

МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ИМЕНИ М.В. ЛОМОНОСОВА
ФАКУЛЬТЕТ БИОИНЖЕНЕРИИ И БИОИНФОРМАТИКИ

Отчёт по качеству расшифровки структуры белка 1A4X
организма *Bacillus Subtilis* методом рентгеноструктурного анализа

Выполнила студентка 4 курса,
Медведева Мария Витальевна

Москва, 2017.

Аннотация

Данный отчёт посвящён анализу качества расшифровки РСА белка PurR организма *Bacillus Subtilis* (идентификатор PDB 1A4X), на основе таких показателей, полученных с помощью различных сервисов и баз данных, например как параметры кристаллической решетки, R-фактор, R-free, карта Рамачандрана.

Введение

Общие сведения об объекте и его структуре

Репрессор оперона Pur (PurR) – ДНК-связывающий белок, который участвует в механизме регуляции транскрипции пуринового (Pur) оперона: ингибирует экспрессию генов, кодирующих белки оперона биосинтеза пуринов [1, 2].

Пуриновый оперон *Bacillus Subtilis*, *purEKBCSQLFMNHD*, кодирует ферменты, которые участвуют в синтезе инозин монофосфата (IMP), предшественника пуриновых нуклеотидов AMP и GMP.

Известно, что гены пуринового метаболизма у *B. subtilis* подвержены множественной регуляции на уровне инициации и терминации транскрипции. В случае *pur*-оперона показано, что их экспрессия регулируется как на уровне инициации транскрипции с участием белка-репрессора PurR, так и на уровне терминации транскрипции [3, 4, 5].

На рисунке 1 изображена регуляция экспрессии оперона с участием белка-репрессора PurR (рис. 1).

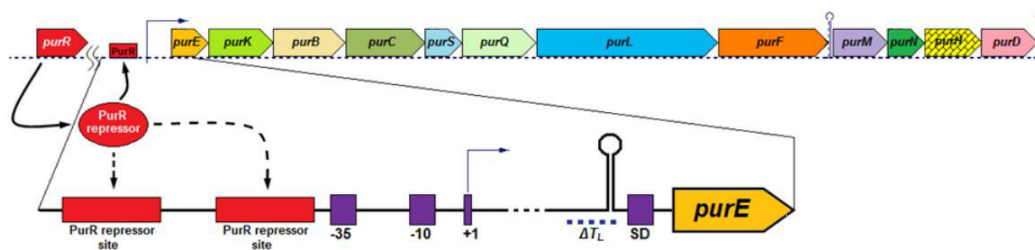


Рисунок 1. Схема структурной организации *pur*-оперона и его регуляции [6].

В верхней части рисунка показано относительное расположение 12 генов, образующих *rig*-оперон и не сцепленный с ними ген *PurR*, кодирующий белок репрессор *rig*-оперона. Внизу представлена область *rig*-оперона с сайтами связывания белка *PurR*, РНК-полимеразы (-35 – -10), старта транскрипции (+1), терминатора транскрипции (в виде шпилечной структуры) и сайта связывания рибосом (SD) [6].

Белок существует в виде димера или гексамера. *PurR* гексамер представляет собой D_3 симметричный тример из димеров [7]. Авторы статьи получили структуры двух различных форм. Для удобства в отчёте я решила остановиться только на гексамере *PurR*.

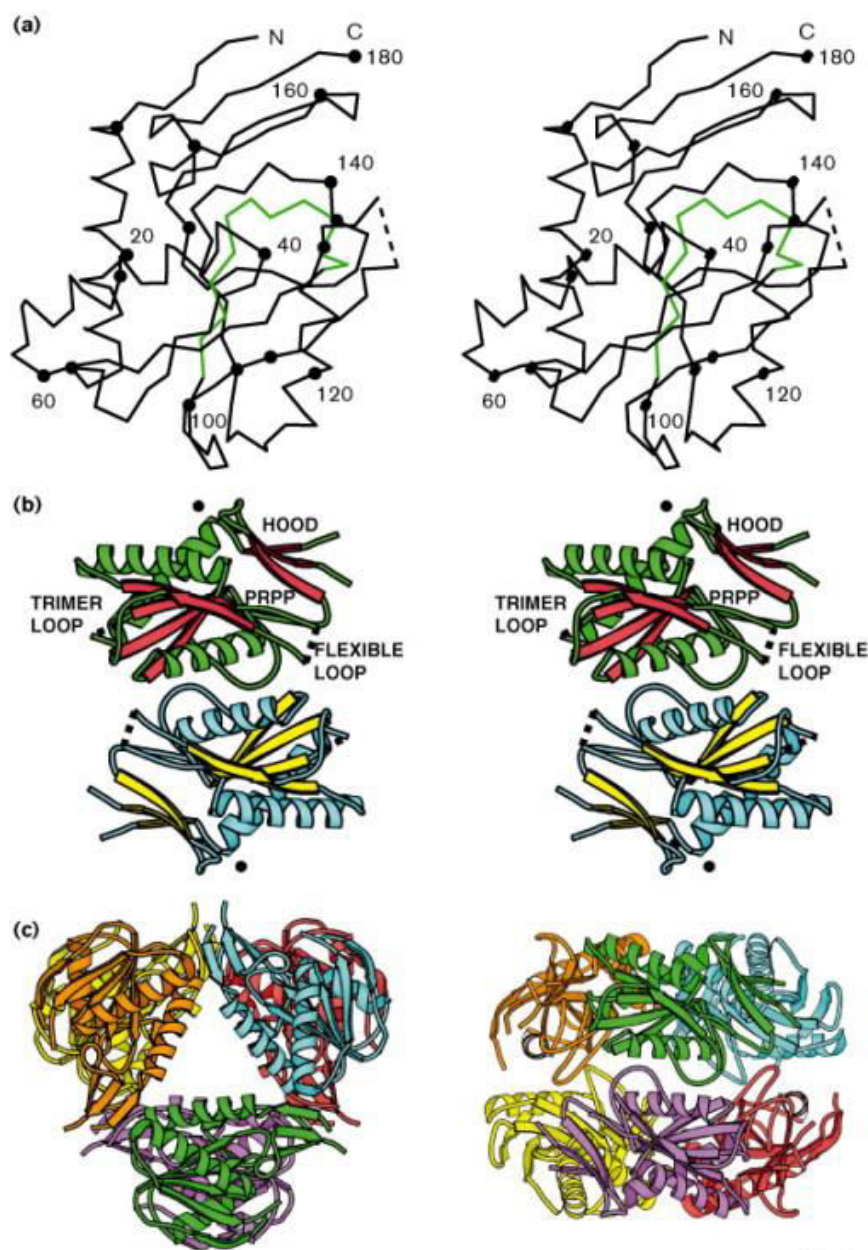


Рисунок 2. Структура PurR . а) мономер из гексамера PurR. в) димер PurR из димерной кристаллической структуры. с) гексамер PurR [7].

Расшифровка кристаллической структуры белка-репрессора PurR важна для понимания его регуляторной активности; механизма того, как PurR узнаёт специфические pur мРНК последовательности, и его связывания с pur-опероном.

Результаты и обсуждения

1. Общая информация о модели

Структура P_{yr}R была получена в 1998 году в ходе совместной работы группы ученых из Университета Пердью, Уэст-Лафейетт, штат Индиана, и Научно-исследовательского института Скриппс, Ля-Холья, штат Калифорния. расшифровку описывает статья «Adaptation of an enzyme to regulatory function: structure of *Bacillus subtilis* P_{yr}R, a pyr RNA-binding attenuation protein and uracil phosphoribosyltransferase» авторов Tomchick DR, Turner RJ, Switzer RL, Smith JL [7].

Метод решения фазовой проблемы в случае димера P_{yr}R был решён с помощью MAD (multiwavelength anomalous diffraction), многоволнового аномального рассеяния, и полученная структура P_{yr}R использовалась для решения фазовой проблемы гексамера с помощью метода MR (molecular replacement), молекулярного замещения.

Всего измерен 23 877 рефлекс.

Разрешение структуры 2.3 Å, диапазон разрешений структурных факторов 15.00 – 2.30 Å. Полнота набора структурных факторов в этом диапазоне 98.5 %.

Некоторые характеристики структуры, содержащиеся в PDB-файле, представлены ниже [7]:

CRYST1 100.580 100.580 275.150 90.00 90.00 120.00 H 3 2

Первые три числа обозначают длины направляющих векторов кристалла в ангстремах (100.580 100.580 275.150), следующие три числа – углы между ними (90.00 90.00 120.00). Кристалл принадлежит к группе симметрии H3. В ячейке находится две молекулы, то есть одна нативная структура.

2. Значения индикаторов качества модели в целом

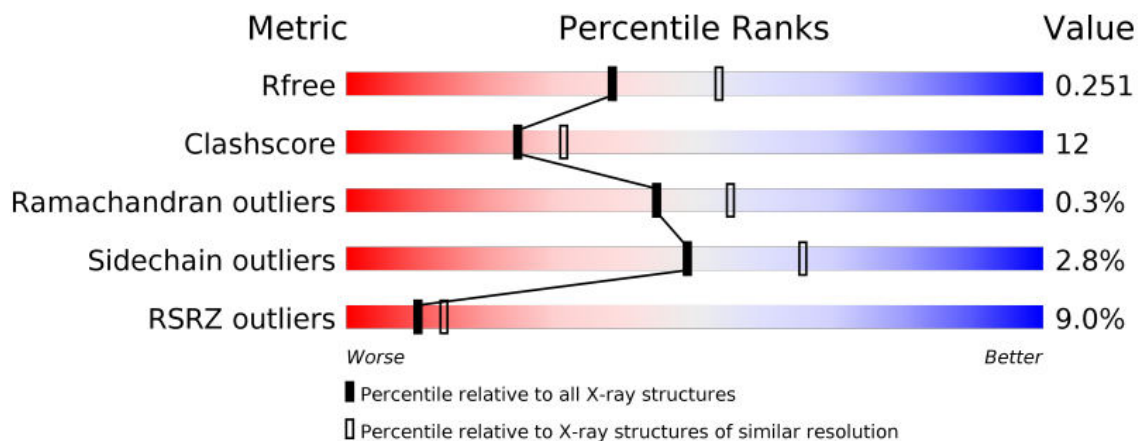


Рисунок 3. Значения различных индикаторов качества модели.

Значения индикаторов модели относительно всех PCA структур и всех структур с таким же разрешением. К сожалению, по рисунку 3 ясно, что все значения принимают средние или не очень хорошие значения, особенно это видно для относительной оценки RSRZ (Рис.3).

R-фактор структуры 1a4x равен 21,1%. Соответствие кристаллографической модели экспериментальным данным описывается R-фактором.

R-free структуры 1a4x равен 24,7%. R-free < 25%, значит, у модели высокое качество. Он более точно показывает качество модели и позволяет избежать переоптимизацию, поскольку он рассчитывается по контрольным рефлексам, не участвующим в построении модели.

RSR, пространственный R-фактор, позволяет найти маргинальные остатки по соответствию боковой цепи экспериментальной электронной плотности (Рис. 4).

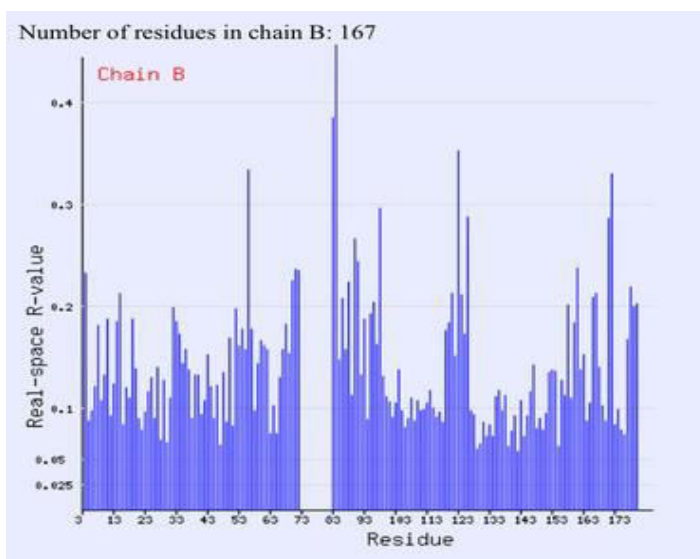
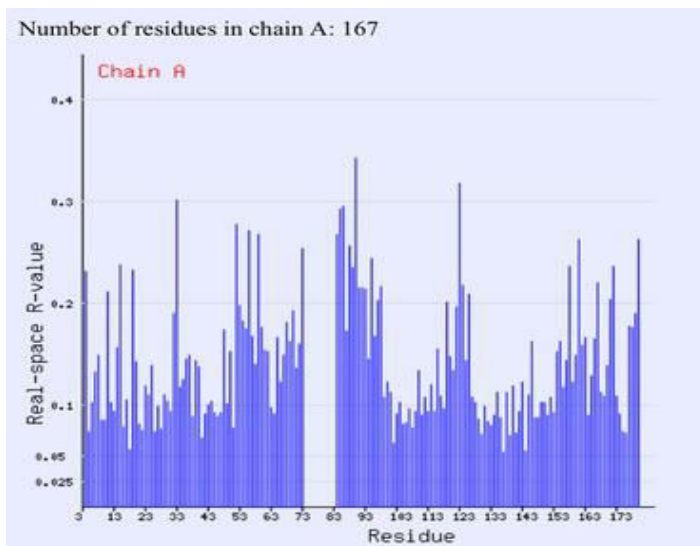


Рисунок 4. Значения RSR аминокислотных остатков, полученные с помощью сервиса EDS.

На цепи А очень выбиваются остатки ALA90, ASP123, ASN33; на цепи В высоким пиком идёт GLU84, а также сильно выделяются ASP83, опять ASP123, ASN172 и GLN56. Наличие таких резких отдельных пиков опять свидетельствует о не очень хорошем качестве структуры (Рис.4).

Относительная оценка RSRZ показывает RSR остатка в равнении с RSR аналогичных остатков. Опять на рисунке 5 видно много маргинальных остатка: ASN35, А; GLY89, А; ASP83, В; VAL122, А и В; и другие.

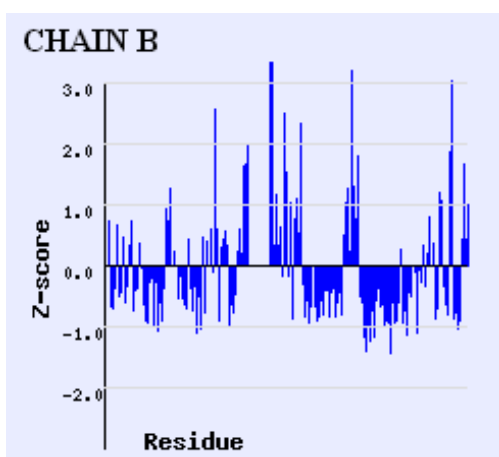
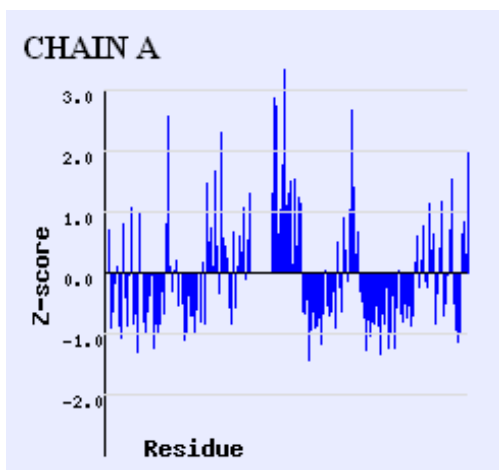
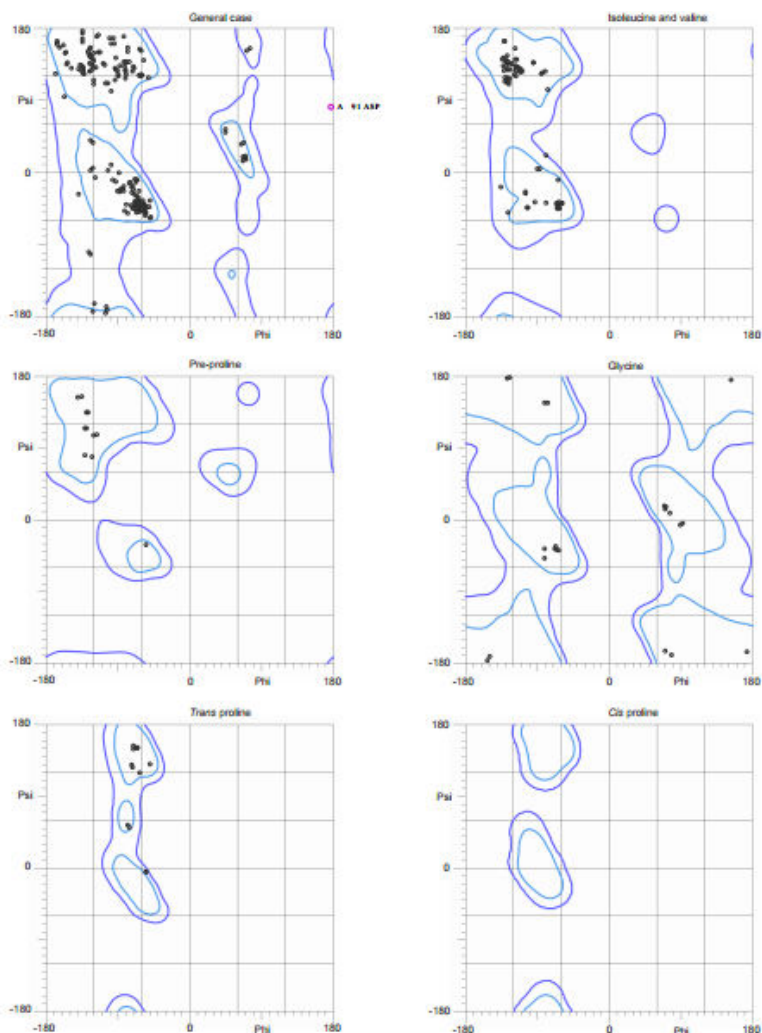


Рисунок 5. Значения RSRZ аминокислотных остатков, полученные с помощью сервиса EDS.

All-Atom Contacts	Clashscore, all atoms:	10.63	93 rd percentile* (N=355, 2.30Å ± 0.25Å)	
	Clashscore is the number of serious steric overlaps (> 0.4 Å) per 1000 atoms.			
Protein Geometry	Poor rotamers	10	3.38%	Goal: <0.3%
	Favored rotamers	281	94.93%	Goal: >98%
	Ramachandran outliers	1	0.31%	Goal: <0.05%
	Ramachandran favored	311	95.40%	Goal: >98%
	MolProbity score [^]	2.26	78 th percentile* (N=8909, 2.30Å ± 0.25Å)	
	Cβ deviations >0.25Å	2	0.62%	Goal: 0
	Bad bonds:	5 / 2690	0.19%	Goal: 0%
	Bad angles:	5 / 3635	0.14%	Goal: <0.1%
Peptide Omegas	Cis Prolines:	0 / 12	0.00%	Expected: ≤1 per chain, or ≤5%

Рисунок 6. Результат выдачи сервиса MolProbity.

По результатам выдачи сервиса MolProbity опять видны нехорошие показатели оценок (Рис.6). Например, довольно высокий Clashscore, но значение «poor rotamers» и «bad bonds» низкие (10 и 5), что хорошо.



95.4% (311/326) of all residues were in favored (98%) regions.
99.7% (325/326) of all residues were in allowed (>99.8%) regions.

Рисунок 7. Карты Рамачандрана, полученные с помощью сервиса MolProbity. Разрешенные области обведены синим цветом, благоприятные – голубым.

По картам Рамачандрана видно, что 95,4% (311/326) всех остатков находятся в благоприятных областях, 99,7% (325/326) находятся в недопустимых.

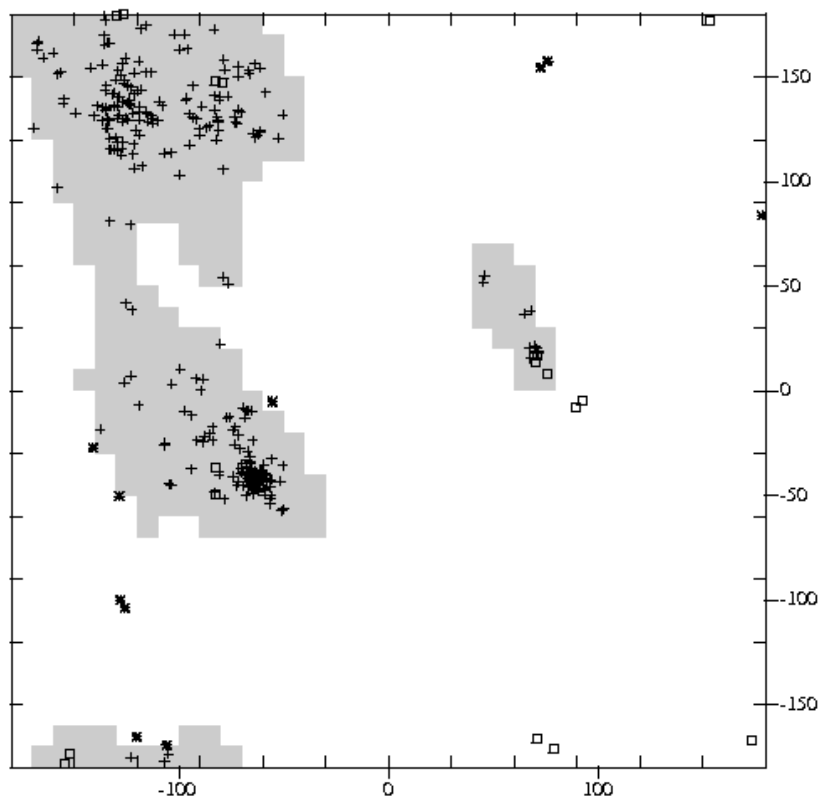


Рисунок 8. Карта Рамачандрана, полученная с помощью сервиса EDS.

Если строить карту Рамачандрана с помощью сервиса EDS, получается примерно 3,6% (11 штук) маргинальных остатков: LYS40,a; ALA90,a; ASP91,a; PRO93,a; TYR109,a; ASP123,a; LYS40,b, PRO93,b, TYR109,b; VAL169,b; ASP170,b.

3. Маргинальные остатки

#	А.к. остаток	цепь	Показатель маргинальности
1	ASN33	A	RSRZ
2	ASP123	A и B	RSRZ
3	ASP123	A	карта Рамачандрана
4	TYR109	B	карта Рамачандрана
5	LYS40	A и B	карта Рамачандрана

Таблица 1. Некоторые примеры маргинальных остатков.

ASN32

Рассмотрим просто для примера. Показатель маргинальности – нормальное значение RSRZ (0.805).

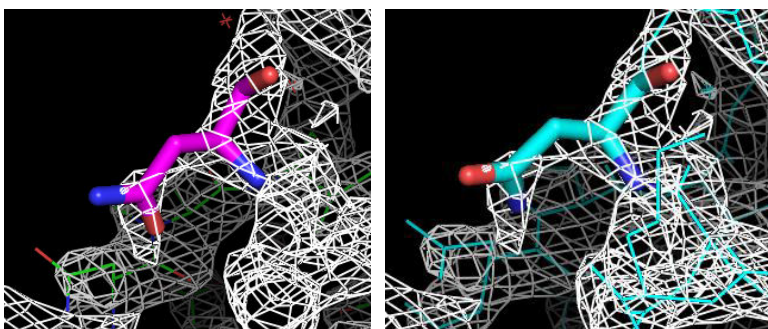


Рисунок 9. Остаток ASN32 (цепь A), фиолетовым показана структура из PDB, а голубым – из PDB_redo.

Структуры остатка из PDB и PDB_redo полностью совпадают и хорошо вписываются в электронную плотность.

ASN33

Показатель маргинальности – слишком высокие RSRZ (2.541).

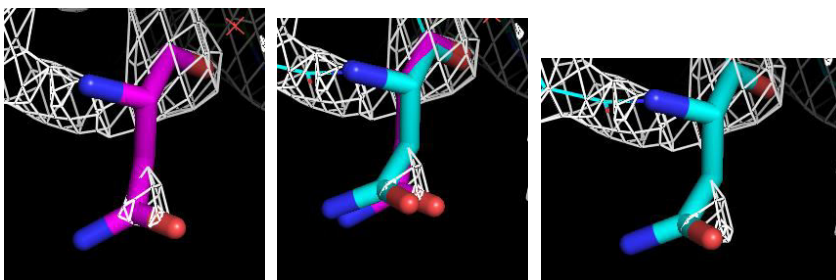


Рисунок 10. Остаток ASN33 (цепь A), фиолетовым показана структура из PDB, а голубым – из PDB_redo.

Структуры остатка из PDB и PDB_redo не очень хорошо совпадают и обе немного не вписываются в электронную плотность. Она проходит рядом с С-альфа атомом.

ASP123

Показатель маргинальности – неправильные углы ϕ и ψ : -141, -27.5 и слишком высокие RSRZ (2.66).

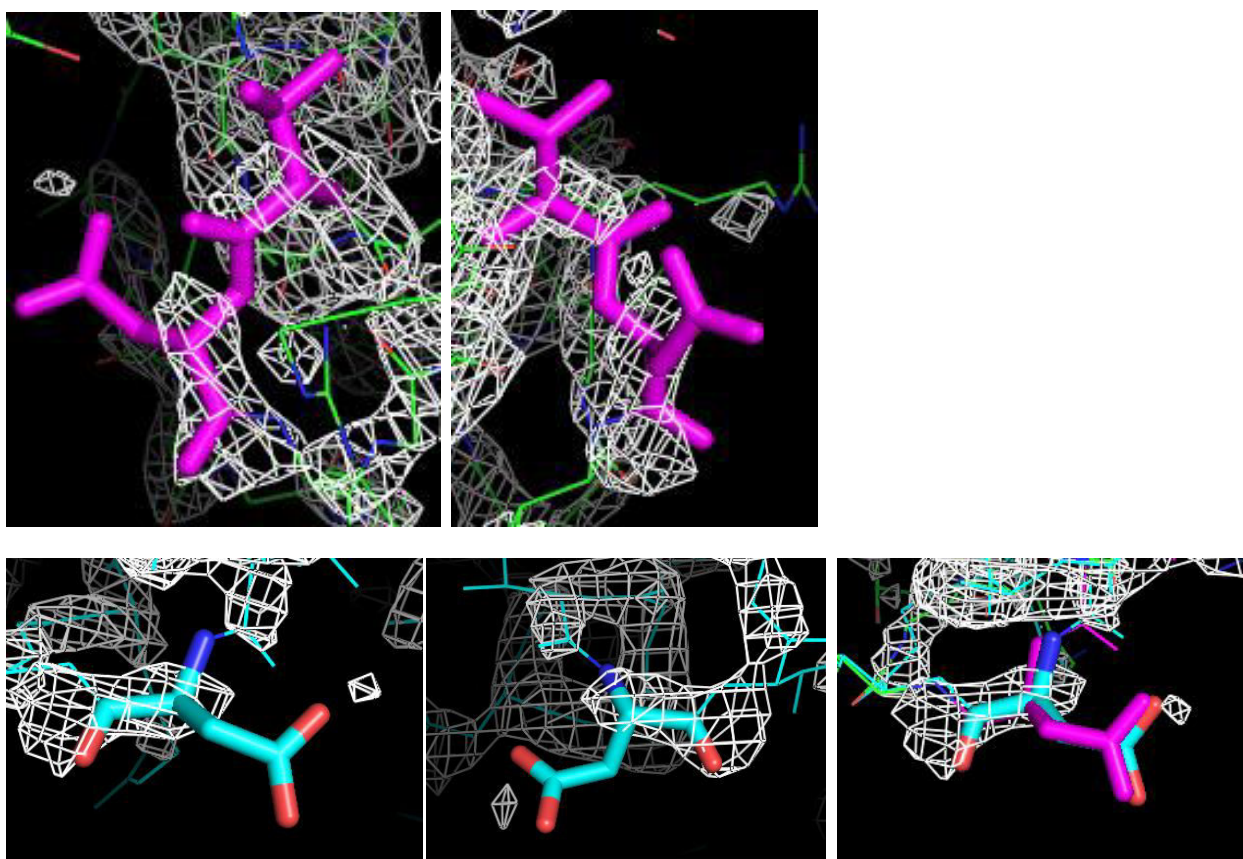


Рисунок 11. Остаток ASP123 (цепь A), фиолетовым показана структура из PDB, а голубым – из PDB_redo.

Структуры остатка из PDB и PDB_redo полностью совпадают и хорошо вписываются в электронную плотность. Причина, почему этот остаток попал в маргинальную область, возможно, связана с тем, что этот остаток находится рядом с другим маргинальным остатком VAL122, у которого довольно высокие RSRZ (1.3).

TYR109

Показатель маргинальности – неправильные углы ϕ и ψ : -125.6 и -103.8.

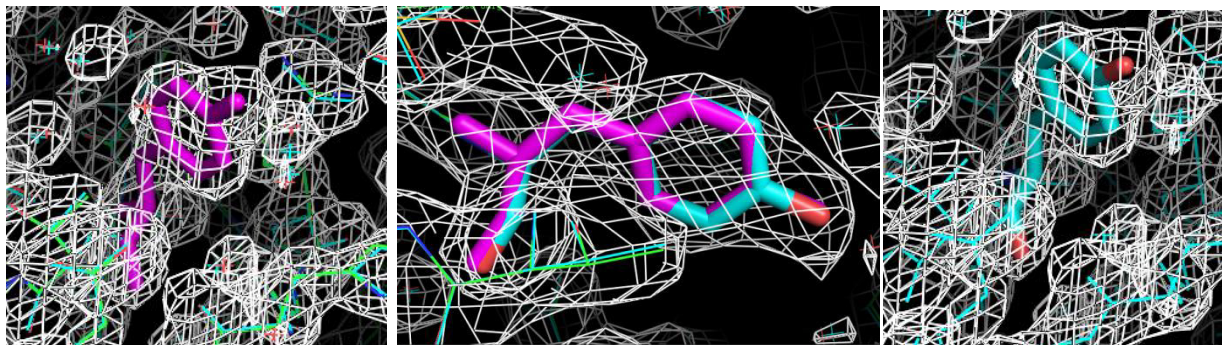


Рисунок 12. Остаток TYR109 (цепь B), фиолетовым показана структура из PDB, а голубым – из PDB_redo.

Структуры остатка из PDB и PDB_redo почти полностью совпадают и хорошо вписываются в электронную плотность.

LYS40

Показатель маргинальности – неправильные углы ϕ и ψ : 76.3 и 157.4.

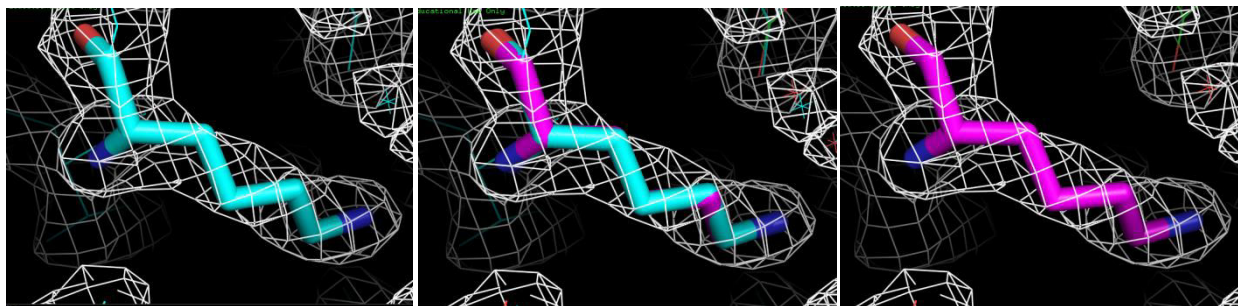


Рисунок 13. Остаток LYS40 (цепь B), фиолетовым показана структура из PDB, а голубым – из PDB_redo.

PRO93

Показатель маргинальности – неправильные углы ϕ и ψ : -55.2 и -5.2.

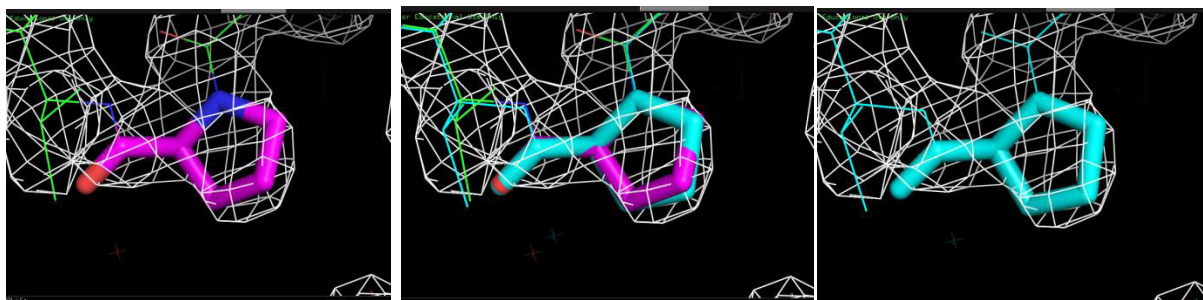


Рисунок 14. Остаток PRO93 (цепь B), фиолетовым показана структура из PDB, а голубым – из PDB_redo.

Структуры LYS40 и PRO93 из pdb и pdb_redo почему-то относятся к маргинальным, хотя хорошо списываются в электронную плотность, находятся на структурированной части белка да и совпадают друг с другом.

4. Сравнение модели из PDB с моделью PDB_redo.

По идее, модель PDB_redo является оптимизированной PDB модель, на рис. 15 видно, что эти две модели совпадают.

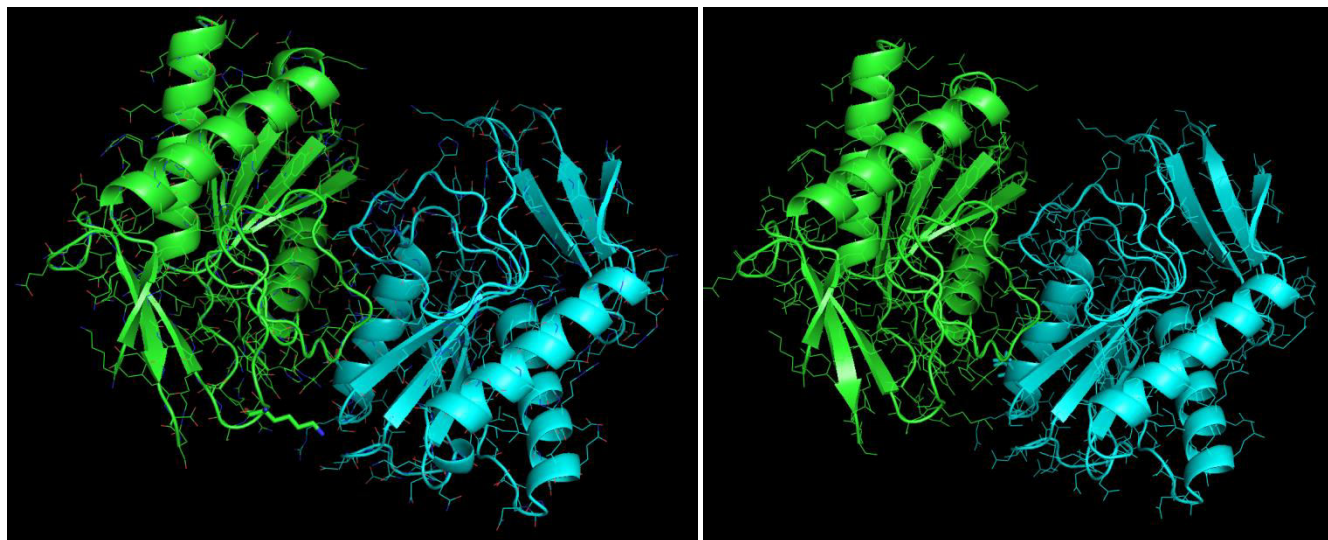


Рисунок 15. Сравнение моделей из PDB (слева) и PDB_redo (справа): структуры раскрашены по цепям, в зелёных и синий.

В базе PDB_REDO приведена сама оптимизация структуры 1a4x и её параметры. По таблице 4 видно, что для оптимизированной структуры уменьшились значения R и R_free, увеличились Z-score, т е общее качество структуры улучшилось.

Validation metrics from PDB-REDO		
	PDB	PDB-REDO
Crystallographic refinement		
<i>R</i>	0.1958	0.1713
<i>R-free</i>	0.2323	0.2117
<i>Bond length RMS Z-score</i>	0.552	0.858
<i>Bond angle RMS Z-score</i>	0.767	0.980
Model quality (raw scores <i>percentiles</i>)		
<i>Ramachandran plot appearance</i>	40	69
<i>Rotamer normality</i>	83	70
<i>Coarse packing</i>	N/A	N/A
<i>Fine packing</i>	79	81
<i>Bump severity</i>	75	70
<i>Hydrogen bond satisfaction</i>	26	32

Таблица 2. Значения при оптимизации структуры 1a4x.

Вывод

Качество расшифровки структуры 1a4x, в частности согласно выдаче сервиса MolProbity, хорошим назвать нельзя.

В структуре имеется много маргинальных остатков. Но рассмотренные остатки не являются ошибками расшифровки, их скорее можно отнести к функциональным особенностям структуры.

Оптимизированная структура, полученная с сайта PDB_REDO, более досотверна, чем первоначальная.

Использованная литература

[1] <http://www.uniprot.org/uniprot/P37551>

[2] Экспрессия генов, Ю. А. Берлин, Издательство "Наука", 2000.

[3] Zalkin, H. and Nygaard, P., Biosynthesis of Purine Nucleotides, in *Escherichia coli* and *Salmonella*, ASM, 1996.

[4] Switzer, R.L., Zalkin, H., and Saxild, H.H., Purine, Pyrimidine, and Pyridine Nucleotide Metabolism, in *Bacillus Subtilis and Its Closest Relatives: From Genes to Cells*, ASM, 2002.

[5] K. V. Lobanov, N. V. Korol'kova, S. Yu. Eremina, L. Errais Lopes, A. S. Mironov, Mutation analysis of the purine operon leader region in *Bacillus subtilis*, *Microbial Genetics*, 2011.

[6] http://www.genetika.ru/sovet2/2011_012_Lobanov_ref.pdf

[7] Tomchick, D.R., Turner, R.J., Switzer, R.W., Smith, J.L., Adaptation of an enzyme to regulatory function: structure of *Bacillus subtilis* PyrR, a pyr RNA-binding attenuation protein and uracil phosphoribosyltransferase, 1998.

[8]

https://files.rcsb.org/pub/pdb/validation_reports/a4/1a4x/1a4x_full_validation.pdf