

Оценка качества расшифровки белка-регулятора транскрипции семейства репрессоров TetR бактерии *Syntrophus aciditrophicus SB*

Подготовлено Погорельской Александрой

Введение

Белки семейства TetR (Tet Repressor Protein) вовлечены в контроль транскрипции генов, отвечающих за проведение биосинтеза антибиотиков (Tet от тетрациклин), участвуют в реакции на изменение осмотического давления [1].

Я анализировала такой репрессор бактерии *Syntrophus aciditrophicus SB*. На сайте UniProt есть ссылка только на одну структуру в базе данных PDB, ее идентификатор 3S5R [2].

Общая информация о структуре

Структура репрессора была размещена на сайте PDB 15 июня 2011 года, в качестве автора указан Joint Center for Structural Genomics [3]. Эксперимент проводился для двух длин волн (0.91837 и 0.97936) так как в качестве метода решения фазовой проблемы использовался метод аномального рассеяния [4]. В структуре есть несколько атомов селена в составе селенметионина.

В файле PDB (3S5R) хранится структура гомодимера, то есть белок состоит из двух идентичных цепей (по 215 аминокислотных остатков), которые сближаются С-концами. Однако количество лигандов не совпадает и, не смотря на то что аминокислотная последовательность двух цепей совпадает, цепь В образует на 1 альфа-спираль меньше (этим объясняется совпадение файлов с биологической и асимметричной единицами - две цепи отличаются по структуре). Всего в белке-регуляторе транскрипции 13 α -спиралей в составе цепи А и 12 - цепи В.

Ссылки на статью с сайта базы данных PDB нет, есть упоминание, что статья будет опубликована. Структура была размещена в базе в 2011 году, а статьи до сих пор нет. Это затрудняет поиск информации о рентгеноструктурном эксперименте, с помощью которого была расшифрована структура.

Основная информация о структуре, полученная с сайта PDB представлена в таблице 1. Разница между значениями R-фактора и свободного R-фактора небольшая, что говорит об отсутствии переоптимизации при расшифровке.

Таблица 1. Общая информация о структуре 3S5R

Параметр	Значение
Разрешение	2.6 Å
R-фактор	0.243
Свободный R-фактор	0.255
Измеренное число рефлексов	16290
Полнота данных	98.2%
Число рефлексов для построения R-free	826
Параметры ячейки	Ортогональная a = 74.43 Å b = 110.72 Å c = 126.03 Å

Оценка качества расшифровки структуры

Судить о качестве расшифровки структуры можно по нескольким показателям, наиболее распространенными являются R-фактор (показатель близости экспериментальных данных к модели), свободный R-фактор (показатели степени оптимизации), а также число маргинальных аминокислотных остатков по разным критериям (то есть остатков с нехарактерными показателями углов и длин связей, RSR, нехарактерным окружением и так далее). Краткое сравнение этих показателей модели с средними по банку PDB и средними по структурам с похожим разрешением представлено на рисунке 1. Приведенная статистика представляет собой перцентиль распределений значений параметра, в которую попадает наблюдаемое для данной структуры значение.

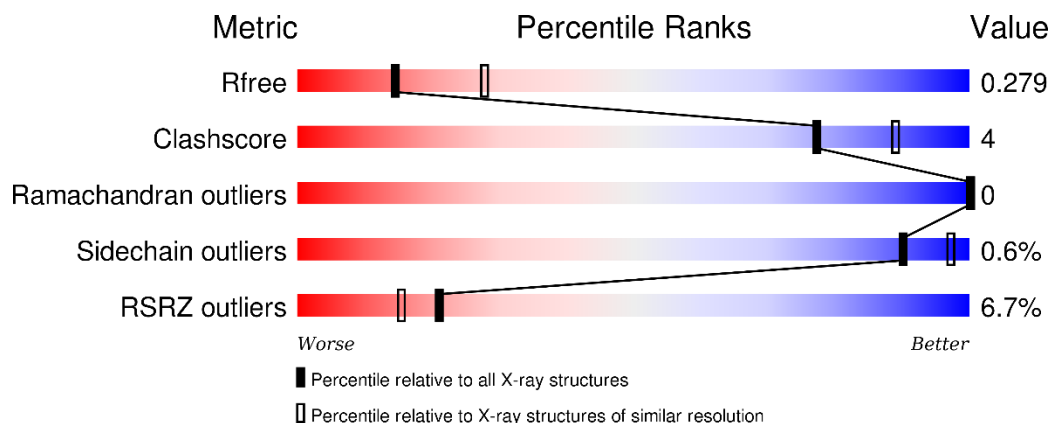


Рисунок 1. Общая характеристика качества модели. Взято с сайта PDB

Структура по этим данным содержит небольшое количество столкновений ван-дер-Ваальсовых поверхностей, незначительное количество несоответствий базам ротамеров (0.6%) и совсем не содержит выбросов согласно картам Рамачандрана. Значение RSRZ (соответствие электронной плотности (ЭП) и полноатомной модели) значительно больше ожидаемого и для всех структур PDB и для структур с таким же разрешением.

Данные валидации, представленные на сайте EMBL-EBI приведены в таблице 2 [5]. Они соответствуют значениям рисунка 1.

Таблица 2. Маргинальные остатки по данным сайта EMBL-EBI

Параметр	Описание
Нетипичные углы связей в белке	0 из 4018 (0%)
Нетипичные длины связей в белке	0 из 2972 (0%)
Несоответствия ЭП и атомов белка	25 из 377 (6.63%)
Несоответствия картам Рамачандрана	0 из 391 (0%)
Несоответствия базе ротамеров	2 из 310 (0.65%)

В более развернутом отчете о валидации есть данные о 8 остатках (5 принадлежащих цепи А и 3 – цепи В), находящихся в возможной, но неблагоприятной конформации согласно картам Рамачандрана [6]. Сервис MolProbity подтверждает эти значения (рисунок 2).

Для цепи А остатков, маргинальных по принадлежности ротамерам не найдено, а для цепи В таких остатков 2 (таблица 3).

Больше всего маргинальных остатков по несоответствию ЭП и расположения атомов (7 остатков из 25 – 6% с наибольшими значениями RSRZ представлены в таблице 3). Из 25 маргинальных остатков 14 расположены на цепи А, однако, как видно в таблице 3, остатки с самыми большими отклонениями принадлежат цепи В. Пример таких остатков (Val39 и Gly40) представлен на рисунке 3. При чем в экспериментальную электронную плотность не списан даже остов остатка глицина. Значения RSRZ для этих аминокислотных остатков составляют 5.0 и 4.7 соответственно.

Также есть достаточно много наложений ван-дер-Ваальсовых радиусов (в таблице 4 приведены первые 10 случаев перекрытия, в которых оно составляет более 0.5 Å). Слишком близкие контакты показаны для обоих включенных в структуру молекул дигидроксиэтилового эфира (PEG), которые образуют 5 наложений оболочек атомов.

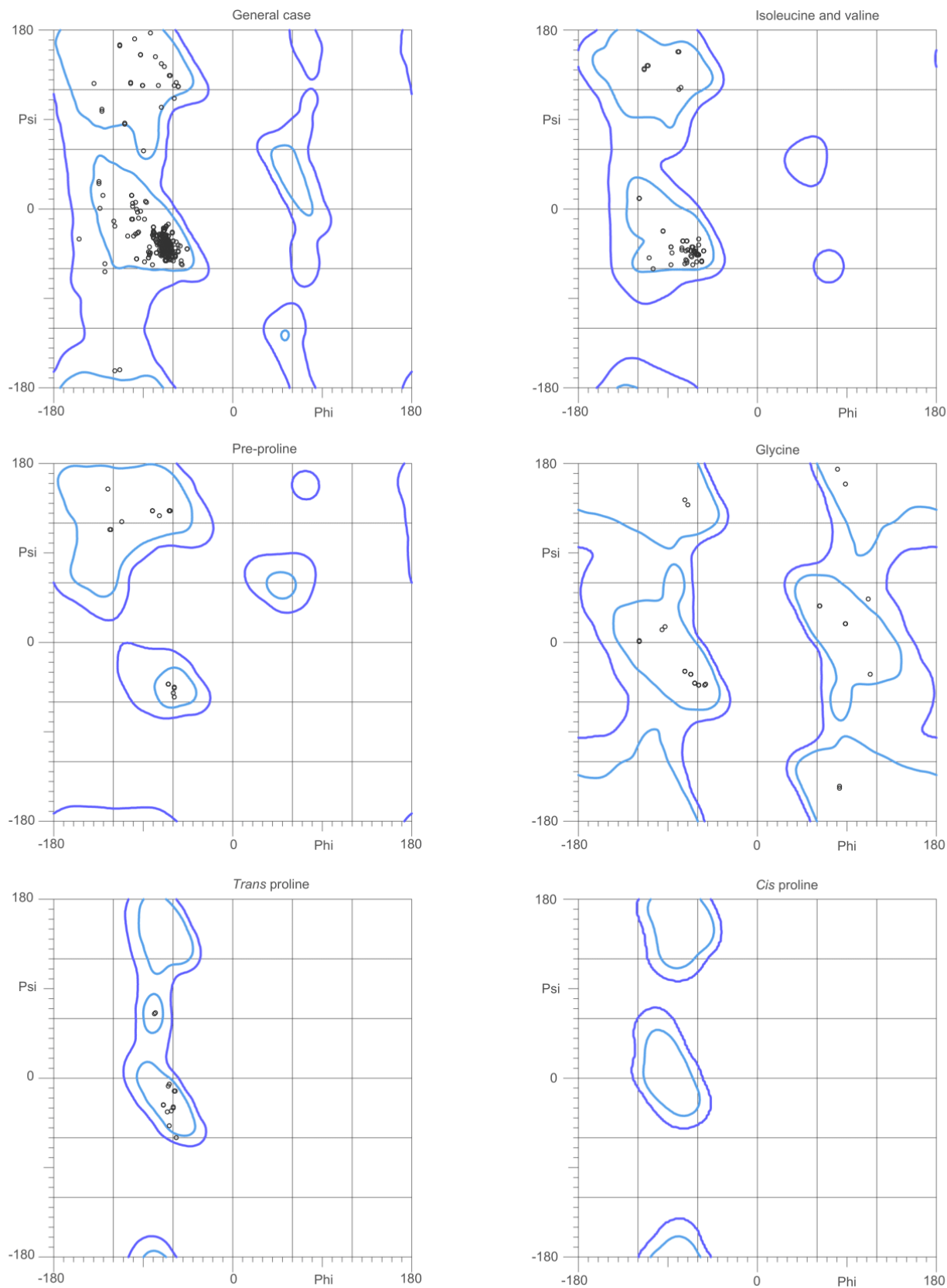


Рисунок 2. Соответствие значений углов связей остова белка картам Рамачандрана. Изображение получено с помощью сервиса MolProbity

Таблица 3. Список маргинальных остатков структуры 3S5R

Цепь	Номер остатка	Остаток	Критерий маргинальности
B	88	Arg	Ротамер
B	181	Mse	Ротамер
B	39	Val	Несоответствие ЭП
B	40	Gly	Несоответствие ЭП
B	41	Val	Несоответствие ЭП
B	100	Thr	Несоответствие ЭП
A	184	Leu	Несоответствие ЭП
A	179	Pro	Несоответствие ЭП
A	79	Gly	Несоответствие ЭП

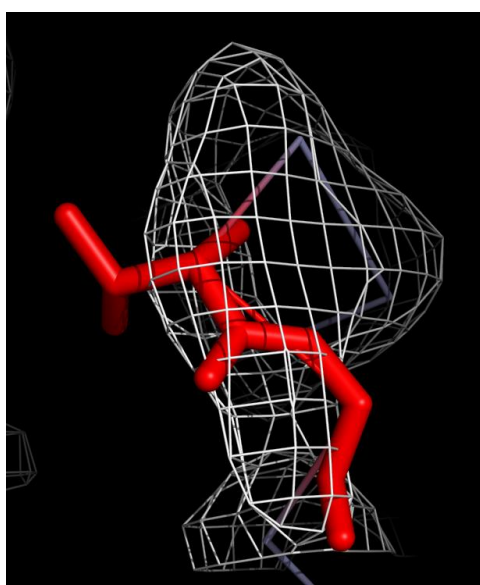


Рисунок 3. Маргинальные по соответствию экспериментальной ЭП Val39 и Gly40. Изображение получено с помощью Rutil, уровень подрезки 1σ

Таблица 4. Случаи перекрытия ван-дер-Ваальсовых радиусов атомов

Атом 1	Атом 2	Межатомное расстояние	Перекрытие
B:68:ILE:HD11	B:101:MSE*:SE	2.23	0.88
A:68:ILE:HD11	A:101:MSE*:SE	2.27	0.85
A:69:ILE:HD12	A:302:PEG** :H22	1.69	0.74
B:68:ILE:CD1	B:101:MSE*:SE	2.94	0.65
A:68:ILE:CD1	A:101:MSE*:SE	2.99	0.61
B:69:ILE:HD12	B:301:PEG** :H32	1.85	0.59
B:69:ILE:HD13	B:123:MSE*:HE1	1.88	0.55
B:76:VAL:HG13	B:86:MSE*:HE2	1.89	0.54
A:194:THR:O	A:198:PHE:HD1	1.92	0.53
B:179:PRO:HA	B:182:LYS:HE3	1.93	0.50

*MSE – селенсетионин

**PEG – Дигидроксиэтиловый эфир

В базе данных PDB-redo есть информация об оптимизации структуры по тем же экспериментальным данным [7]. Согласно этой базе, в заголовке PDB-файла 3S5R приведены не совсем честные значения R-факторов, а значительно ниже (возможно, это произошло по причине разного подбора структурных факторов, использованного для построения R-free). И при незначительной оптимизации и после полной оптимизации рабочий R-фактор существенно снизился, однако, возросла разница между ним и свободным R-фактором (таблица 5).

Таблица 5. Значения R-факторов по базе PDB-redo

	Заголовок PDB	Пересчитанные	После консервативной оптимизации	После полной оптимизации
R	0.2430	0.2701	0.2312	0.2287
R-free	0.2550	0.2843	0.2787	0.2924

Всего было улучшено RSRZ для 127 аминокислотных остатков (60%), со значительным ухудшением всего 2 (0.9%) [8]. В том числе было оптимизировано расположение лигандов – молекул дигидроксиэтилового эфира. Но, например, плохое соответствие остатков Val39 и Gly40 и ЭП (рисунок 3) исправлено не было.

В случае консервативной оптимизации было также уменьшено количество доноров и акцепторов водородной связи, которые эту связь не образуют (10 вместо 17). Для полной оптимизации этот показатель был хуже, осталось 16 таких доноров и акцепторов.

Различия вторичной структуры для двух идентичных цепей димера

Структуры, соответствующие аминокислотным остаткам с 178 по 182 (Ile-Pro-Thr-Mse-Lys), которые аннотированы как α -спираль только в одной из цепей, можно увидеть на рисунке 4. Водородные связи, которые автоматически рисуются в Jmol для обеих цепей выделены черным цветом, в В-цепи они короче. В А-цепи есть третья водородная связь, выделенная зеленым (ее длина 3,145Å). Расстояние между аналогичными атомами В-цепи (на рисунке соединены более тонкой зеленой линией) равно 3,380Å.

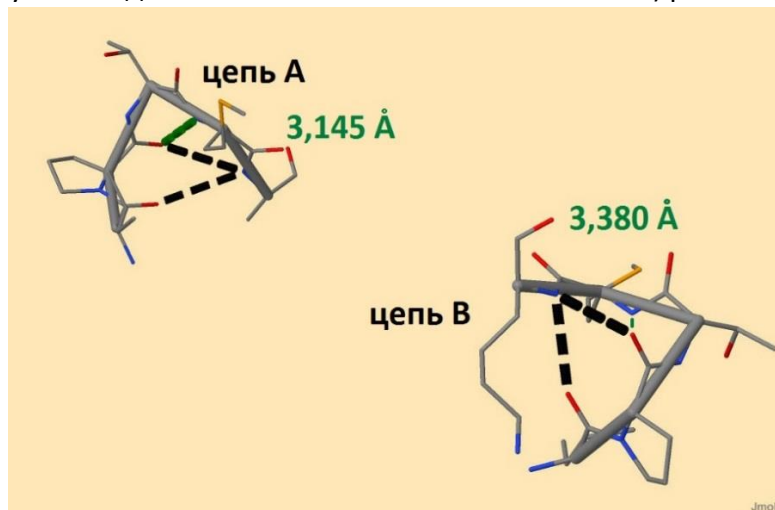


Рисунок 4. Аминокислотные остатки с 178 по 182 на обеих цепях репрессора. Черными и зелеными пунктирными линиями выделены водородные связи

Выводы

Структура 3S5R не является достаточно хорошей, ее можно значительно оптимизировать, используя те же экспериментальные данные, как это было продемонстрировано в базе PDB-redo. Количество маргинальных остатков, отобранным по геометрическим соображениям мало, но соответствие модели экспериментальным данным недостаточно. Этот параметр значительно хуже, чем в структурах с таким же разрешением.

Список литературы

1. Juan L. Ramos, Manuel Martínez-Bueno, Antonio J. Molina-Henares, Wilson Terán, Kazuya Watanabe, Xiaodong Zhang, María Trinidad Gallegos, Richard Brennan, and Raquel Tobes. The TetR Family of Transcriptional Repressors, *Microbiol Mol Biol Rev.* 2005 Jun
2. <http://www.uniprot.org/uniprot/Q2LS68>
3. <http://www.rcsb.org/pdb/explore/explore.do?structureId=3s5r>
4. <http://www.rcsb.org/pdb/explore/materialsAndMethods.do?structureId=3S5R>
5. <http://eds.bmc.uu.se/cgi-bin/eds/uusfs?pdbCode=3s5r>
6. http://ftp.wwpdb.org/pub/pdb/validation_reports/s5/3s5r/3s5r_full_validation.pdf
7. http://www.cmbi.ru.nl/pdb_redo/s5/3s5r/index.html
8. http://www.cmbi.ru.nl/pdb_redo/s5/3s5r/3s5r_dRSCC.png