

МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ  
имени М.В.ЛОМОНОСОВА

---

ФАКУЛЬТЕТ БИОИНЖЕНЕРИИ И БИОИНФОРМАТИКИ

Отчет по качеству рентгеноструктурного анализа белка  
репарации и рекомбинации RadA *Methanococcus voltae*

Работа студентки 4 курса  
Васюткиной Ольги Николаевны

Москва

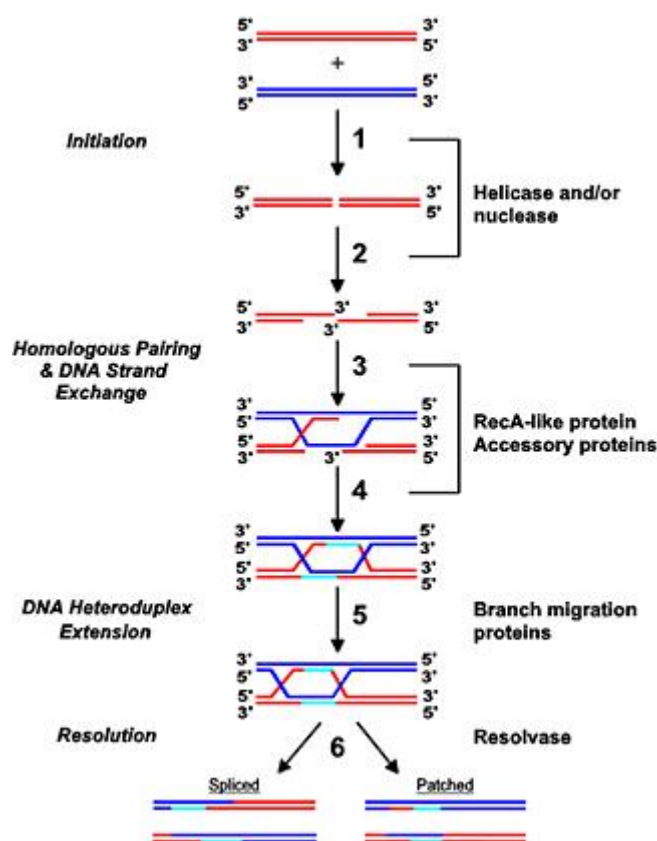
2016 г.

## Аннотация

В отчёте рассмотрены некоторые индикаторы качества модели структуры 3NTU, а также проведён анализ индикаторов локального качества структуры для ряда аминокислотных остатков. В ходе работы над отчётом удалось познакомиться с возможностями, которые предоставляют серверы для оценки качества структур белковых молекул.

## Введение

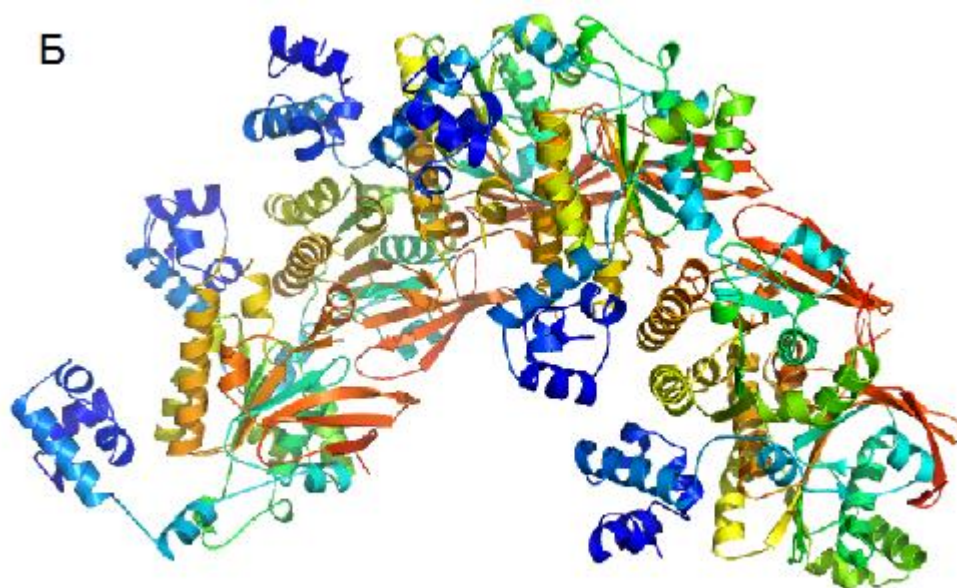
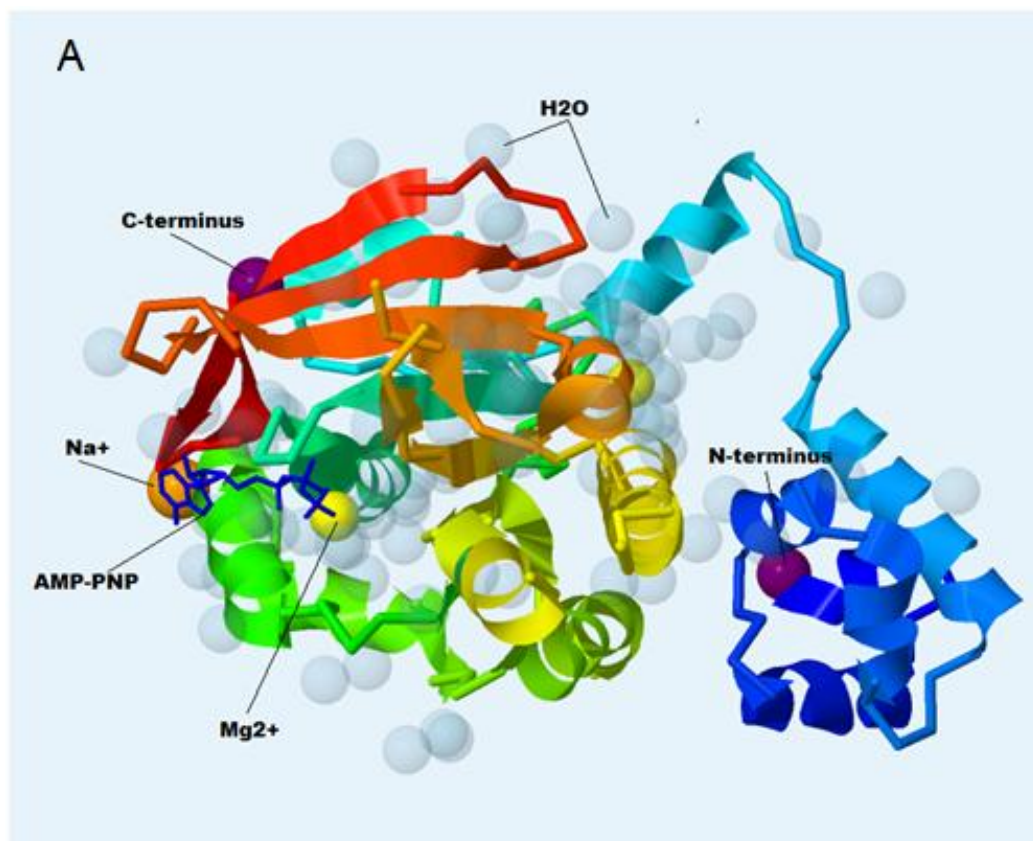
В данной работе рассмотрена структура белка репарации и рекомбинации RadA археи *Methanococcus voltae*, идентификатор PDB 3NTU [1]. При воздействии химических агентов или ионизирующего излучения в ДНК появляются двуцепочечные разрывы. Чтобы восстановить цепь ДНК и информацию, которую она несет, используется гомологичная рекомбинация с другой молекулой ДНК. Белок в присутствии АТФ связывается с одной цепью поврежденной молекулы ДНК и инициирует поиск ее двуцепочечной копии. Если она найдена, происходит гомологичная рекомбинация, в которой RadA также принимает участие, и исходная молекула ДНК восстанавливается [2]. Процесс показан на рис. 1. Данные электронной микроскопии показали, что комплекс RadA, одноцепочечной ДНК и АТФ представляет собой нуклеофиламент, состоящий из шести субъединиц RadA и 18 нуклеотидов [3].



**Рисунок 1.** Репарация двуцепочечного разрыва ДНК с помощью гомологичной рекомбинации [4].

Гомологи RadA принимают участие в аналогичных процессах у бактерий (RecA) и эукариот (Rad51). Было показано, что последовательности RadA и Rad51 идентичны на 40%, а RadA и RecA – на 20% [2].

Структура 3NTU (рис. 2) была получена, чтобы определить значение консервативного остатка аспартата 302 в связывании АТФ. Этот остаток наблюдается в последовательностях RadA и Rad51. Судя по структуре, он нужен для образования солевого мостика с АТФ. Но у RecA в этой позиции находится консервативный остаток лизина, который не может образовать солевой мостик. Оказалось, что мутантный белок RadA D302K обладает сниженной АТФазной активностью, но образует более стабильные нуклеофиламенты [5].



**Рисунок 2.** Структура белка RadA, идентификатор PDB 3NTU: мономер с подписанными лигандами (A), биологическая единица из шести мономеров (Б).

## Результаты и обсуждение

### *Общая информация о модели*

Асимметрическая единица ЗНТУ представлена одной цепью (А) из 319 аминокислотных остатков. Биологическая единица состоит из шести идентичных субъединиц. В структуре белка есть лиганд AMP-PNP (идентификатор PDB ANP), который представляет собой структурный аналог АТФ и ионы  $K^+$ ,  $Mg^{2+}$ .

Модель структуры ЗНТУ была размещена в PDB в 2010 году, выпущена в 2011 и опубликована в 2012 году [5]. Общая информация о модели была получена с использованием сервиса EDS [6] и систематизирована в таблице 1.

Таблица 1. Общая информация о модели

Разрешение, Å	1.90	
Диапазон разрешения, Å	39,35 – 1,88	
Число измеренных рефлексов	34519	
Число использованных рефлексов	31633	
Полнота данных	99,0%	
Кристаллографическая группа	P 6 <sub>1</sub>	
Параметры ячейки	Длина, Å	Угол, °
	84,970	90,00
	84,970	90,00
	104,320	120,00

Для решения фазовой проблемы использовался метод молекулярного замещения с использованием структуры нативного белка, полученной ранее (1T4G, [2]).

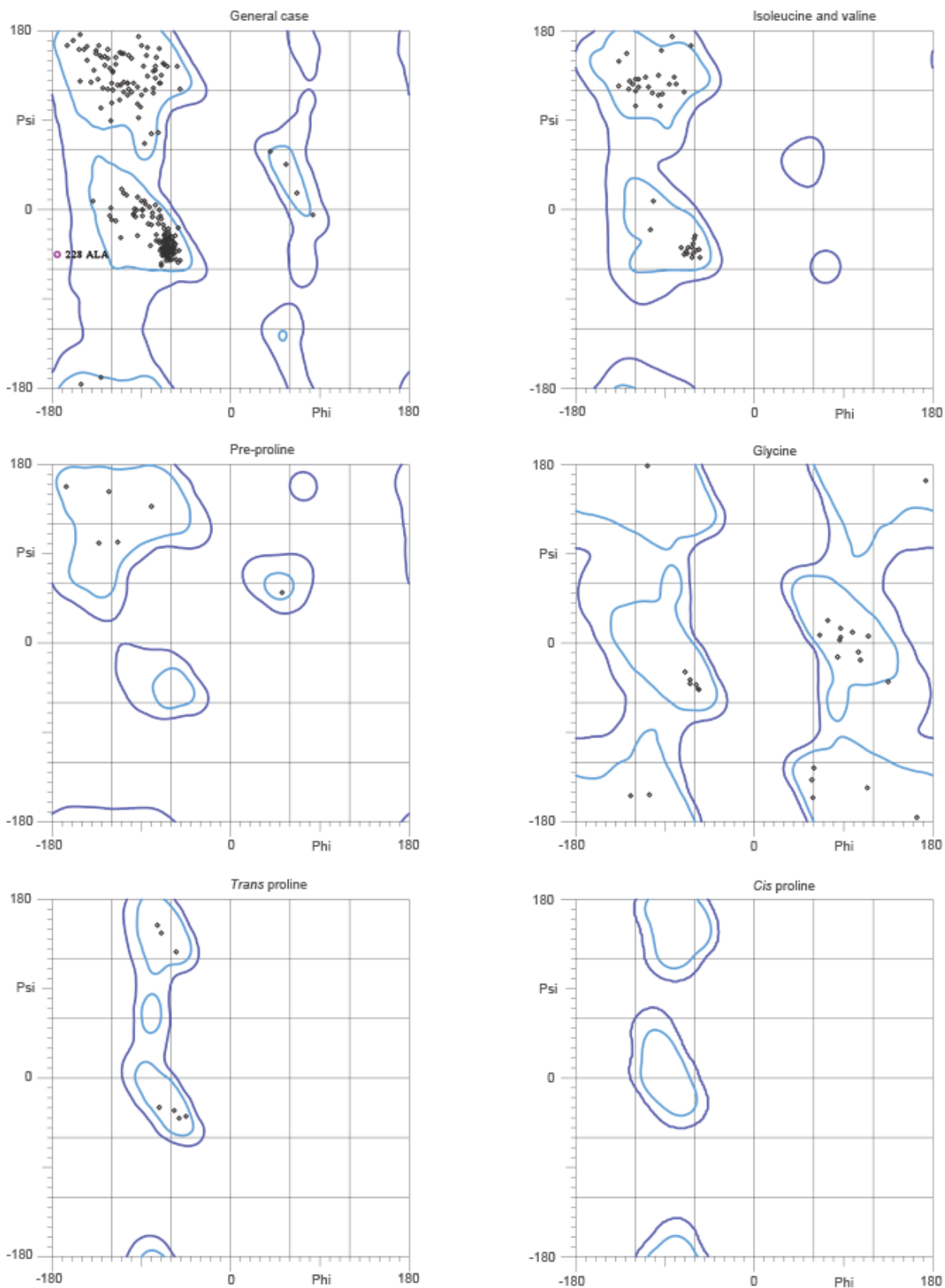
### *Индикаторы качества модели в целом*

Для данной структуры значение R-фактора равно 0,188, а R-free составляет 0,214. Если принять, что хорошие значения R-free – меньше 20%, а плохие – больше 40%, то такое значение R-free можно назвать

удовлетворительным. Разность R-free и R-фактора составляет 0,026, что свидетельствует о том, что модель не была переоптимизирована.

Для построения карты Рамачандрана был использован сервис MolProbity [7]. Карты приведены на рис. 3. Обнаружился один остаток из запрещенной области карты Рамачандрана, это аланин с порядковым номером 228, это 0,3% от всех остатков. MolProbity анализировал только 304 из 319 остатков структуры. 298 остатков из 304 (98%) попали в предпочитаемую область карты Рамачандрана, 303 из 304 (99,7%) находятся в разрешенной зоне. Процент остатков в предпочитаемой области соответствует критерию хорошей модели, но это не исключает возможность подгонки модели. Также MolProbity выдал количество «неправильных» торсионных углов и связей (таблица 2).

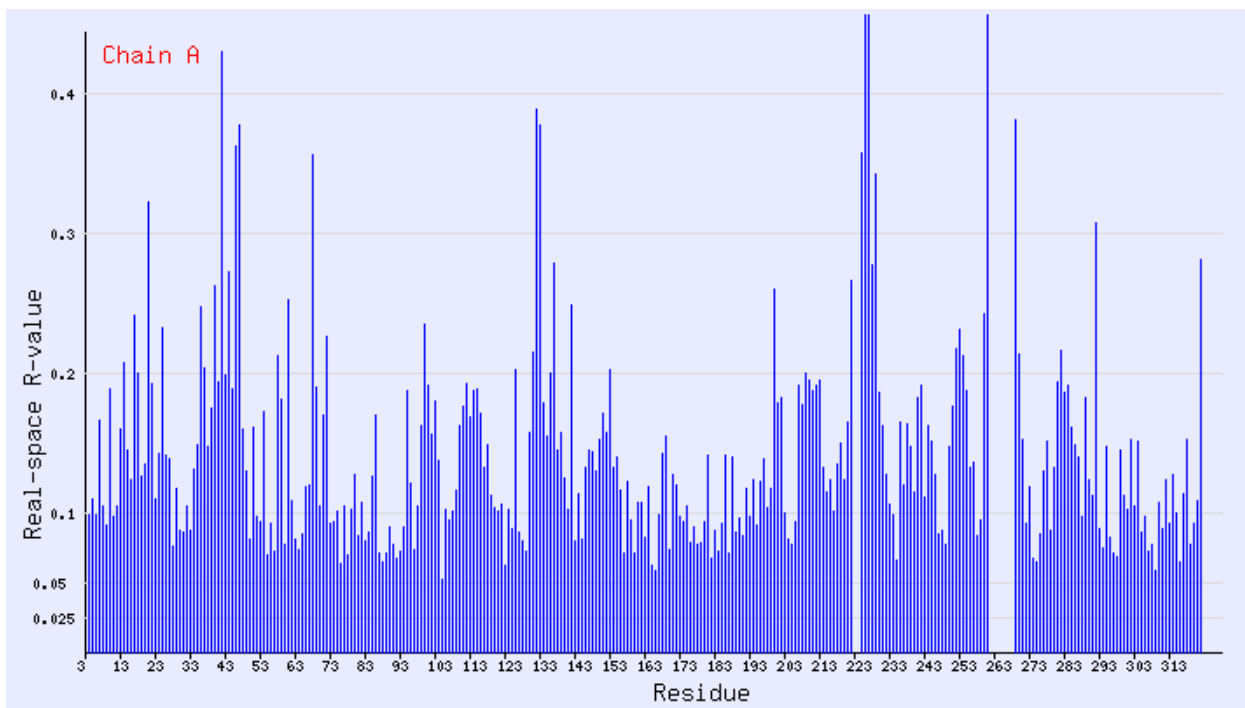
Стоит отметить, что в файле PDB есть информация о двух остатках, которые попадают в запрещенную зону карты Рамачандрана. Это ALA 228, найденный MolProbity, и SER 83. У остатка серина угол  $\phi$  равен 83,9; угол  $\psi$  равен -5,9. Это значение находится на самой границе разрешенной и запрещенной зон.



**Рисунок 3.** Карты Рамачандрана для структуры 3NTU. Получено с помощью MolProbity [7].

Пространственный R-фактор (Real Space R, RSR) характеризует, насколько модель соответствует «экспериментальной» электронной

плотности. Среднее значение RSR составляет 0,186, что опять можно назвать удовлетворительным значением. Все остатки со значением RSR больше 0,2 можно рассматривать как маргинальные. Таких обнаружилось 43 остатка, это 13,9% всей последовательности. Карта RSR для каждого аминокислотного остатка структуры приведена на рис. 4.



**Рисунок 4.** Карта RSR для структуры 3NTU. Получено с помощью EDS [6]

Для RSR каждого остатка можно определить Z-score, взяв параметры распределения (среднее значение и стандартное отклонение) из RSR всех остатков из всех структур с похожим разрешением. В данном случае использовались все структуры с разрешением 1,8 Å – 2,0 Å. Если RSR плохой, а RSR-Z – хороший, то значит, координаты атомов расшифрованы плохо, но не хуже, чем в других подобных структурах. Высокие значения Z-score ( $Z > 2$ ) свидетельствуют о том, что остаток плохо вписан в электронную плотность по сравнению с другими структурами с близким разрешением. Карта RSR-Z для каждого аминокислотного остатка структуры приведена на рис. 5. Средний RSR-Z равен 0,5. Это значит, что некоторые из остатков вписаны в электронную плотность не хуже, чем в других моделях с близким разрешением. В данном случае маргинальных остатков 29, это 9,35%.



В таблицу 2 сведены все описанные выше индикаторы качества модели для структуры 3NTU. В таблице 3 приведено 12 примеров маргинальных остатков: по карте Рамачандрана, или  $RSR > 0,4$ ; или  $RSR-Z > 3$ .

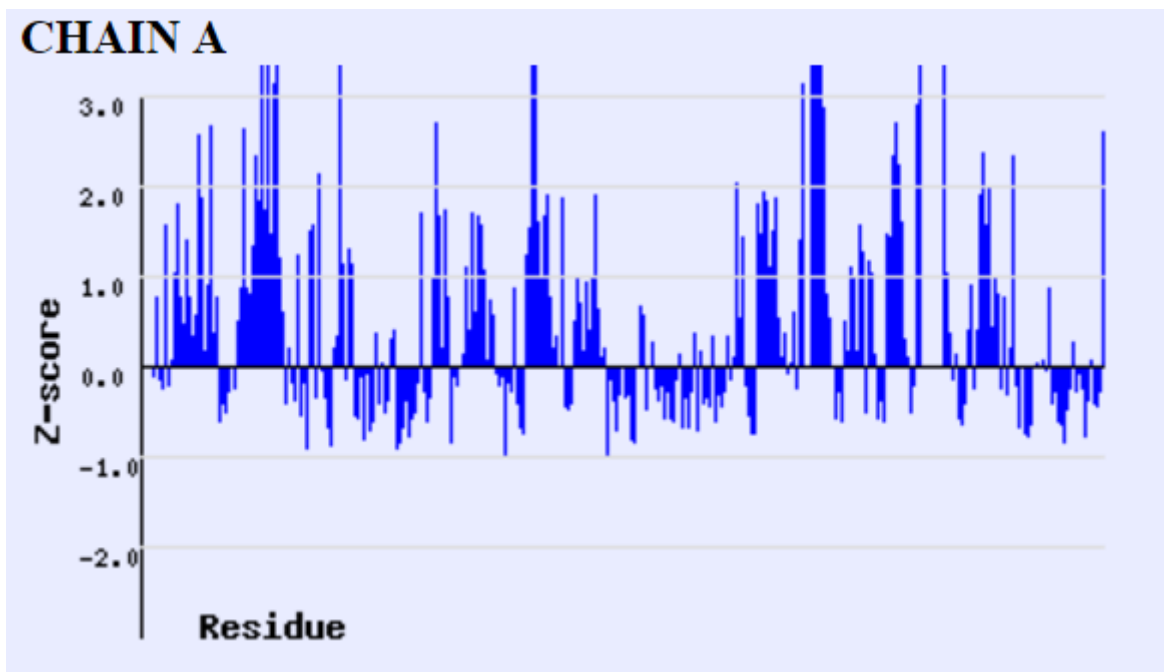


Рисунок 5. Карта RSR-Z для структуры 3NTU. Получено с помощью EDS [6]

Таблица 2. Величины некоторых индикаторов качества структуры в целом для модели 3NTU

R-фактор	0,188
R_free	0,214
R - R_free	0,026
Ротамеры в предпочитаемой зоне карты Рамачандрана	298/304 (98%)
Ротамеры в разрешенной зоне карты Рамачандрана	303/304 (99,7%)
Ротамеры в запрещенной зоне карты Рамачандрана	1/304 (0,3%)
Отклоняющиеся от нормы ковалентные связи	12/2441 (0,49%)
Отклоняющиеся от нормы валентные углы	10/3292 (0,30%)
RSR	0,186
Остатков с $RSR > 0,2$	43/310 (13,9%)
RSR-Z	0,5
Остатков с $RSR-Z > 2$	29/310 (9,4%)

Из рисунков 3–4 следует, что, как правило, маргинальные остатки образуют кластеры. Области с недостаточной электронной плотностью имеют размеры около 5 остатков.

Каждая программа по неизвестным причинам не анализировала несколько аминокислотных остатков. Сервис MolProbity выдал результат для 304 из 319 остатков, а EDS – для 310. На картах EDS заметны области, где RSR не посчитан: пропущены остатки 223–224, 262–268. Рядом с пропущенными остатками наблюдаются большие значения RSR и RSR-Z.

Оказалось, что в файле структуры PDB действительно нет информации об остатках 223–224, 262–268. В статье авторов структуры нет никакого упоминания этого факта. Только написано, что участок от Asp-256 до Arg-285 является высокоупорядоченным [5].

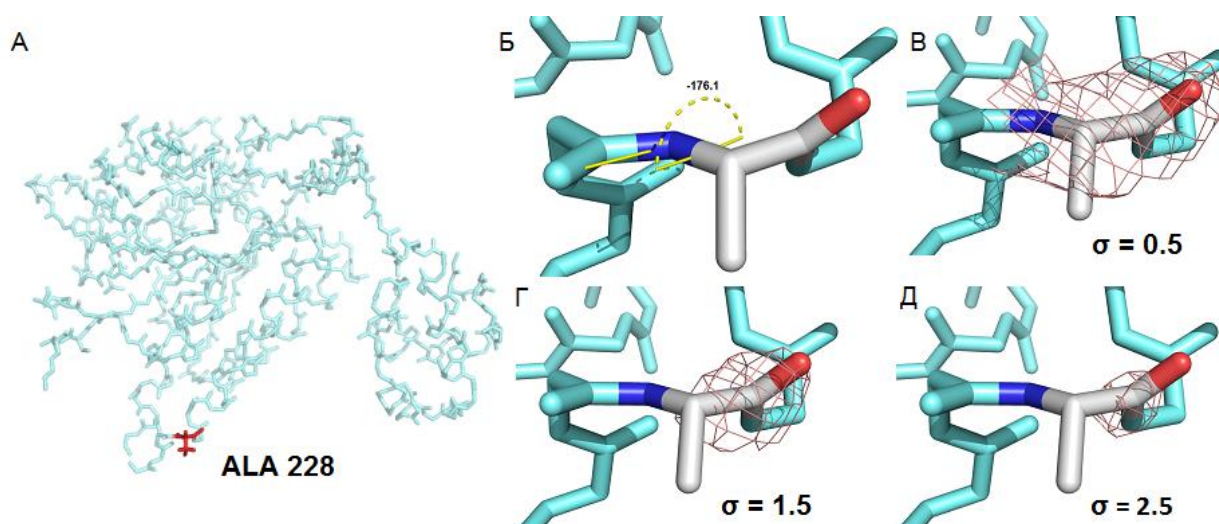
Таблица 3. Маргинальные аминокислотные остатки

Остаток	RSR	RSR-Z	Комментарий
ALA 228	0,278	3,63	В запрещенной зоне карты Рамачандрана. Выбран для детального изучения.
SER 83	0,080	-0,52	В запрещенной зоне карты Рамачандрана (файл PDB, EDS). Выбран для детального изучения.
GLU 42	0,430	4,11	
LYS 226	0,485	4,90	
LEU 227	0,467	8,00	
LYS 261	0,476	4,77	Рядом с разрывом в структуре. Выбран для детального изучения.
ALA 269	0,381	5,74	Рядом с разрывом в структуре. Выбран для детального изучения.
ILE 68	0,356	5,43	Выбран для детального изучения.
ILE 44	0,273	3,58	
LYS 47	0,377	3,33	
GLU 132	0,389	3,52	
GLU 133	0,378	3,37	

### *Детальный анализ некоторых маргинальных аминокислотных остатков*

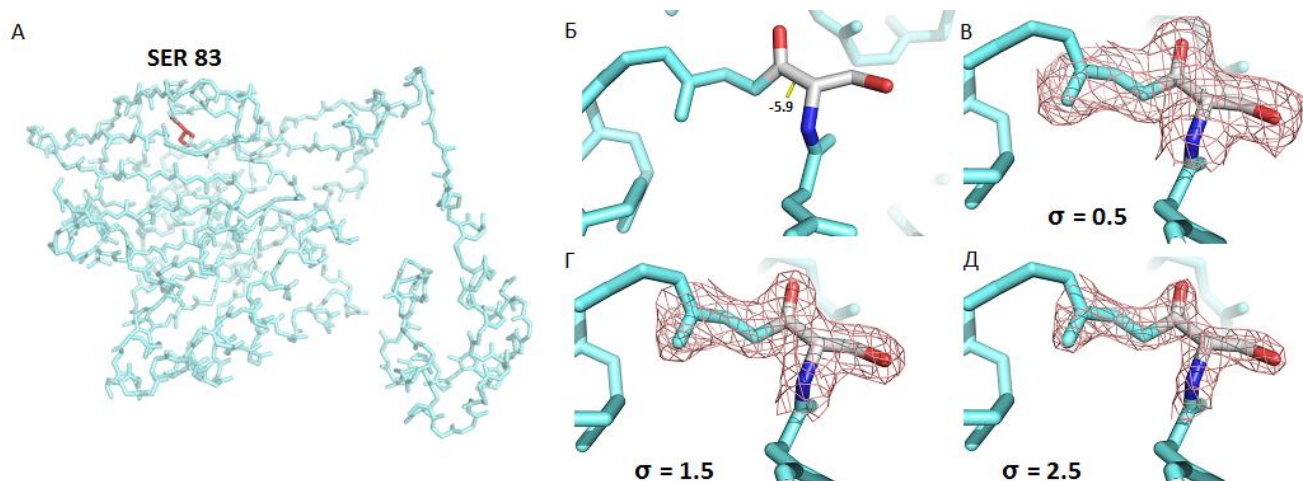
Рассмотрим, насколько маргинальные аминокислотные остатки, выделенные в таблице 3, вписываются в электронную плотность структуры.

Первый остаток – ALA 228. Он находится в запрещенной зоне карты Рамачандрана и имеет высокие значения RSR и RSR-Z. Как правило, остатки в запрещенной зоне карты Рамачандрана предсказаны ошибочно, но бывают и случаи, когда это есть на самом деле по определенной причине. Как показано на рис. 6А, остаток ALA 228 находится во внешней петле белка, что ставит под сомнение значимость его напряженной структуры. Более того, этот остаток плохо вписан в электронную плотность (рис. 6Б–Д), что также доказывает ошибочность модели в этом месте.



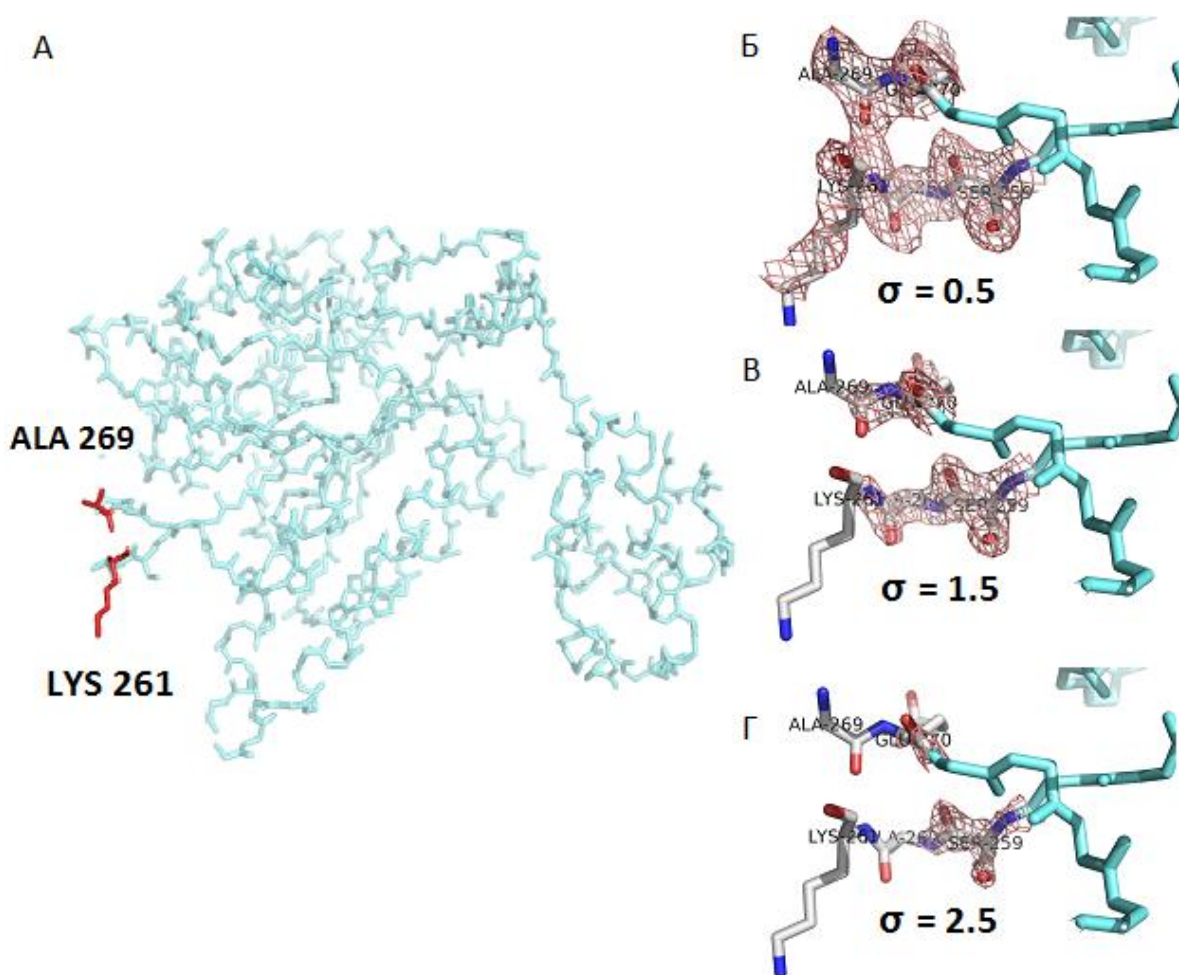
**Рисунок 6.** Маргинальный остаток ALA 228 в структуре 3NTU: положение на структуре (А), значение угла  $\phi = -176.1$ , из-за которого остаток относится к запрещенной зоне карты Рамачандрана (Б), электронная плотность вокруг остатка при различных порогах: 0,5 (В), 1,5 (Г), 2,5 (Д).

Второй остаток – SER83. Сервис EDS относит его к запрещенной зоне карты Рамачандрана, а MolProbity – к разрешенной. Находится в структуре на изгибе между бета-листом и альфа-спиралью. На рисунке 7 видно, что остаток хорошо вписан в электронную плотность, это подтверждают и хорошие значения RSR и RSR-Z. Скорее всего, остаток не до конца оптимизирован.



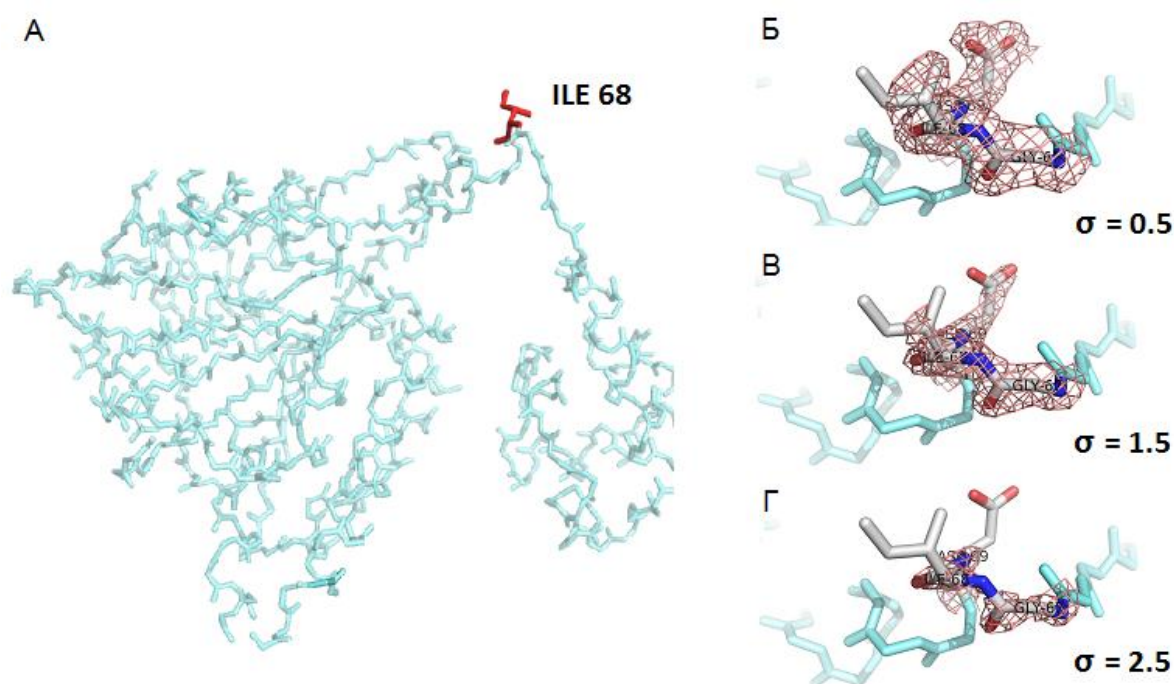
**Рисунок 7.** Маргинальный остаток SER 83 в структуре 3NTU: положение на структуре (А), значение угла  $\psi = -5,9$ , из-за которого остаток относится по версии EDS к запрещенной зоне карты Рамачандрана (Б), электронная плотность вокруг остатка при различных порогах: 0,5 (В), 1,5 (Г), 2,5 (Д).

Далее рассмотрим остатки LYS 261 и ALA269. Они находятся в петле на поверхности глобулы, на границе с разрывом в полипептидной цепи (рис. 8А), который не объяснен авторами структуры [5]. На уровне  $0,5\sigma$  остатки вписаны в электронную плотность, а при  $1,5\sigma$  и  $2,5\sigma$  совсем не вписаны (рис. 8Б–Г). При этом на месте разрыва также наблюдается электронная плотность при  $0,5\sigma$ . Таким образом, для остатков LYS 261 и ALA 269 положение определено неточно, так как сигнал неотличим от шума. В действительности они могут располагаться вдали от определенных в модели мест, тем более что где-то между ними располагаются 7 аминокислотных остатков.



**Рисунок 8.** Маргинальные остатки LYS 261 и ALA269 на границе с разрывом в структуре 3NTU: положение на структуре (А), электронная плотность вокруг остатков при различных порогах: 0,5 (Б), 1,5 (В), 2,5 (Г).

Пятый остаток – ILE 68. Он интересен тем, что соседние с ним остатки имеют хорошие значения RSR и RSR-Z: у остатка GLY 67 RSR=0,120; RSR-Z=0,33; у остатка ASP 69 RSR=0,190; RSR-Z=1,14. Располагается этот остаток в петле, соединяющей два домена. На рисунке 9 видно, что у гидрофобного изолейцина не очень комфортное окружение, он экспонирован наружу. На уровне  $0,5\sigma$  заметно, что боковая цепь не полностью вписана в электронную плотность, при увеличении  $\sigma$  до  $2,5$  сигнал у боковой цепи пропадает полностью (рис. 9Б–Г). Хорошие значения GLY 67 можно объяснить тем, что небольшой остаток проще вписать в электронную плотность. Положение остатка может быть связано с решением фазовой проблемы методом молекулярного замещения, или с оптимизацией его по, например, предпочитаемой области карты Рамачандрана.



**Рисунок 9.** Маргинальный остаток ILE 68 в структуре 3NTU: положение на структуре (А), электронная плотность вокруг остатка при различных порогах: 0,5 (Б), 1,5 (В), 2,5 (Г).

#### *Сравнение модели из PDB с моделью из PDB\_REDO*

Было проведено сравнение модели 3NTU из PDB с оптимизированной моделью из PDB\_REDO [8]. По выложенным в PDB структурным факторам сервис не смог определить параметр  $R_{\text{free}}$ , так как авторы структуры выложили только рабочий сет структурных факторов. Это стоит иметь в виду при оценке качества модели по  $R_{\text{free}}$ . Поэтому  $R_{\text{free}}$  определялся заново по 5% случайно выбранным структурным факторам. Значения некоторых параметров для оценки качества двух моделей приведены в таблице 4.

Таблица 4. Некоторые параметры модели 3NTU из PDB и из PDB\_REDO

Индикатор качества структуры	PDB	PDB_REDO
R-фактор	0,188	0,1695
$R_{\text{free}}$	0,214	0,2023
$R_{\text{free}} - R$	0,026	0,328
$Z(R_{\text{free}})$	5,43	-0,73
Доноров/акцепторов протона без	13	19



Значения R-фактора и R<sub>free</sub> улучшились, но разница их возросла. Впрочем, она все равно невелика. Любопытно заметить, что значительно улучшился параметр Z-score (R-free), хотя неизвестно, как соотносятся исходный тестовый сет и новый, определенный программой PDB\_REDO. Также 60 аминокислотных остатков заметно улучшили свое положение относительно электронной плотности, а 2 остатка заметно ухудшили. Таким образом, можно утверждать, что новая модель лучше подогнана под экспериментальные данные. Но при этом увеличилось число доноров и акцепторов протона, не образующих водородной связи.

3NTU представляет собой структуру с хорошим разрешением. Однако её качество можно оценить как «удовлетворительное», о чём свидетельствуют индикаторы оценки качества этой структуры. Возможно, это связано с тем, что авторы хотели изучить роль точечной мутации в белке. Из наблюдаемых аномалий можно отметить высокие показатели RSR-фактора и Z-score (RSR), никак не объясненные разрывы в полипептидной цепи. В модели структуры удалось также обнаружить остатки, находящиеся в запрещенной области карты Рамачандрана. Стоит отметить, что сервер PDB\_REDO значительно улучшил такие показатели модели, как R-фактор и R<sub>free</sub>, но при этом улучшились не все параметры структуры.

## Ссылки

- [1] RSCB Protein Data Bank, (n.d.).
- [2] Y. Wu, Y. He, I.A. Moya, X. Qian, Y. Luo, Crystal structure of archaeal recombinase RADA: a snapshot of its extended conformation, *Mol. Cell*, 15 (2004) 423–435.
- [3] S. Yang, X. Yu, E.M. Seitz, S.C. Kowalczykowski, E.H. Egelman, Archaeal RadA protein binds DNA as both helical filaments and octameric rings, *J. Mol. Biol.*, 314 (2001) 1077–1085.
- [4] [Frontiers in Bioscience 3, d570-603, June 17, 1998], (n.d.).
- [5] R. Amunugama, Y. He, S. Willcox, R.A. Forties, K.-S. Shim, R. Bundschuh, Y. Luo, J. Griffith, R. Fishel, RAD51 Protein ATP Cap Regulates Nucleoprotein Filament Stability, *J. Biol. Chem.*, 287 (2012) 8724–8736.
- [6] G.J. Kleywegt, M.R. Harris, J.Y. Zou, T.C. Taylor, A. Wählby, T.A. Jones, The Uppsala Electron-Density Server, *Acta Crystallogr. D Biol. Crystallogr.*, 60 (2004) 2240–2249.
- [7] V.B. Chen, W.B. Arendall, J.J. Headd, D.A. Keedy, R.M. Immormino, G.J. Kapral, L.W. Murray, J.S. Richardson, D.C. Richardson, MolProbity: all-atom structure validation for macromolecular crystallography, *Acta Crystallogr. D Biol. Crystallogr.*, 66 (2010) 12–21.
- [8] R.P. Joosten, F. Long, G.N. Murshudov, A. Perrakis, The PDB\_REDO server for macromolecular structure model optimization, *IUCr*, 1 (2014) 213–220.