром меньше -5, что соответствовало бы маргиналам по окружению. Есть 4 остатка со значением меньше -3, следовательно, их можно считать достаточно сомнительными и рассмотреть позже. Среднее значение Z-score = -0.97, что соответствует хорошей структуре. Все оценочные суждения основаны на инструкции в конце выдачи программы.

Белок не содержит нестандартных аминокислот [6], все остатки, которые должны быть планарными, таковыми и являются [8]. Также были проверены искажения Pro по углу phi: несколько остатков оказались подозрительными, однако создатели программы предупреждают, что данный показатель имеет низкий доверительный интервал [8].

3. Примеры маргинальных остатков

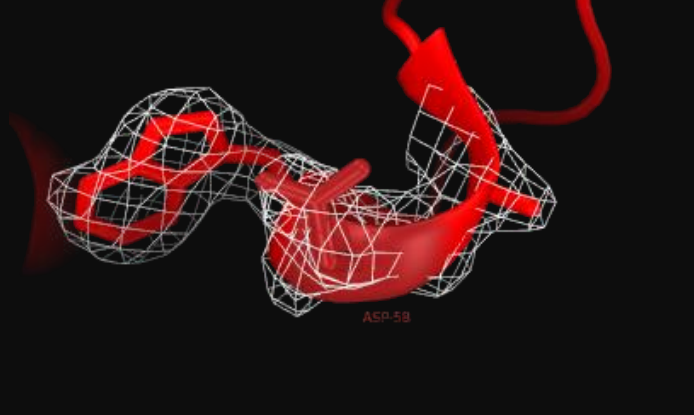
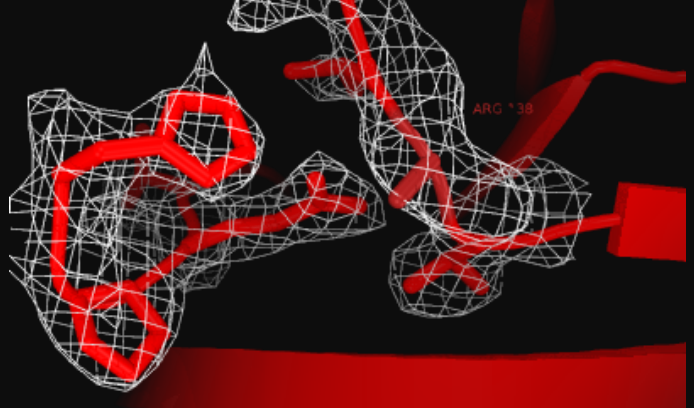
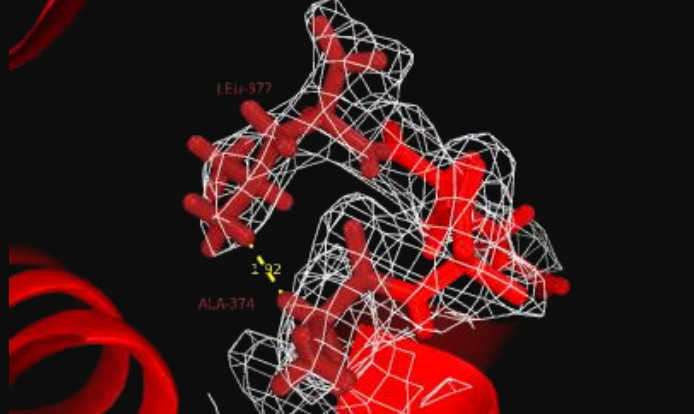
Теперь рассмотрим конкретные примеры маргинальных остатков. Данные взяты из Validation report, выдачи MolProbity и WHAT_CHECK.

#	Остаток	Причина маргинальности
1	Asp 58 (B)	RSR-Z = 3.02, что >2, следовательно, остаток плохо вписан в электронную плотность. Является «выбросом» по валентным углам (см. Validation report, PDB)
2	Arg 138 (A)	Некомфортное окружение: параметр < -3. Значение RSR = 0,244 (>0.2)
3	Leu 377 (A,B)	RSR-Z = 0.331, RSR = 0.245. Перекрытие 0.502 Å с Ala 374 (B)
4	Asn 136 (A)	RSR = 0.212. Отклоняются углы chi: 225.5, 282.5 – poor rotamer
5	Ile 275 (A)	Имеет перекрытие 0.541 Å с His278. Находится в significant regions
6	Tyr 362 (все цепи)	Poor rotamer (отклонение Cβ на 0.28 Å), образует плохой угол CA-CB-CG: 4.1 σ
7	Met 326 (D)	Отброс по углу CG-SD-CE: 5.1 σ. Перекрытие 0,409 Å с Gln 291 (D)
8	Glu 367 (C)	Выброс по длине связи: CD--OE1: 4.6 σ
9	Met 11 (C)	Отклоняются углы chi: 188, 242.1, 298 – poor rotamer
10	Ile 219 (D)	Отклоняются углы chi: 68.5, 201.9 – poor rotamer
11	Glu 222 (D)	Отклоняются углы chi: 46.9, 271.9, 97.8 – poor rotamer
12	Gln 155 (A)	Отклоняются углы chi: 190.1, 91.1, 118.6 – poor rotamer. Перекрытие 0.528 Å с Tyr 341 (B). Скорее всего, присутствует flip.

Таблица 1. Примеры маргинальных остатков

4. Анализ маргинальных остатков и гетеромолекул

Рассмотрим некоторые маргинальные остатки из Таблицы 1 подробнее, отдавая предпочтение тем остаткам, что являются «отбросами» по нескольким критериям. Возьмем Asp-58(B), Arg-138(A), Leu-377(A, B), Tyr-362(одна из цепей) и Gln-155(A), и объединим их в одну сводную таблицу.

Остаток	Показатели маргинальности	Иллюстрация
Asp-58(B)	<p>RSR-Z = 3.02, что >2, следовательно, остаток плохо вписан в электронную плотность.</p> <p>Является «выбросом» по валентным углам (см. Validation report, PDB)</p>	
<p>Вывод: действительно, значение RSR-Z и полученное изображение характеризуют то, что остаток плохо вписан в электронную плотность (изображение на уровне $\sigma=1.0$). Скорее всего, это просто ошибка расшифровки.</p>		
Arg-138(A)	<p>Некомфортное окружение: параметр < -3. Значение RSR = 0,244 (>0.2)</p>	
<p>Вывод: на изображении действительно представлен Arg, вернее, его обрубок, который, к слову, будет достроен в PDB_redo. Видно, что электронная плотность (изображение на уровне $\sigma=1.0$) в месте, где он должен находиться, отсутствует. Это также можно отнести к ошибке расшифровки – неправильный расчет функции электронной плотности на этом участке, т.к. в PDB_redo, который тоже использует экспериментальные данные, ЭТОТ остаток восстановлен.</p>		
Leu-377(B)	<p>RSR-Z = 0.331, RSR = 0.245. Перекрывание 0.502 Å с Ala 374 (B) атомами Н.</p>	
<p>Вывод: на изображении видно, что перекрывание с Ala действительно есть. Пространственный R-фактор слишком большой, однако для всех подобных остатков в моделях сходного разрешения значение RSR-Z также очень велико. Следовательно, остаток расшифрован плохо, и даже хуже, чем в других подобных структурах. Еще одна ошибка дешифровки.</p> <p>Изображение на уровне $\sigma=1.0$</p>		

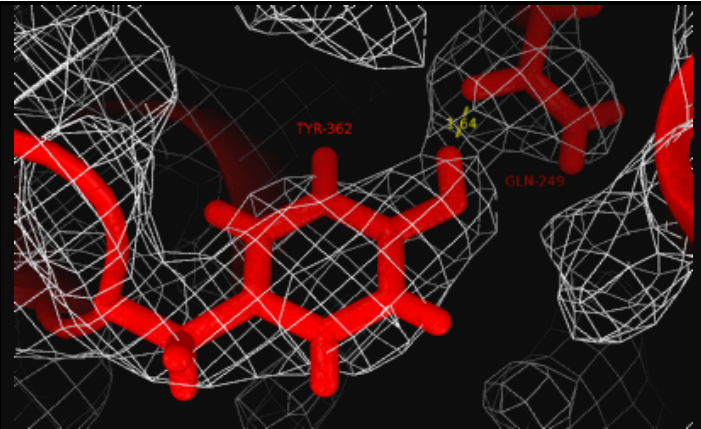
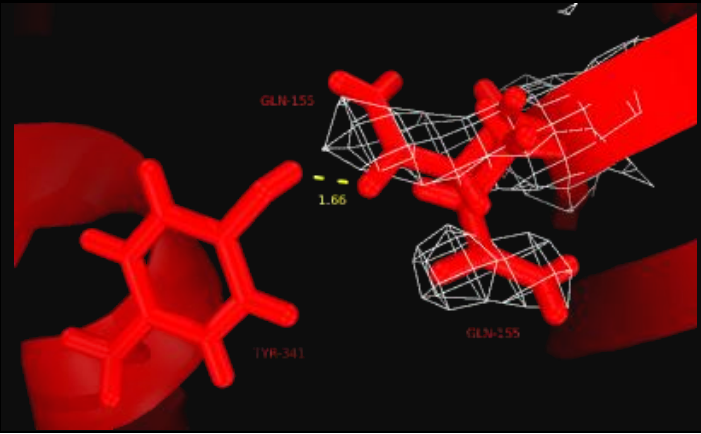
Tyr-362(A)	Poor rotamer (отклонение C β на 0.28 Å), образует плохой угол CA-CB-CG: 4.1 σ	
<p>Вывод: интересный случай; несмотря на искажение угла, данное положение радикала скорее всего выгодно из-за наличия водородной связи с Gln-248(A). Можно считать, что подобная конформация является особенностью. Изображение на уровне $\sigma=1.0$</p>		
Gln-155(A)	Отклоняются углы chi: 190.1, 91.1, 118.6– poor rotamer. Перекрывание 0.528 Å с Tyr 341 (B). Скорее всего присутствует flip.	
<p>Вывод: здесь присутствует ошибка расшифровки - flip, так как в плохом случае Gln и Tyr-341(B) перекрываются атомами водорода, чего быть не должно. Верна «нижняя» (по рисунку) конформация, несмотря на неоднозначность карты электронной плотности (изображение на уровне $\sigma=1.0$).</p>		

Таблица 2. Характеристика маргинальных остатков. Изображения получены в PyMol, причем структура с атомами водорода получена средствами MolProbity.

5. Сравнение модели из PDB с моделью из PDB_redo

PDB_redo позволяет оптимизировать модель, используя различные параметры уточнения, выбирая между изо- и анизотропным B-фактором, перестраивая конформации ротамеров, вращая боковые цепи остатков для оптимизации сети водородных связей, пересматривая молекулы воды и конформации лигандов.



Рис.8. Консервативно- (синий) и полностью оптимизированная (красный) структура по сравнению с исходной (оранжевый), стрелками показаны некоторые видимые различия. На рис.8 стрелками показаны участки, различающиеся у исходной структуры и у оптимизированной: а – положение бета-тяжа, б/с – отсутствие тяжа как элемента вторичной структуры в оптимизированном файле. Итак, рассмотрим подробнее, какие конкретно характеристики подверглись изменению и в какую сторону.

Сначала рассмотрим общие характеристики. В переработанном файле были улучшены конформация скелета, вид карты Рамачандрана (меньше остатков в сравнительно неблагоприятных областях), оптимизированы длины и углы связей, уменьшено число столкновений (bumps/clashes) атомов. Практически неизменным осталось количество водородных связей [10].

PDB_redo дополнительно выдает огромный плот с произведенными изменениями. Значительно ухудшилось положение Leu-4(A), Arg-109(A), Met-171(A), Gln-221(A), Glu-22(C), Pro-128(C), Arg-179(C), Arg-205(C), Leu-377(C), Asn-216(D), Lys-273(D); значительно улучшилось положение Пе-361(D), Пе-12(C). Об изменениях судится по изменению пространственного коэффициента корреляции с данными эксперимента. Возможно, данный подход не совсем верен для различных «огрызков», например, для Arg-179(C) – по-моему, не стоит считать это ухудшением, так как электронная плотность на него не распространяется (рис.9А). То же самое относится и к Arg-109(A), так как карта электронной плотности на данном участке достаточно неоднозначна (рис.9В).

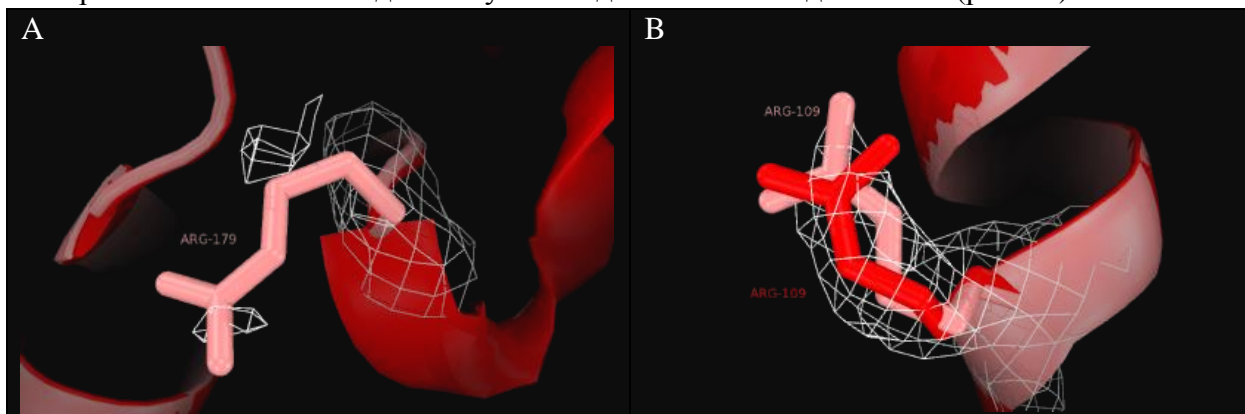


Рис.9. Изображения некоторых изменений. Розовым цветом показана переработанная структура, красным – исходная.

Несмотря на некоторое количество локальных ухудшений, модель в целом, согласно PDB_redo, улучшилась.

Заключение

Данная структура определена сравнительно неплохо относительно всех структур с похожим разрешением, о чем можно судить по R_free фактору, небольшому количеству перекрываний (clashes), RSR-Z для подавляющего большинства остатков. На карте Рамачандрана нет «ярких» выбросов в запрещенных областях. Кроме того, в PDB_redo в структуры было внесено достаточно небольшое число изменений, что также может судить о ее качестве и об отсутствии грубых дефектов структуры.

Исходя из изображения электронной плотности как целого белка [11], так и отдельных аминокислот, можно сделать вывод о том, что разрешения структуры 4M9A недостаточно для точного определения положения отдельных атомов, однако полипептидная цепь и положение большинства радикалов аминокислотных остатков уже достаточно хорошо видны для восстановления трехмерной структуры белка с учетом его последовательности. Однако не обошлось и без ошибок – часть остатков просто «обрезаны» и были восстановлены только с помощью PDB_redo, встречались Gln с неправильной конформацией, а также прочие остатки, атомы которых так или иначе слишком сближались с атомами остатков-соседей (clashes).

В целом, основываясь на вышесказанном, модель можно считать достаточно хорошей, несмотря на некоторые огрехи в виде несоответствия координат атомов карте электронной плотности и т.д., так как их немного и после оптимизации, например, в PDB_redo, они могут быть устранены.

Приложение

Рис.1. Карты Рамачандрана

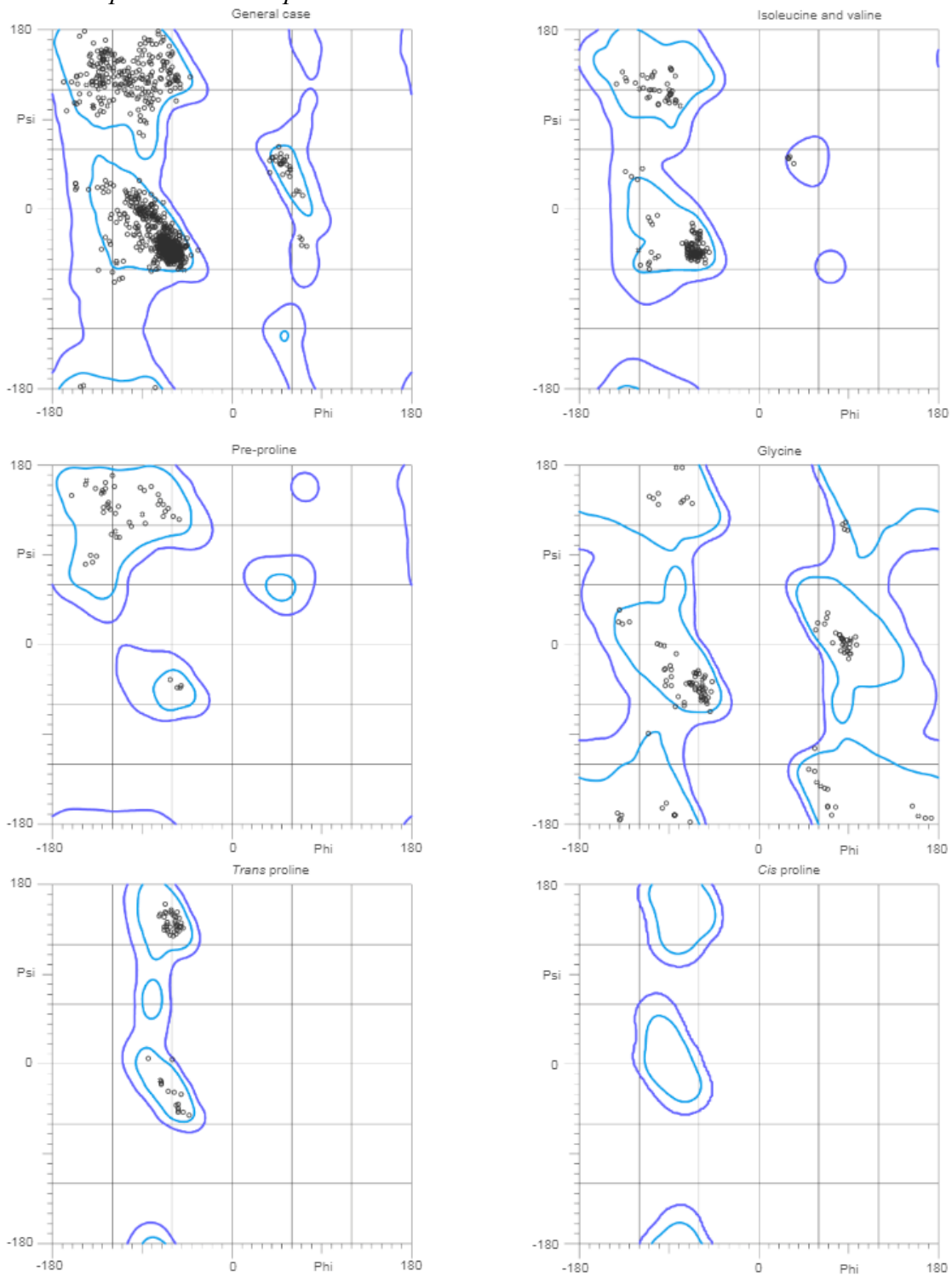


Рис.2. RSR для всех остатков белка

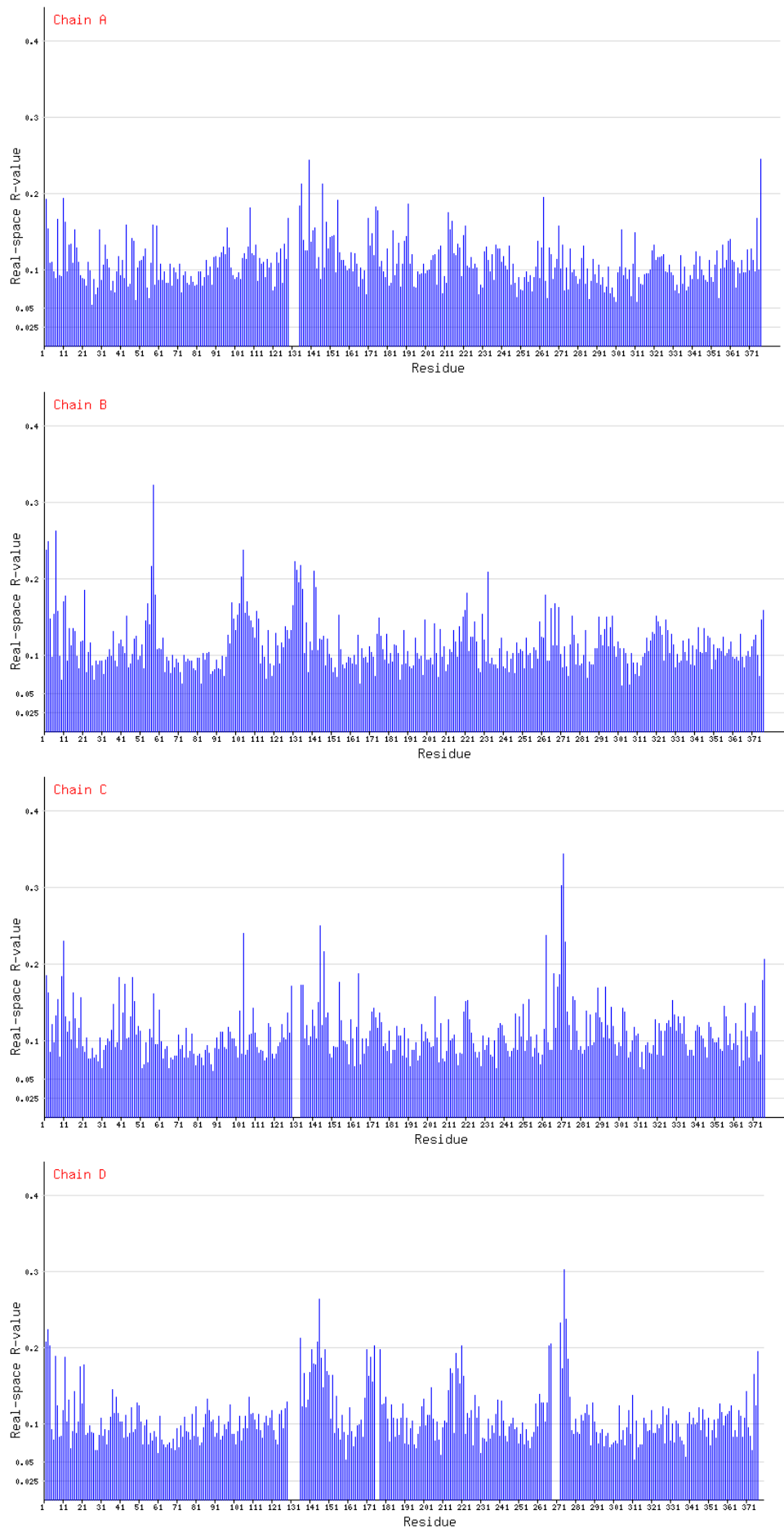


Рис.3. Significant regions (последовательно цены A-D)

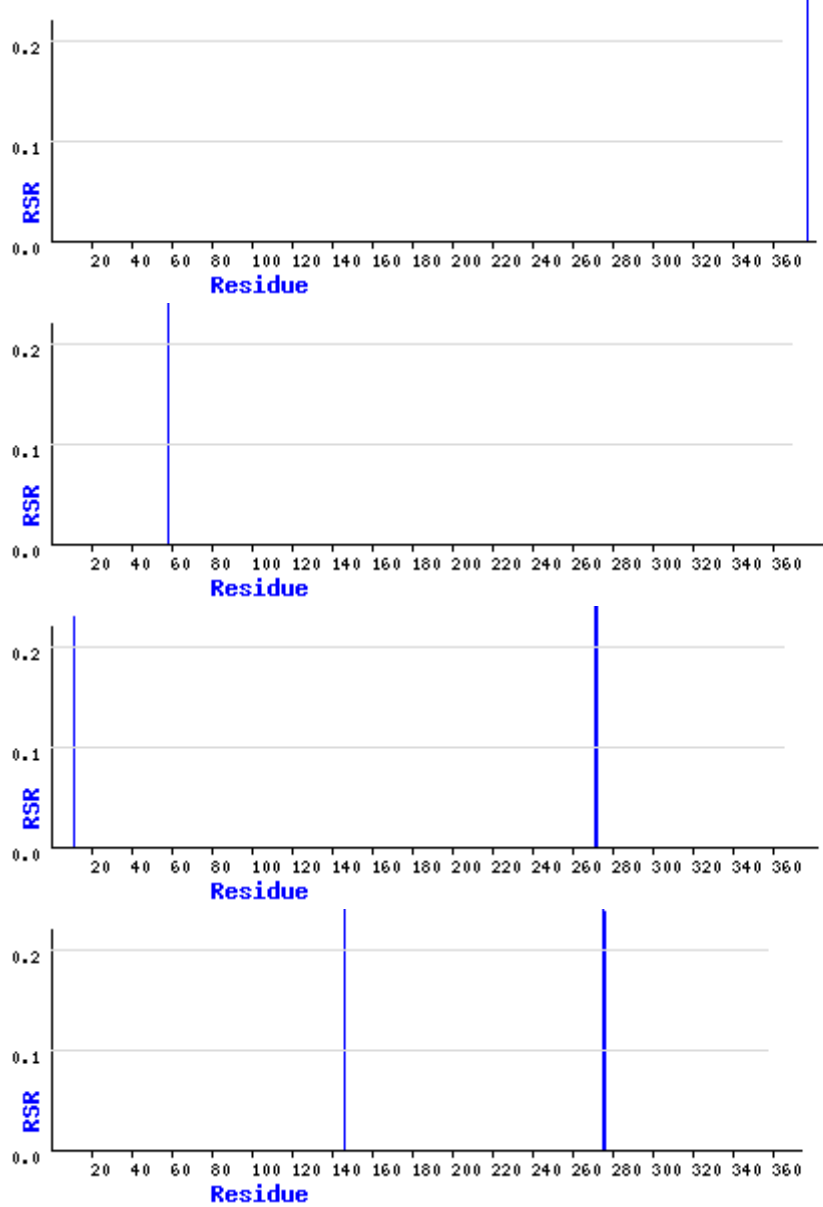
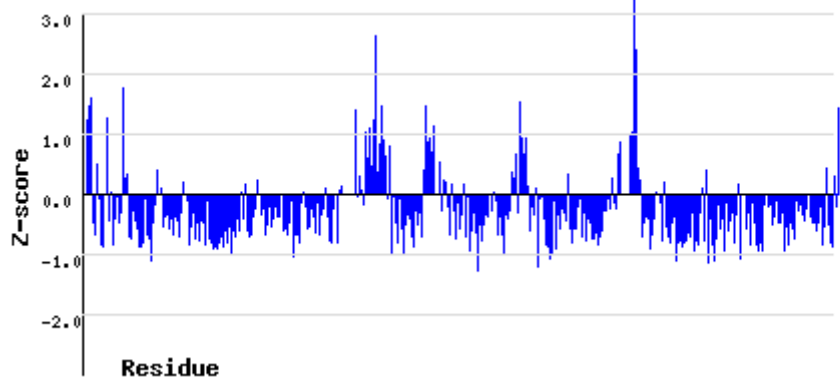
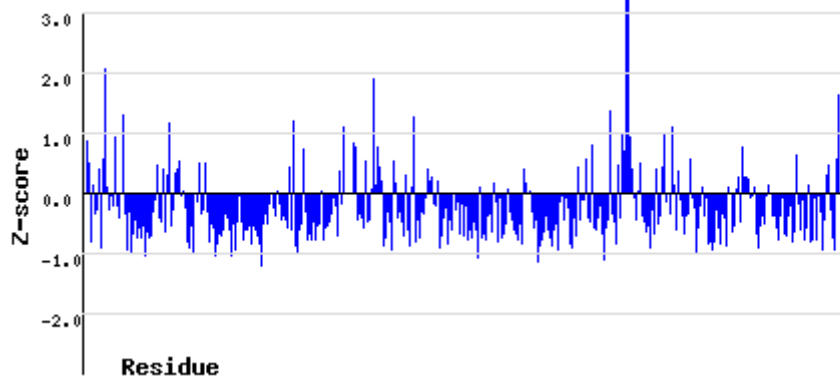
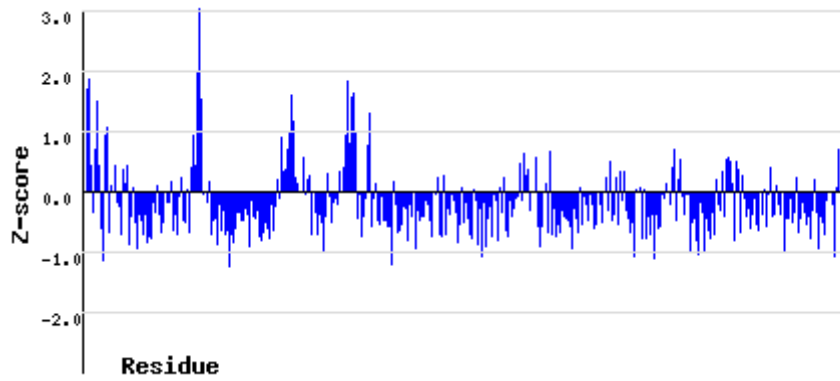
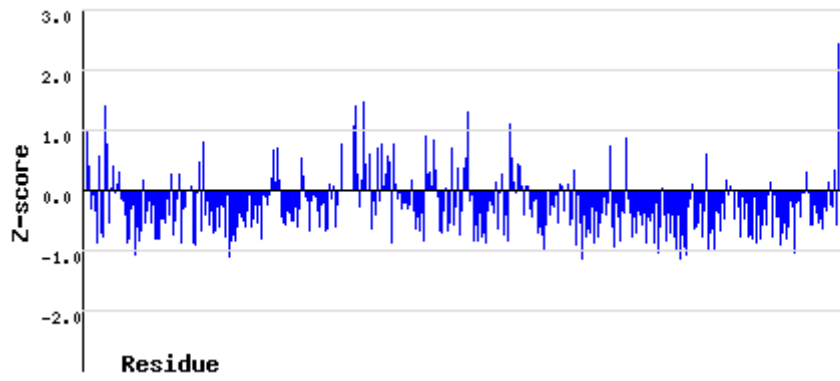


Рис.4. RSR-Z (последовательно цепи A-D)



Список литературы и сервисов

- [1] «EC:4m9a,» [В Интернете]. Available: <http://eds.bmc.uu.se/cgi-bin/eds/uusfs?pdbCode=4m9a>.
- [2] «RSCB PDB,» [В Интернете]. Available: <http://www.rcsb.org/pdb/explore/explore.do?structureId=4m9a>.
- [3] R. S., «Изучение файла структурных факторов,» [В Интернете]. Available: http://kodomofbb.msu.ru/~iltarn/term7/pr_4/index.html.
- [4] R. S., «Восстановление кристалла из PDB файла,» [В Интернете]. Available: http://kodomofbb.msu.ru/~iltarn/term7/pr_3/index.html.
- [5] *PyMol*.
- [6] «wwPDB Validation report на сайте PDB со структурой 4m9a,» [В Интернете]. Available: <http://www.rcsb.org/pdb/explore/explore.do?structureId=4m9a>.
- [7] «MolProbity,» [В Интернете]. Available: <http://molprobity.biochem.duke.edu>.
- [8] «WHAT_CHECK,» [В Интернете]. Available: <http://swift.cmbi.ru.nl/servers/html/index.html>.
- [9] «EDS,» [В Интернете]. Available: <http://eds.bmc.uu.se/eds/>.
- [10] «Выдача pdb_redo,» [В Интернете]. Available: http://www.cmbi.ru.nl/pdb_redo/m9/4m9a/index.html.
- [11] R. S., «Построение и визуализация электронной плотности,» [В Интернете]. Available: http://kodomofbb.msu.ru/~iltarn/term7/pr_1/index.html.