

**МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ  
имени М.В.ЛОМОНОСОВА**

**ФАКУЛЬТЕТ БИОИНЖЕНЕРИИ И БИОИНФОРМАТИКИ**

**Отчёт по качеству расшифровки структуры с PDB-кодом 5DQY,  
полученной методом рентгеноструктурного анализа**

Работа студента 4 курса  
Старикова Сергея

**Москва  
2017**

# Оглавление

Аннотация	2
Введение	2
Результаты	3
1. Общая информация о модели	3
2. Значения индикаторов качества модели в целом	4
3. Анализ маргинальных остатков	7
4. Анализ пяти маргинальных остатков	9
5. Сравнение моделей из PDB и PDB-REDO	12
Выводы	12
Список литературы	13

## Аннотация

В этой работе представлен анализ качества расшифровки структуры 5DQY, рассмотрены основные индикаторы качества модели, а также несколько маргинальных остатков.

## Введение

Для анализа был выбран белок тиоредоксин из организма человека.

Тиоредоксины - семейство маленьких белков, представленных во всех организмах от архей до человека[1][2]. Они участвуют во многих важных биологических процессах, включая определение ОВР потенциала клетки и передачу сигнала. У человека тиоредоксин кодируется геном TXN. Главной частью белка является дисульфидная связь. При помощи этой неё он может восстанавливать дисульфидные связи других белков:



# Результаты

## 1. Общая информация о модели

Целью [работы](#) [1] (2015 г), из которой взята структура 5DQY, было определение структуры полностью окисленного тиоредоксина для того, чтобы понять, как этот белок может регулироваться клеткой.

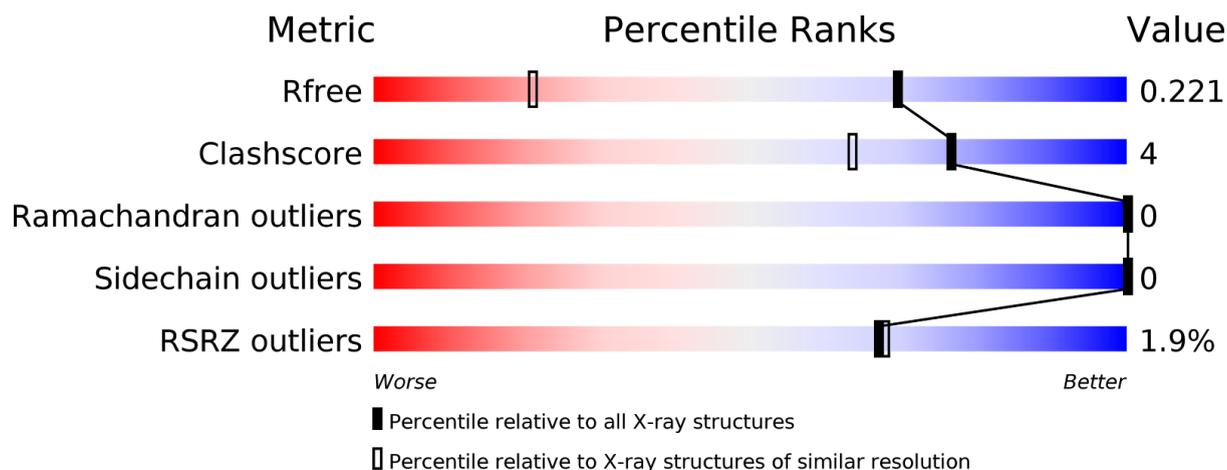
В таблице 1 представлена краткая характеристика структуры белка [2].

**Таблица 1.** Основные параметры структуры 5DQY.

Параметр	Значение
Разрешение структуры	1.40 Å
Количество измеренных рефлексов	20454
Полнота данных	96.5 %
Диапазон разрешений	14.93 - 1.40 Å
Тип симметрии	I 2 2 2
Параметры кристаллографической ячейки	a=53.20 Å, b=61.23 Å, c=64.84 Å вектора ортогональны
Метод решения фазовой проблемы	молекулярное замещение

## 2. Значения индикаторов качества модели в целом

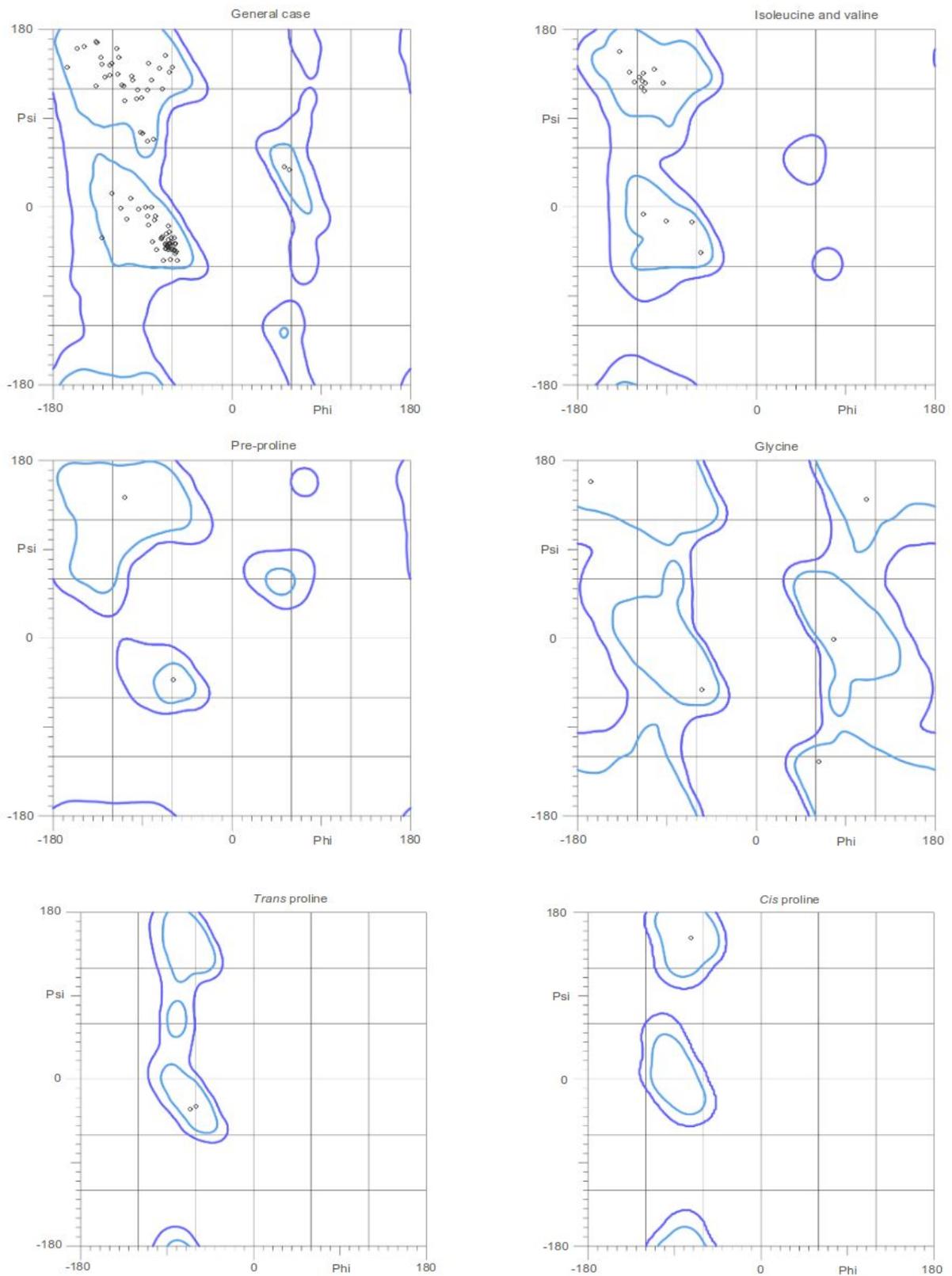
На рисунке 1 показаны значения индикаторов относительно всех структур PCA. Из всех значений хорошими можно назвать все, а лучше всех получились Ramachandran outliers и Sidechain outliers - их нет.



**Рис. 1.** Значения различных индикаторов качества модели.

Значения R-фактора и R-free составляет 0,158 и 0,209, соответственно [2]. Так как значения R-free <25%, а R-free – R <10%, можно сделать вывод, что у модели высокое качество.

## Карта Рамачандрана:



**Рис. 2** Карта Рамачандрана для 5DQY

Маргинальных остатков на карте Рамачандрана не выявлено.

**Таблица 2.** Критерии качества модели, выданные сервисом MolProbity.

Зеленым цветом выделены значения, удовлетворяющие условиям хорошей модели, желтым – значения, отклоняющиеся от условий хорошей модели и красным – значения, не удовлетворяющие условиям хорошей модели.

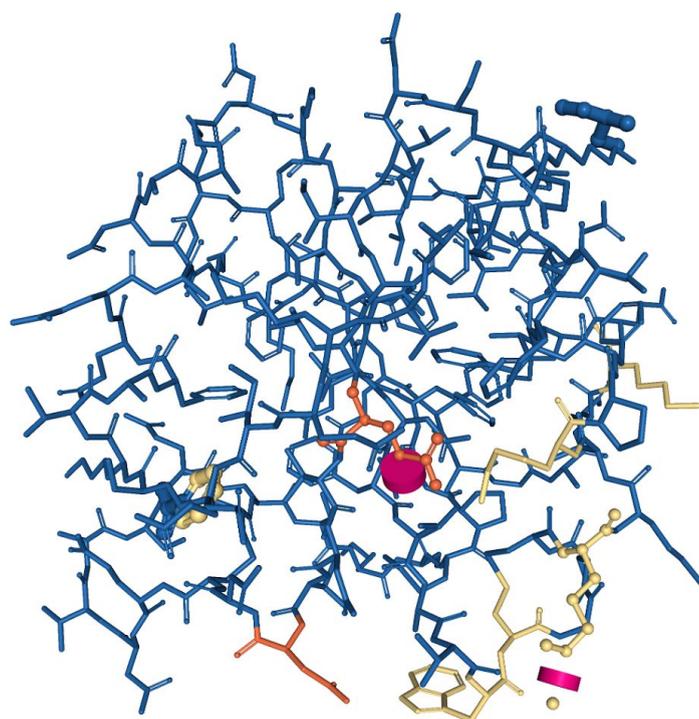
All-Atom Contacts	Clashscore, all atoms:	3.54		96 <sup>th</sup> percentile* (N=480, 1.40Å ± 0.25Å)
	Clashscore is the number of serious steric overlaps (> 0.4 Å) per 1000 atoms.			
Protein Geometry	Poor rotamers	0	0.00%	Goal: <0.3%
	Favored rotamers	91	95.79%	Goal: >98%
	Ramachandran outliers	0	0.00%	Goal: <0.05%
	Ramachandran favored	106	99.07%	Goal: >98%
	MolProbity score	1.14		98 <sup>th</sup> percentile* (N=3363, 1.40Å ± 0.25Å)
	Cβ deviations >0.25Å	2	1.92%	Goal: 0
	Bad bonds:	3 / 872	0.34%	Goal: 0%
	Bad angles:	3 / 1177	0.25%	Goal: <0.1%
Peptide Omegas	Cis Prolines:	1 / 3	33.33%	Expected: ≤1 per chain, or ≤5%

### 3. Анализ маргинальных остатков

Для поиска маргинальных остатков использовался результат анализа с помощью MolProbity (таблица 2). Были найдены выбросы по длинам и углам нескольких связей. Некоторые из найденных остатков указаны в таблице 3.

**Таблица 3.** Некоторые маргинальные остатки в структуре 5DQY.

N	Остаток	Критерий маргинальности
1	GLU47	выброс по углу связи CB-CG-CD
2	CYS32	выброс по длине связи с CYS35 (SG--SG)
3	MET37	выброс по длине связи SD--CE
4	ASP60	выброс по длине связи CB-CG
5	GLN63	выброс по углу связи C-CA-CB
6	LYS94	выброс по углу связи CD-CE-NZ



**Рис. 3.** Изображение структуры 5DQY с показанными красным и желтым цветом остатками с нестандартной конформацией (изображение соответствует отчету по валидации, приведенному на сайте PDB). Clashes показаны красными дисками. Далее в этом отчете эти данные не используются, используются только результаты анализа с помощью сервиса MolProbity.

#### 4. Анализ пяти маргинальных остатков.

На рисунке 4 изображен остаток GLU47, маргинальный по углу связи CB-CG-CD. Электронная плотность недостаточно хорошо ложится на остаток, а согласно электронной плотности угол связи действительно странный. Скорее всего данный остаток является маргинальным из-за особенностей структуры.

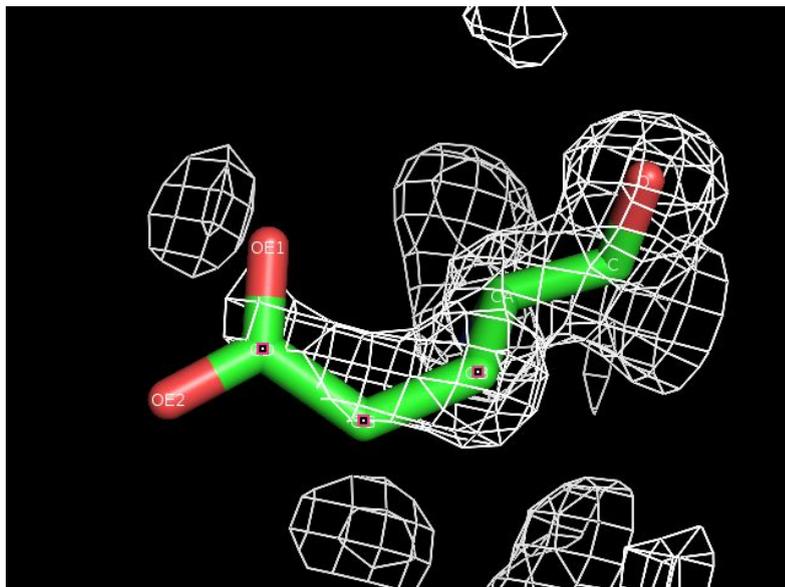


Рис. 4. Остаток GLU47 в структуре 5DQY.

На рисунке 5 изображен остаток MET37, маргинальный по длине связи SD-CE. Электронная плотность достаточно хорошо ложится на остаток. Несовпадения не выявлено, скорее всего это не маргинальный остаток.

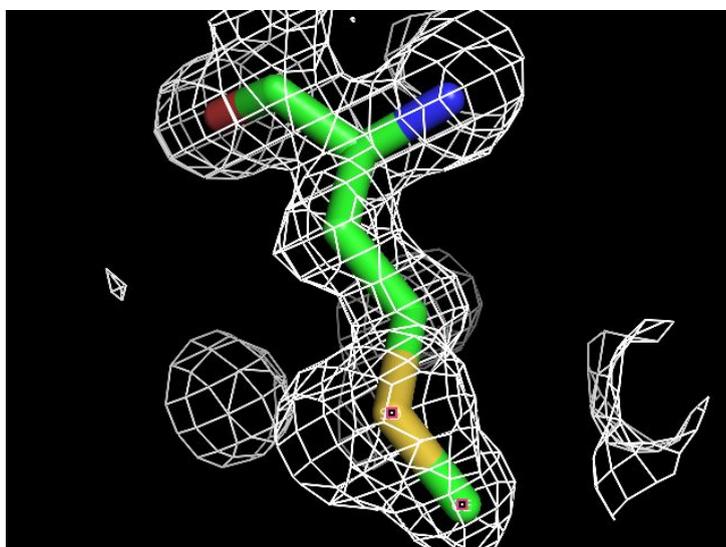
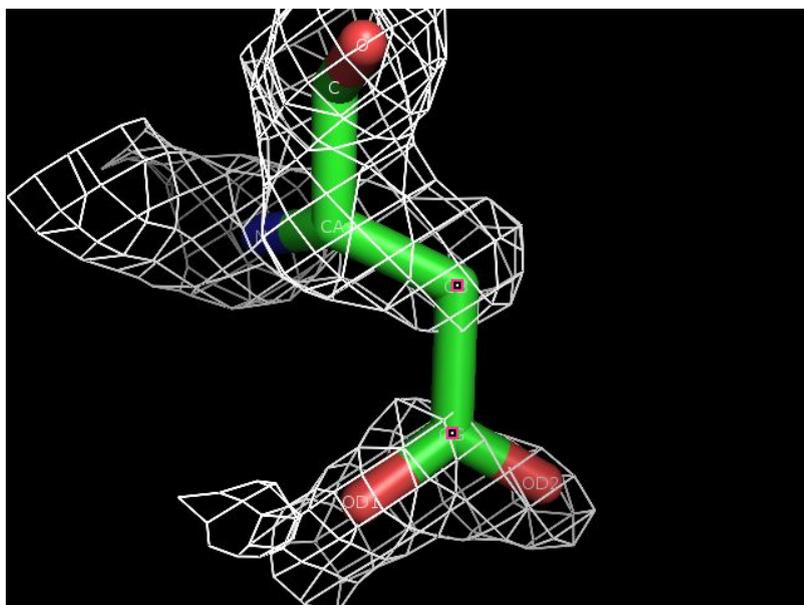


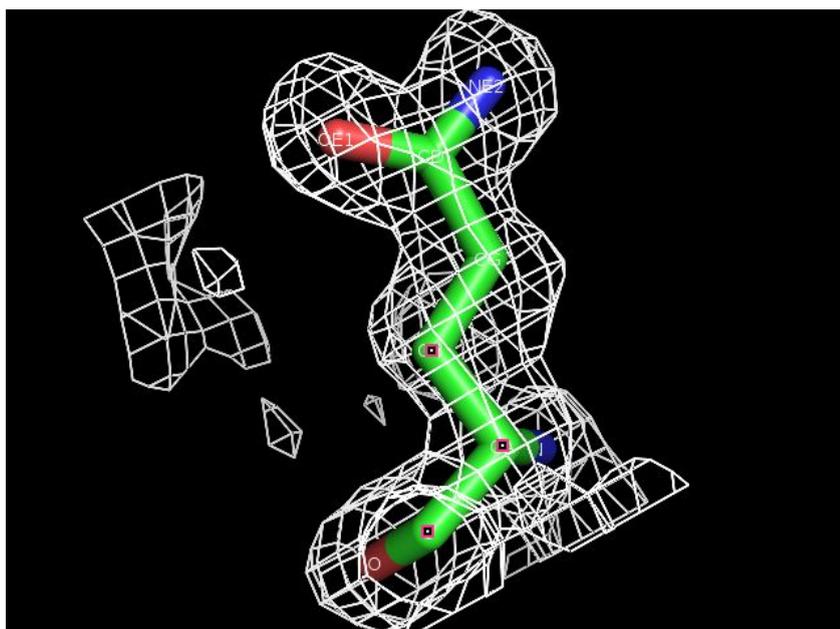
Рис. 5. Остаток MET37 в структуре 5DQY.

На рисунке 6 изображен остаток ASP60, маргинальный по длине связи CB-CG. Электронная плотность плохо ложится на остаток. Возможно, что присутствует ошибка расшифровки данных.



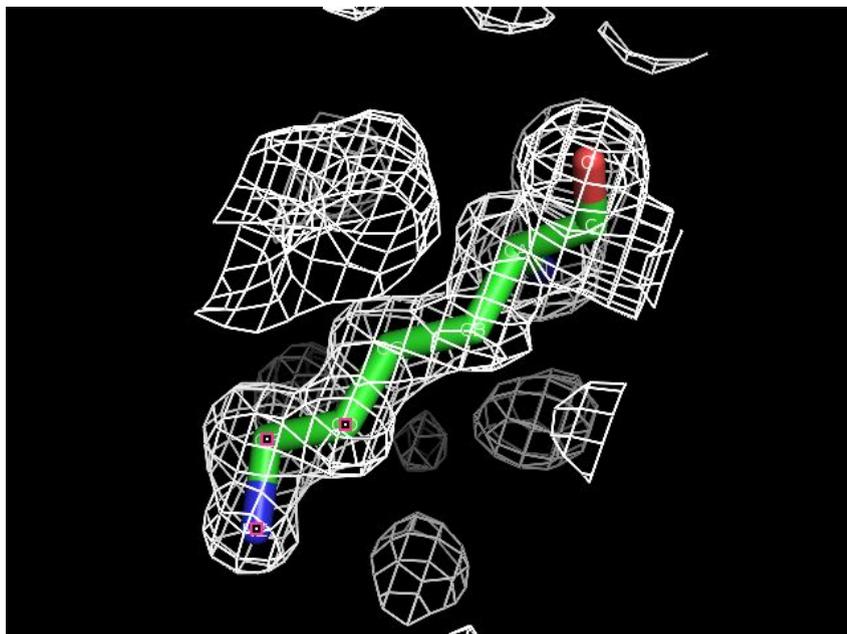
**Рис. 6.** Остаток ASP60 в структуре 5DQY.

На рисунке 7 изображен остаток GLN63, маргинальный по углу связи C-CA-CB. Электронная плотность хорошо ложится на остаток. Несовпадения данных по электронной плотности и структурой не выявлено.



**Рис. 7.** Остаток GLN63 в структуре 5DQY.

На рисунке 8 изображен остаток LYS94, маргинальный по углу связи CD-CE-NZ. Электронная плотность хорошо ложится на остаток. Несовпадения данных по электронной плотности и структурой не выявлено.



**Рис. 8.** Остаток LYS94 в структуре 5DQY.

## 5. Сравнение моделей из PDB и PDB-REDO

В базе PDB-REDO [3] проведена оптимизация структуры 5DQY, что не привело к сильному улучшению качества вписывания остатков в электронную плотность, значения R-фактора и R-free стали заметно выше.

Таблица 4. Оптимизация 5DQY на PDB-REDO.

Validation metrics from PDB-REDO		
	PDB	PDB-REDO
<b>Crystallographic refinement</b>		
<i>R</i>	0.1775	0.2070
<i>R-free</i>	0.2076	0.2377
<i>Bond length RMS Z-score</i>	1.183	0.844
<i>Bond angle RMS Z-score</i>	1.088	0.899
<b>Model quality (raw scores   percentiles)</b>		
<i>Ramachandran plot appearance</i>	98	98
<i>Rotamer normality</i>	99	99
<i>Coarse packing</i>	N/A	N/A
<i>Fine packing</i>	94	94
<i>Bump severity</i>	98	95
<i>Hydrogen bond satisfaction</i>	37	50
<i>WHAT_CHECK</i>	<a href="#">Report</a>	<a href="#">Report</a>

## Выводы

Качество структуры отличное, обладает высоким разрешением и малым количеством маргинальных остатков. Рекомендовать к использованию структуру из PDB-REDO для биоинформатических исследований вместо структуры в PDB не стоит.

## Список литературы

1. Jungwon Hwanga, Loi T. Nguyenab, Young Ho Jeon, Chan Yong Lee, Myung Hee Kima (2015), Crystal structure of fully oxidized human thioredoxin, *Biochemical and Biophysical Research Communications*, 467, 2
2. <http://eds.bmc.uu.se/cgi-bin/eds/uusfs?pdbCode=5DQY>
3. <https://pdb-redo.eu/db/5dqy>