

Отчёт по качеству расшифровки структуры
глицеральдегид-3-фосфатдегидрогеназа *Bacillus*
stearothermophilus (PDB: 1GD1)

Выполнил: студент 4 курса ФББ МГУ

Медведев Дмитрий

Аннотация

Данный отчет представляет собой анализ качества расшифровки структуры 1GD1 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

Глицеральдегид-3-фосфатдегидрогеназа (ГАФД, GAPDH) – ключевой фермент гликолиза, осуществляющий реакцию образования 1,3-дифосфоглицеральдегида из глицеральдегид-3-фосфата и молекулы неорганического фосфата. Помимо гликолиза ГАФД принимает участие в других важных для клетки процессах: фиксация углерода в фотосинтезирующих организмах, различных метаболических путях, биосинтезе вторичных метаболитов, вовлечен во многие важные клеточные процессы. По последнему пункту можно выделить, что белок также участвует в процессе репарации ДНК у человека и бактерий [1,2].

Глицеральдегид-3-фосфатдегидрогеназа – фермент гликолиза, катализирующий обратимое окислительное фосфорилирование D-глицеральдегид-3-фосфата(G3P) до 1,3-дифосфоглицерата, с участием молекулы NAD+, являющейся кофактором реакции.

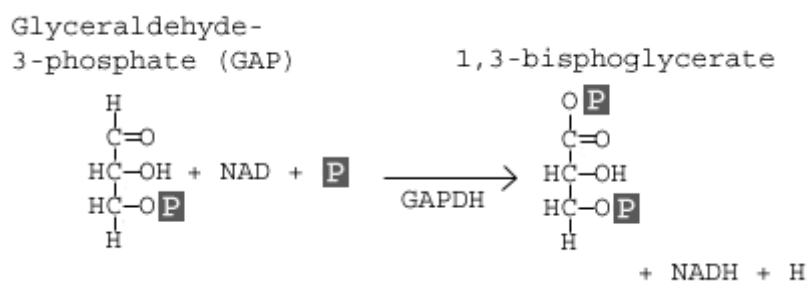


Схема 1. Общая схема реакции, катализируемая ферментом ГАФД [2]

Реакция, катализируемая GAPDH, является полностью обратимой. Из общей схемы видно, что для реакции требует ко субстрата (неорганические фосфат) и кофермент (NAD+), который находится в окисленной форме, способной принимать гидрид-ион и переходить в состояние NADH.

Глобула белка ГАФД является тетрамером общей массой ~ 150kDa. Структура мономера ГАФД имеет 2-белковых домена $\alpha\beta$ области (в основном антипараллельные β -листы), выложенных после 6-ой цепи Rossmann-подобного домена. Укладка Россмана(The Rossmann fold) — структурный мотив в третичной структуре белка, который связывает нуклеотиды (в частности, кофактор NAD).[4] В разных семействах ГАФД-подобных белков этот дополнительный домен изменяется по размеру и количеству структурных элементов. ГАФД имеет следующее устройство фолда: ядро представлено $\alpha\text{-}\beta\text{-}\alpha\text{-}\beta$ последовательностью и содержит 4-скрученных смешанных листа с характерными пси-петлями: 2134, лист 2 параллелен 1 листу [3]. Каждая укладка Россмана может связать один нуклеотид, в ферменте ГАФД она состоит из двух параллельных укладок, поскольку связывающий домен взаимодействует с динуклеотидом(NAD+) [4].

Мономер фермента состоит из трех доменов: аминокислотные остатки 1-147 и 312-332 относятся к NAD-связывающему домену, остатки с 148 по 311 входят в состав каталитического домена. Доменная организация тетрамерного белка представлена на Рисунок 8. Каталитический центр расположен, как бы в щели между двумя NAD-связывающими доменными структурами (непосредственно создающие “укладка Россмана”). Созданная таким образом щель, имеет характерные особенности у ряда организмов. Её ширина у человеческого фермента составляет 4-5 ангстрема, а у паразитических организмов таких как: *Leishmania mexicana*, *Trypanosoma brucei*, *Trypanosoma cruzi*, *Plasmodium falciparum* 7-8 ангстрем. Различия в размерах являются следствием того, что человеческий ГАФД образует за счет водородных связей с участием молекул воды, которая делает щель более узкой [5].

Результаты и обсуждения

Общая информация о модели

Здесь представлены результаты использования различных сервисов для оценки качества структуры и сбора информации о модели белка 1GD1 *Bacillus stearothermophilus* [6].

Фермент закристаллизован в виде гомотетрамера, связанного с FAD, служащий кофактором; присутствуют молекулы воды и ионы сульфата. Эту работу выполнили английские ученые Т. Skariyfiski, Р. С. Е. Moody и А. J. WonacotQ из Blackett Laboratory Имперского колледжа Лондона в 1986. Для решения фазовой проблемы использовали комбинацию двух методов: изоморфного замещения со структурой разрешением 2.7Å и последующего применения некристаллической симметрии для двух производных, чтобы улучшить фазы. Молекула NAD строилась со «скалкой» в структуре: никотинамидное кольцо взято в син-конформации, а рибозимное вместе с аденином в анти-конформации. Далее результат для кофактора был просчитан с помощью программы REALSPACE(Diamond. 1974, 1976) Первая модель имела R-factor 0.446 с начальным разрешением 2.7Å.

Всего было получено 130 797, из них в конечное формирование модели были взяты лишь 116 031. Теоретически наблюдаемое разрешение для структуры 1.8 Å было достигнуто с 81% рефлексов.

Разрешение структуры : 1.80 Å;

Полноте данных(набор полных рефлексов): 80.1 % ;

Диапазон расширения: 18.81 - 1.80 Å.

Параметры кристаллической структуры белка глицеральдегид-3-фосфатдегидрогеназы(ГАФД):

CRYST1 82.440 124.100 82.540 90.00 108.98 90.00 P 1 21 1 8

Первые три числа – значения длин направляющих векторов кристалла: 82.440 124.100 82.540; следующие три – углы между направляющими векторам: 90.00 108.98 90.00; следующее обозначение группы симметрии P 1 21 1, означает что кристаллическая решетка Браве имеет моноклинную симметрию, число ячеек. Симметрии: X,Y,Z и -X,Y+1/2,-Z. Число молекул в ячейке равно 8.

Таблица 1. Параметры кристаллографической ячейки

Длина (Å)	Угл (°)
a = 82.44	$\alpha = 90.00$
b = 124.10	$\beta = 108.98$
c = 82.54	$\gamma = 90.00$

Значения индикаторов качества модели в целом

Стандартный фактор достоверности (R-фактор) – это величина, которая характеризует модель после оптимизации. Это мера различия оптимизированной модели и экспериментальных данных.

R-фактор: 0.177;

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.

R_free: ---;

Хорошими значения принято считать R-фактор < 0.25 и R-free < 0.20. Соответственно, R-фактор говорит о хорошем качестве модели. в структуре 1GD1 R-free отсутствует и нам не может сказать об качестве оптимизации модели.

RSR (real-space R-фактор) показывает, насколько построенная по модели электронная плотность сочетается с «экспериментальной». Пространственный R-фактор позволяет оценить соответствие модели группы атомов, сделать оценку локального качества. Его среднее значение составляет по аминокислотам: 0.081(8.1%) (со стандартным отклонение 0.036(3.6%)). Хорошими

значениями принято считать $RSR < 10\%$, плохими – $RSR > 20\%$. Значение RSR для исследуемой модели говорит о ее хорошем качестве.

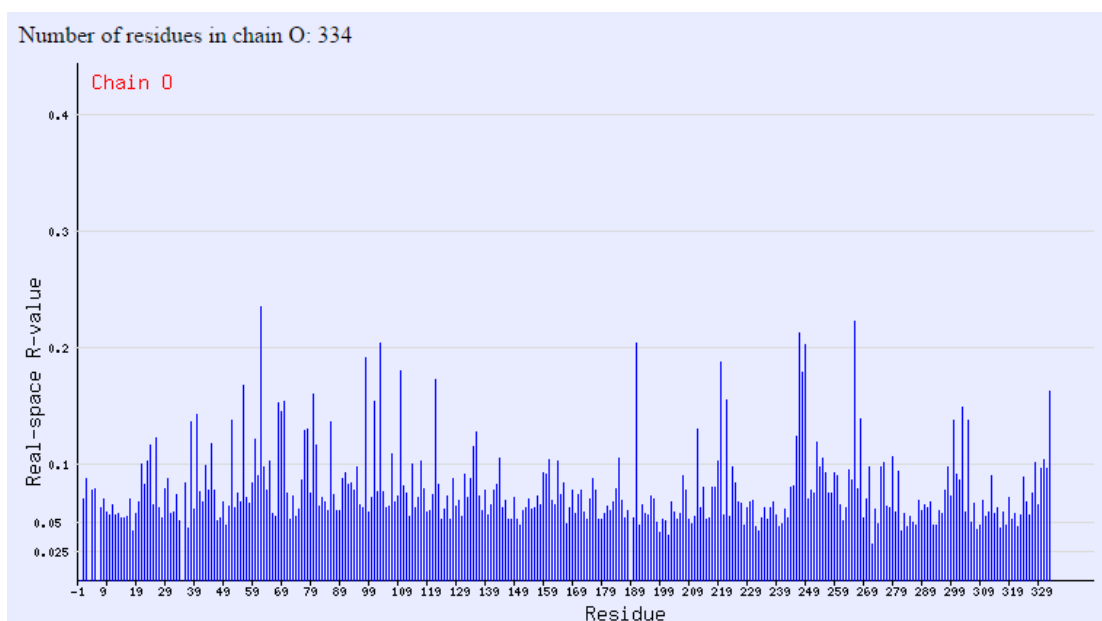


Рисунок 1. Пространственный R-фактор для всех аминокислотных остатков структуры 1GD1. Цепь O.

Как видно из диаграммы, что есть ряд аминокислот, чей $RSR > 10\%$: ASN62 – 23.5%, LYS191-20.4%, GLU266 – 22.3% и т.д. Определенной зависимости в значениях RSR и положениях аминокислот в полипептидной цепи не видно. В других цепях картина распределения значений RSR иная, количество аминокислот, имеющих этот параметр больше 10% выше:

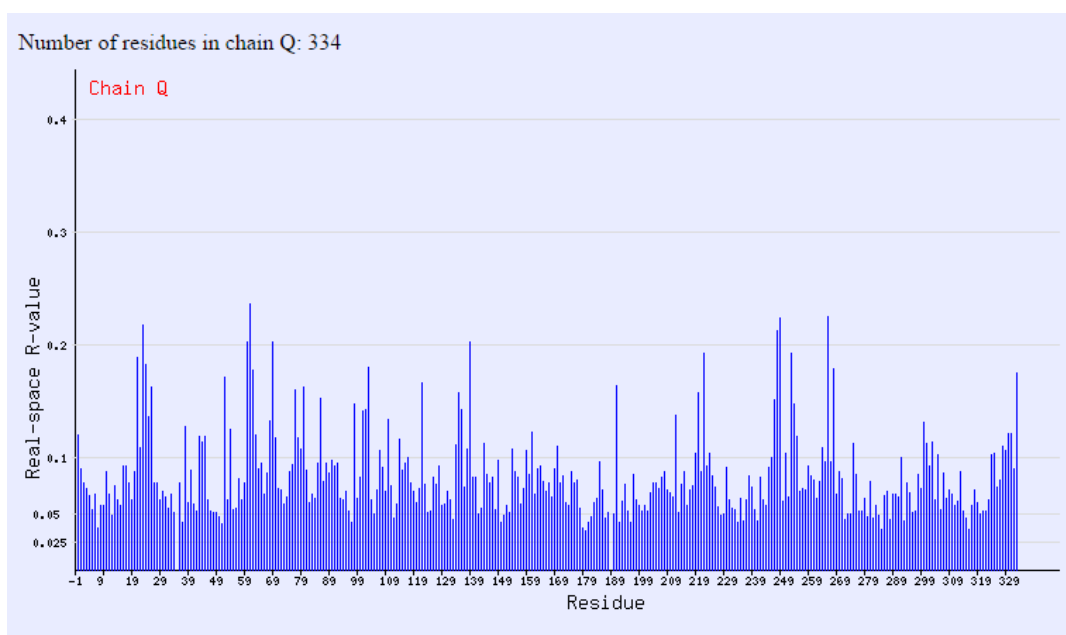


Рисунок 2. Пространственный R-фактор для всех аминокислотных остатков структуры 1GD1. Цепь Q.

Температурный фактор (B-фактор)

Температурный фактор (B-фактор) характеризует меру динамической неупорядоченности кристалла, обусловленную тепловым движением, а также небольшими отличиями в расположении молекул белка в кристалле. Так как с повышением температуры увеличиваются тепловые колебания атомов, этот фактор увеличивается с увеличением температуры.

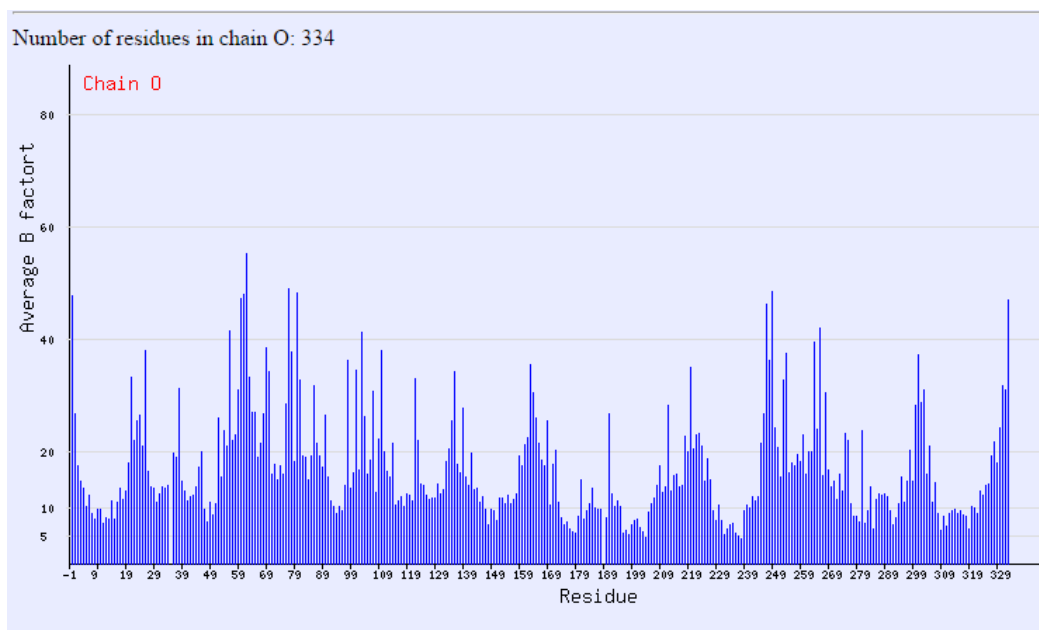


Рисунок 3. В-фактор (температурный фактор) для цепи O структуры 1GD1.

Наибольшее значение получила аминокислота ASN62 – 55.280. У большинства аминокислот значение температурного фактора меньше 40.000. Стоит отметить некую цикличность образования пиков возрастание, и соответственно увеличение динамичности, значений фактора. Возможно, это обусловлено расположением аминокислот в пространстве (находятся на поверхности структуры и ее концах). Для всех диаграмм разных цепей, В-фактор представлен схожими пиками.

Z-score

$Z\text{-score} = (RSR - \langle RSR \rangle) / \sigma$. 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. С его помощью можно определить маргинальный ли остаток: Z-score больше 2.

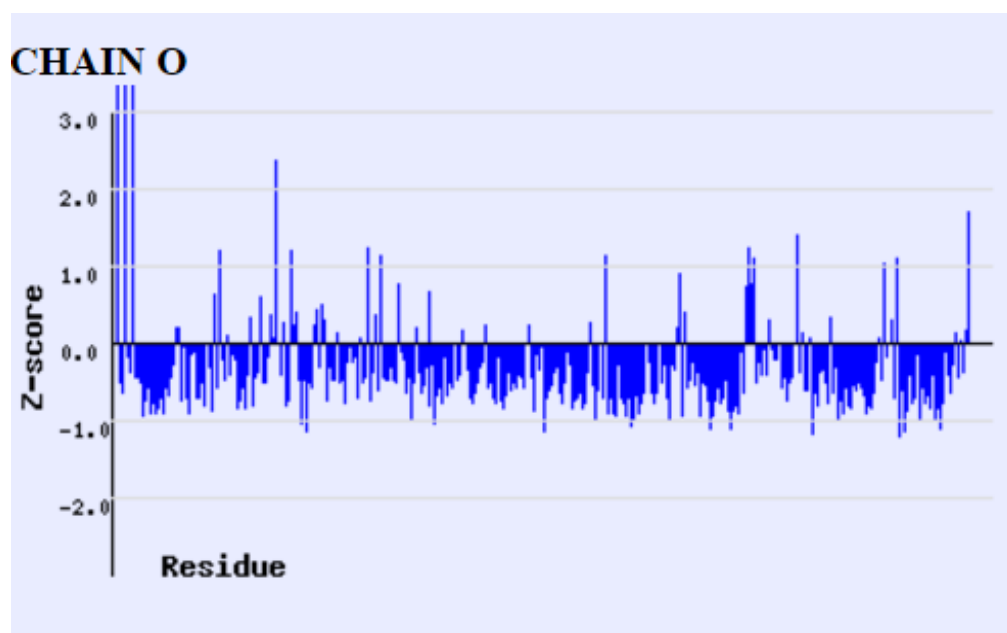


Рисунок 4. Z-score для аминокислотных остатков структуры 1GD1 цепи O.

Как видно из рисунка, есть аминокислоты, чей Z-score больше 2: ASN6 – 172.012, VAL3 – 241.40 и т.д.

Это параметр сильно отличается у всех цепей. И если среди цепей P, Q, R все выглядит хорошо, то цепь O выбивается по своим значениям среднего Z и sigma.

Таблица 2. Значение Z-score для разных цепей.

Chain	Nres	<Z>	Sigma	%<-1	%<0	%>1	%>2
O	334	1.46	19.94	5.09	83.53	4.49	1.20
P	334	-0.35	0.70	6.59	79.94	3.59	0.60
Q	334	-0.31	0.68	5.39	76.35	3.59	0.90
R	334	-0.41	0.68	3.59	83.53	3.89	0.60
All	1336	0.10	9.99	5.16	80.84	3.89	0.82

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.

Для получения соответствующей информации был использован сервис MolProbity. 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. В начале, для получения информации о модели 1GD1 были добавлены водород (с параметрами по умолчанию: добавить электронную плотность водородов, детектировать все возможные инверсии боковых цепей Asn, Gln, His).

Таблица 3. Данные сервиса MolProbity для структуры 1GD1.

All-Atom Contacts	Clashscore, all atoms:	5.75	94 th percentile* (N=837, 1.80Å ± 0.25Å)	
	Clashscore is the number of serious steric overlaps (> 0.4 Å) per 1000 atoms.			
Protein Geometry	Poor rotamers	55	5.06%	Goal: <0.3%
	Favored rotamers	941	86.49%	Goal: >98%
	Ramachandran outliers	8	0.61%	Goal: <0.05%
	Ramachandran favored	1247	95.05%	Goal: >98%
	MolProbity score [^]	2.19		46 th percentile* (N=11444, 1.80Å ± 0.25Å)
	C β deviations >0.25Å	51	4.10%	Goal: 0
	Bad bonds:	15 / 10277	0.15%	Goal: 0%
Bad angles:	353 / 13952	2.53%	Goal: <0.1%	
Peptide Omegas	Cis Prolines:	0 / 44	0.00%	Expected: ≤1 per chain, or ≤5%

В первой строке (Clashscore, all atoms) представлено число недопустимых наложений атомов >0.4 Å на 1000 атомов. В данном случае – это 5.75. Ramachandran favored - число и процент остатков в предпочитаемой области – 941(86.49%) Poor rotamers (число остатков с маргинальными по отклонению от ротамеров боковыми цепями) составляет 5.06%, что значительно превышает

допустимое значение 0.3%. $C\beta$ deviations > 0.25 Å - число $C\beta$ атомов с неприемлемым отклонением от ожидаемого положения – равняется 51 (4.10%), число плохих валентных связей равняется 15/10277 (0.15%), число плохих валентных углов, существенно отклоняющихся от теории, составило 353/13952 (2.53%).

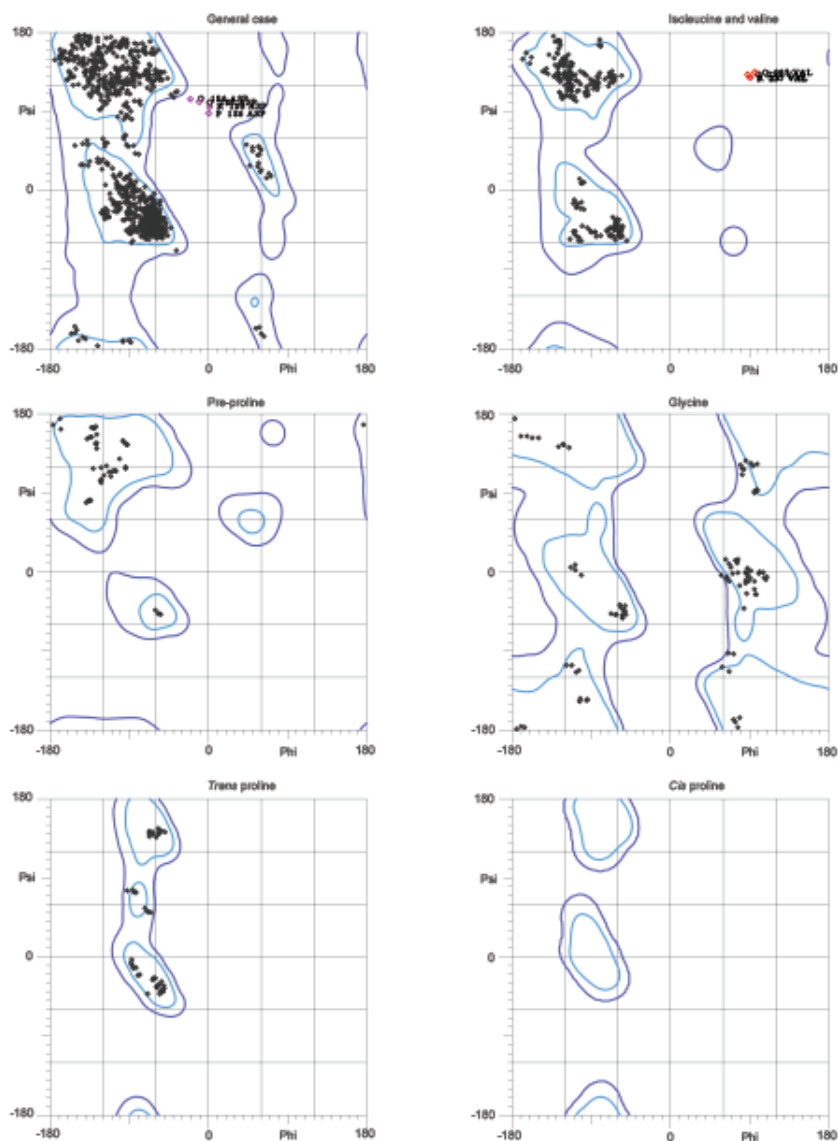


Рисунок 5. Карты Рамачандрана для структуры 1GD1, построенные с помощью сервиса MolProbity.

По результат 95.1% (1263/1328) аминокислот находятся в предпочитаемой области, 99.4% (1320/1328) – в разрешенной. У структуры есть остатки, которые находятся в запрещенной области и, соответственно, имеют сильно отклоняющиеся конформации – это те 0.6% все аминокислот (8 остатков):

- O 186 ASP (-20.0, 104.6)
- O 237 VAL (88.6, 132.1)
- P 186 ASP (1.2, 88.5)
- P 237 VAL (92.0, 129.6)
- Q 186 ASP (-10.1, 101.8)
- Q 237 VAL (96.9, 135.3)
- R 186 ASP (2.3, 95.7)
- R 237 VAL (91.6, 129.8)

Видно, что внутри одной цепи одни и те же остатки имеют сильное отклонение в конформации.

Затем с помощью сайта WHAT_CHECK опцией Fine Packing Quality Control была оценена комфортность окружения для всех остатков. Нормальные PRO кольца имеют поворота амплитуду(Q) от 0,20 и 0,45 Å. Если Q меньше, чем 0.20 Å для PRO это может означать, расстройтва между двумя различными нормальных формами. Если Q выше, чем 0.45 Å то есть предположение, что проблема в оптимизации структуры. Для 1GD1 было найдено 12 пролинов(по 3 в каждой из 4 цепей). 13 аминокислот из 4 цепей были выделены, как имеющие ненормальное уплотнение окружения. Так же были обнаружены необычные остатки в структуре, которые нуждаются в рассмотрении, их Z-score имел значение меньше -2.5:

```
416 LEU ( 82-) P -3.31
750 LEU ( 82-) Q -3.26
82 LEU ( 82-) O -3.25
1084 LEU ( 82-) R -3.22
55 ALA ( 55-) O -2.54
```

Одно из наиболее интересных особенностей являются наличие среди HIS, ASN, GLN поверхности аминокислоты с расшифровкой flips:

```
62 ASN ( 62-) O
63 ASN ( 63-) O
154 ASN ( 152-) O
257 ASN ( 256-) O
372 ASN ( 38-) P
397 ASN ( 63-) P
488 ASN ( 152-) P
512 HIS ( 176-) P
591 ASN ( 256-) P
706 ASN ( 38-) Q
731 ASN ( 63-) Q
822 ASN ( 152-) Q
925 ASN ( 256-) Q
1040 ASN ( 38-) R
1064 ASN ( 62-) R
1065 ASN ( 63-) R
1156 ASN ( 152-) R
1259 ASN ( 256-) R
```

Поиск и анализ маргинальных остатков

Выберем конкретные остатки и рассмотрим их. Данные взяты из Validation report, выдачи MolProbity и WHAT_CHECK.

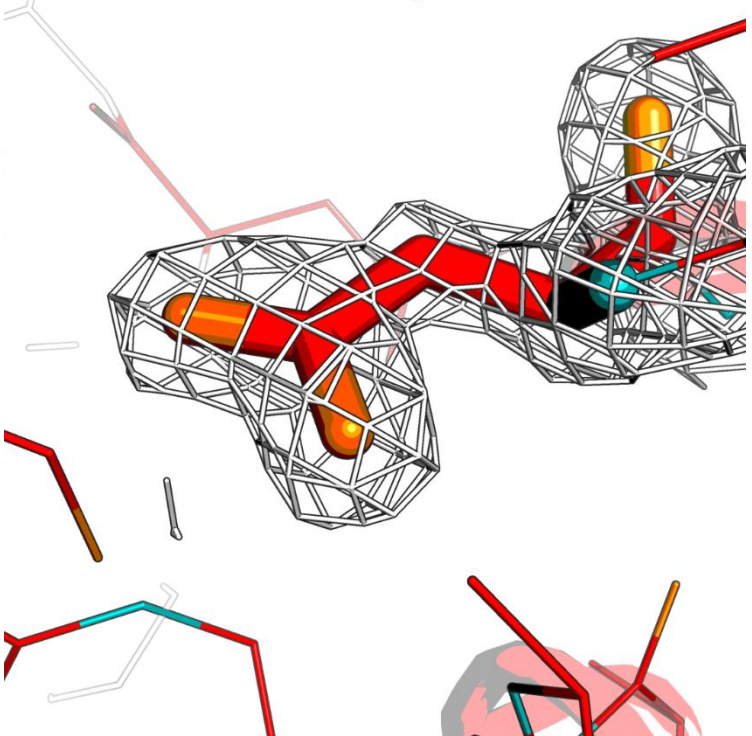
Таблица 4. Примеры маргинальных остатков для структуры 1GD1

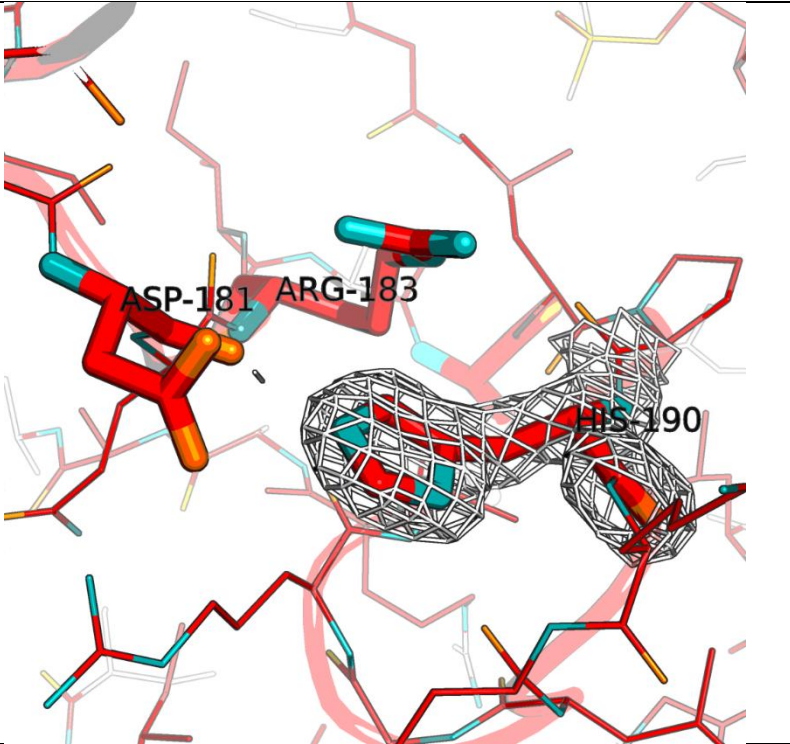
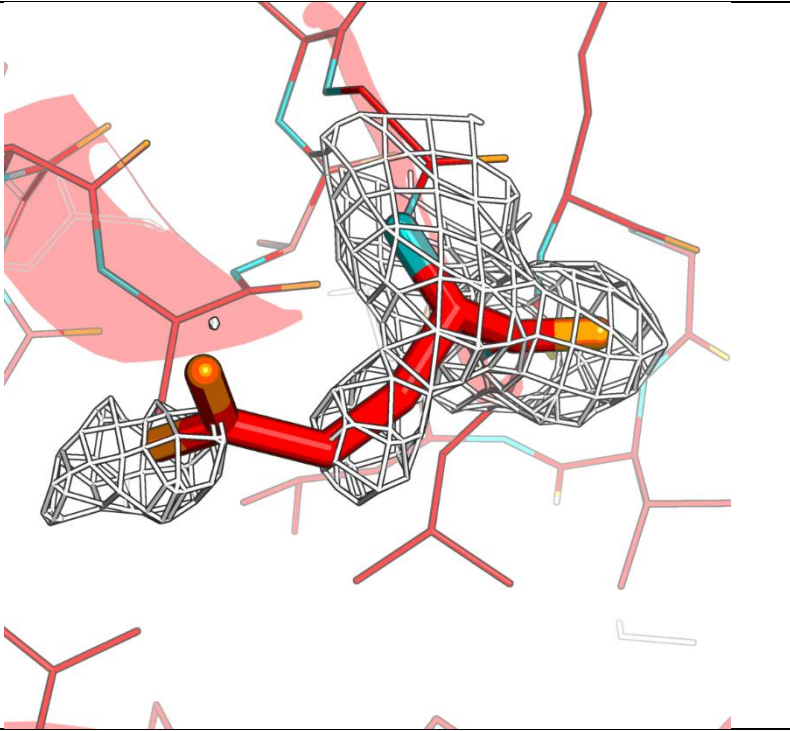
	Остаток(цепь)	Причина маргинальности
1	65 VAL (Q)	Имеет недопустимое перекрытие с 60 ASN (0.63Å); RSR = 0.203.
2	186 ASP(O)	Является «выбросом» по валентным углам; RSR=0.069
3	172 MET(R)	Имеет недопустимое перекрытие с 208 THR (0.68Å); RSR = 0.097. Отклоняются углы chi: 68.9, 138.6, 230.6 – poor rotamer
4	248 LYS(R)	Имеет недопустимое перекрытие с 165 PHE (0.60Å); RSR = 0.203. Разрешённые углы chi: 09.1, 162.9, 125.6, 215.9.
5	237 VAL(P)	Является «выбросом» по валентным углам; RSR=0.076
6	266 GLU(Q)	Отклоняются углы chi: 310.9, 48.4, 96.4 – poor rotamer; RSR=0.225, высокий B-фактор(49.880).
7	23 PRO(O)	Образует плохой угол C-N-CA: 6.9 σ; RSR=0.102, большой B-фактор(25.600)

8	25 ILE(Q)	Имеет недопустимое перекрытие с 19 ALA (0.54Å); Отклоняются углы chi: 83.3, 114.5 – poor rotamer; Образует плохой угол CA-CB-CG2: 5.1 σ; RSR=0.136; высокий B-фактор(48.650)
9	190 HIS(O)	Выброс по длине связи CD2--NE2: 4.6 σ; образует плохой угол CA-C-O: 5.4 σ; RSR=0.054; низкий B-фактор(8.350)

Рассмотрим некоторые маргинальные остатки из Таблицы 4 подробнее. Возьмем в рассмотрение такие остатки как 186 ASP(O), 266 GLU(Q), 190 HIS(O).

Таблица 5. Характеристика маргинальных остатков структуры 1GD1

186 ASP(O)	Является «выбросом» по валентным углам; RSR=0.069	
<p>На рисунке изображен 186 ASP. Видно, что электронная плотность, изображённая на уровне $\sigma=2.0$, находится на своем месте. Является выброс возможно из-за окружения.</p>		

<p>190 HIS(O)</p>	<p>Выброс по длине связи CD2--NE2: 4.6 σ; образует плохой угол CA-C-O: 5.4 σ; RSR=0.054; низкий B-фактор(8.350)</p>	
<p>На рисунке изображен 190 HIS. Видно, что электронная плотность изображённая на уровне $\sigma=2.0$, находится на своем месте. Как видно из рисунка, радикал расположен довольно близко к ASP (с которым он может образовать водородные связи) и ARG (который образует с ним пи-сопряжённую систему). Возможно именно поэтому он имеет такие плохие углы и длины связи. B- фактор, указывает на то, что аминокислота находится в стабильном положении.</p>		
<p>266 GLU(Q)</p>	<p>Отклоняются углы χ_1: 310.9, 48.4, 96.4 – poor rotamer; RSR=0.225, высокий B-фактор(49.880).</p>	
<p>На рисунке изображен 266 GLU. Видно, что электронная плотность, изображённая на уровне $\sigma=1.0$, находится не на своем месте и в целом плохо покрывает структуру. Ввиду высокого B- фактора, есть предположение, что структура испытывала большое количество колебаний.</p>		

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. После оптимизации структуры, сразу стоит отметить, что появился R-free фактор, равный 0.1619, что очень хорошо. Также улучшился сам R-фактор он стал равен 0.1390.

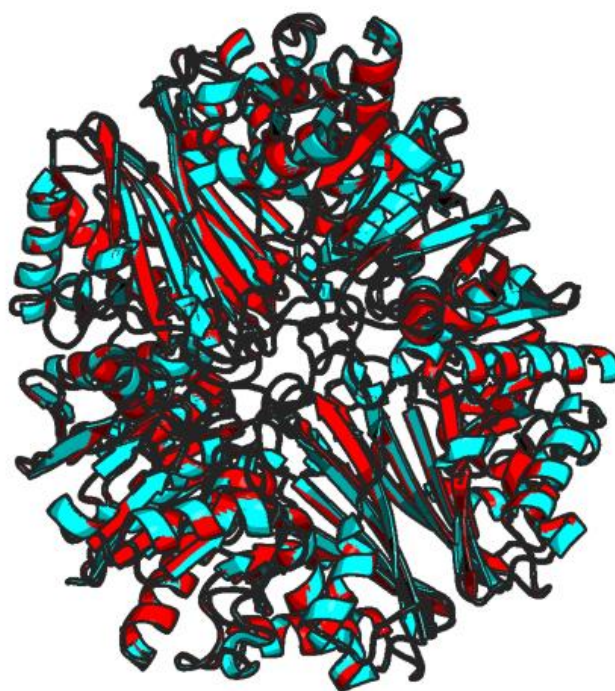
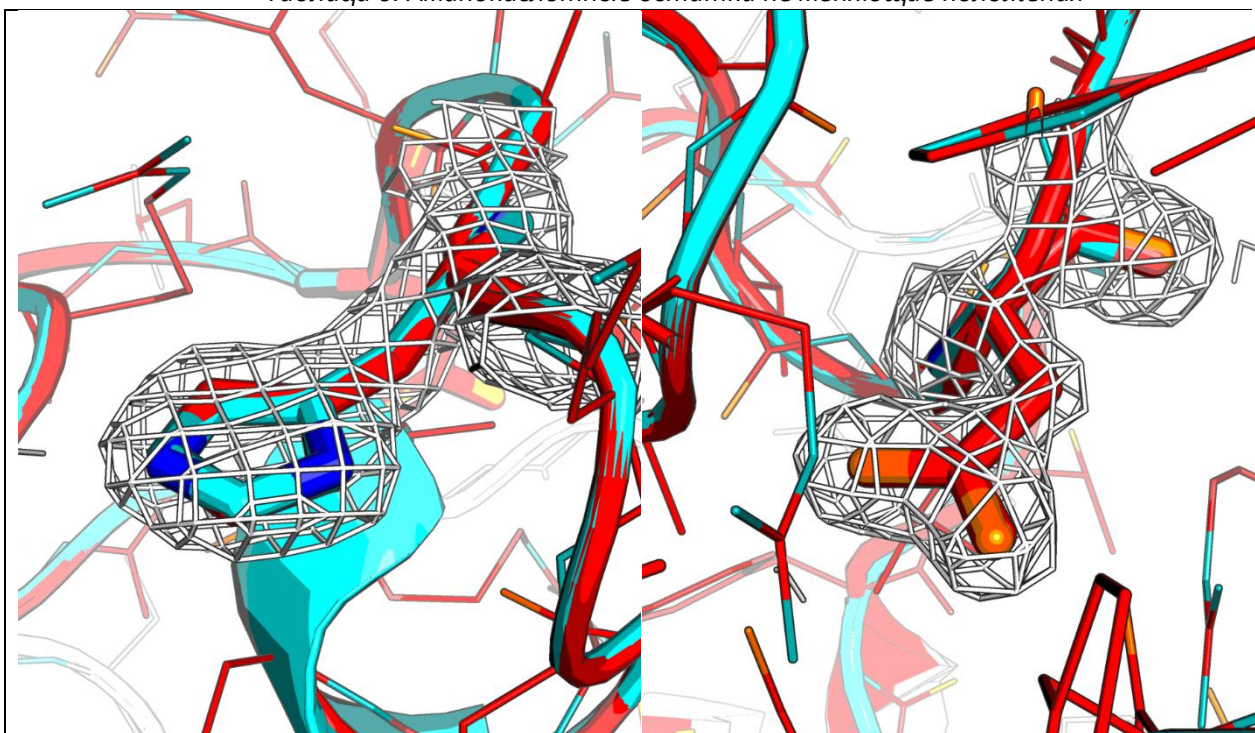


Рисунок 6 Наложение структуры 1GD1 до обработки(красная), на структуру после(сияняя)

К координальным изменения в структурной организации улучшение с помощью PDB_redo не привело. Но на уровне расположение аминокислот можно заметить, что некоторые поменяли положение.

Таблица 6. Аминокислотные остатки не меняющие положения



На рисунке HIS(слева) и ASP(справа) имеют такое же положение, что и в структуре после обработки

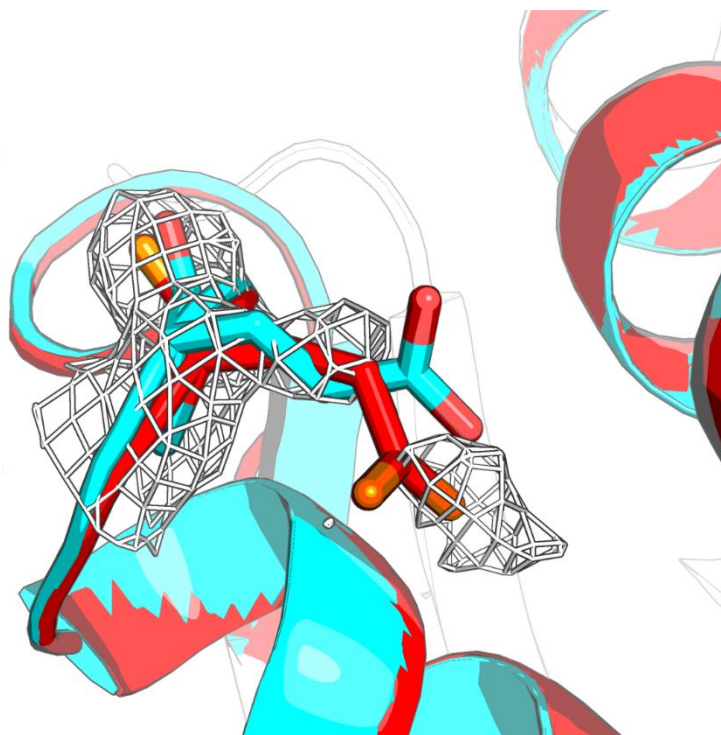


Рисунок 7. Изменение положения GLU в структуре 1GD1 до обработки(красная) и после(синяя). Как видно изменённое положение все равно не покрывается электронной плотностью.

Выводы

По результатам проведенного анализа структуры 1GD1 можно сказать, что качество последней является хорошим, но не отличным. Для данной структуры нет R-free фрактура(его можно получить с помощью PDB_redo). На карте Рамачандрана есть ряд выбросов в запрещенных областях и немалое число аминокислот с неправильной конформацией боковой цепи. Z-score в среднем, является хорошим, но различные в разных цепях и для некоторых аминокислот довольно большой (что плохо). Положение электронной плотности относительно атомов полипептидной цепи не всегда совпадает. Средний температурный фактор имеет нормальное значение.

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. Но в целом, модель тяжело кардинально улучшить с помощью сервисов переоптимизации структур.

Список литературы:

1. Seidler N.W. GAPDH : Biological Properties and Diversity Advances in Experimental Medicine and Biology.
2. Ferreira E. et al. Glyceraldehyde-3-phosphate dehydrogenase is required for efficient repair of cytotoxic DNA lesions in Escherichia coli. // Int. J. Biochem. Cell Biol. 2015. Vol. 60. P. 202–212.
3. Structural Classification of Proteins 2 [Online]. URL: <http://scop2.mrc-lmb.cam.ac.uk/3000043.html>.
4. Reis M. et al. The catalytic mechanism of glyceraldehyde 3-phosphate dehydrogenase from Trypanosoma cruzi elucidated via the QM/MM approach. // Phys. Chem. Chem. Phys. The Royal Society of Chemistry, 2013. Vol. 15, № 11. P. 3772–3785.
5. Zubrzycki I.Z. Homology modeling and molecular dynamics study of NAD-dependent glycerol-3-phosphate dehydrogenase from Trypanosoma brucei rhodesiense, a potential target enzyme for anti-sleeping sickness drug development. // Biophys. J. 2002. Vol. 82, № 6. P. 2906–2915.
6. Skarzynski, T., Moody, P.C., Wonacott, A.J. Structure of holo-glyceraldehyde-3-phosphate dehydrogenase from Bacillus stearothermophilus at 1.8 Å resolution. 1987// J.Mol.Biol. 193: 171-187

В работе также использовались:

- сервис MolProbity(<http://molprobity.biochem.duke.edu/>);
- сервис PDB(<http://www.rcsb.org/pdb/home/home.do>);
- сервер EDS(<http://eds.bmc.uu.se/eds/index.html>);
- сервиса «WHAT_CHECK,» (<http://swift.cmbi.ru.nl/servers/html/index.html>)

Визуализация производилась с помощью PyMol-1.7.2.1.