Отчет по расшифровке структуры 4GUW

Выполнен студентом 4 курса ФББ МГУ Пензаром Дмитрием

Москва 2016 год

Введение

Кристалл

Структура 4GUW – одна из нескольких структур RIP 1 (рибосом-инактивирующего белка типа 1), полученных в результате эксперимента в 2012 группой ученых - Singh, A., Pandey, S., Kushwaha, G.S., Bhushan, A., Sinha, M., Kaur, P., Sharma, S., Singh, T.P. К сожалению, на данный момент нельзя найти ни статьи, не ее черновой версии, в которых бы рассказывалось о целях и способе получения этих структур (однако параметры эксперимента доступны на сайте PDB (http://www.rcsb.org/pdb/), о чем речь пойдет несколько дальше).

Белок был кристаллизован целиком, удалось установить положение всех тяжелых атомов аминокислот. Кристаллизация производилась в растворе pH 6.7 при температуре 298 градусов Кельвина. Белок кристаллизировался в виде комплекса с липосахаридом.

Фазовая проблема решалась молекулярным замещением (известно большое количество гомологов данного белка, из которых и была получена модель).

В ходе эксперимента было получено 33243 рефлексов, с разрешением от 1.6А до 65.12 А. Полнота данных составляет 99.7%, что позволяет также доверять указанному разрешению.

Далее было использовано 31148 рефлексов, 1681(5 % всех рефлексов) из которых в ходе оптимизации не использовались, а использовались для расчета R-free.

В результате было получено R = 0.193 и R-free = 0.223. Отсутствие сильного различия между ними позволяет доверять полученным цифрам. Полученное значение R-free приемлемо, хотя и превышает порог для очень хороших структур (0.2).

В ассиметричной ячейке, как и в биологической единице, представлен мономер белка, что соответствует данным о RIP-1.

Симметрия в кристалле НЗ (в современной номенклатуре - РЗ)

Параметры элементарной ячейки приведены в таблице 1.

Таблица 1. Параметры элементарной ячейки кристалла

Unit Cell	
Length (Å)	Angle (°)
a = 130.24	$\alpha = 90$
b = 130.24	$\beta = 90$
c = 39.82	$\gamma = 120$

О белке

RIP (Ribosome Inactivating Protein) - семейство белков, инактивирующих рибосому за счет расщепления в ней N-гликозидной связи определенного нуклеотида (в случае белков 2-го типа это A4324). Данная группа белков широко распространена у растений (печально известный рицин, находящийся, к примеру, в семенах клещевины), а также встречается у некоторых типов бактерий.

Выделяют два типа RIP – RIP-1, состоящий из одной цепи, обеспечивающей как проникновение в клетку, так и каталитическую активность (N-гликозид гидролаза), и RIP-2, в котором за эти функции отвечают две разные белковый цепи.

Как уже говорилось выше, белок, структура которого была получена, относится как раз к RIP первого типа.

По данным САТН белок состоит из двух доменов, данные о которых приведены в таблице 2.

Изображение структуры белка, покрашенной по доменам приведено на рис.1.

Коорд инаты	Домен	Класс	Архитектура	Топология	Гомология
A, 1- 162	4guwA0 1	<u>Альфа и</u> <u>Бета</u>	<u>3-</u> <u>слойный(aba)</u> <u>сэндвич</u>	1 домен А- субъединицы 1 рицина	1 домен А-субъединицы рицина
A2, 163- 246	4guwA0 2	<u>Несколько</u> вторичных структур	<u>Иррегулярная</u>	2 домен А- субъединицы рицина	2 домен А-субъединицы рицина

Таблица 2. Доменный состав белка 4guw по данным САТН



Рисунок 1. Раскраска белка 4guw по доменам. Зеленым покрашен домен 4guwA01, синим – 4guwA02.

Следует отметить, что для белка 4guw характерно наличие сравнительно большого количества спиралей типа 3-10 (пример на рисунке 2) и бета-мостов – «бета-листов», состоящих из двух аминокислот соединенных связями, как в бета-листе (рисунок 3). На рисунке 4 показано предсказание вторичной структуры при помощи STRIDE, т.к для DSSP хорошей визуализации найдено не было, а выдачи программ практически не отличаются.



Рисунок 2. 3-10 спираль, образующийся между остатками 120-125 белка, предсказание получено при помощи сервиса DSSP (<u>http://www.cmbi.ru.nl/dssp.html</u>).



Рисунок 3. Бета-мост, образующийся между 127ILE и GLU 179, предсказание получено при помощи сервиса DSSP (<u>http://www.cmbi.ru.nl/dssp.html</u>)



Рисунок 4. Разметка, полученная при помощи сервиса STRIDE (<u>http://webclu.bio.wzw.tum.de/cgi-bin/stride/</u>)



По данным Clud(<u>http://mouse.belozersky.msu.ru/npidb/cgi-bin/hftri.pl</u>) (порог 5A) структура имеет один гидрофобный кластер (рисунок 5).

Рисунок 5. Гидрофобный кластер в структуре, результаты получены при помощи программы Clud(<u>http://mouse.belozersky.msu.ru/npidb/cgi-bin/hftri.pl</u>).

Главные параметры для оценки качества расшифровки структуры

Для оценки качества структуры мною использовался генерируемый самим сайтом PDB Validation report и сервис с. Наиболее распространенные общие индикаторы качества приведены на **рисунке 5**. Из него видно, что наша структура во многом лучше, чем остальные структуры из PDB-банка.



Рисунок 6. Параметры качества структуры 4guw по сравнению с другими структурами из PDBбанка.

В случае с сервисом MolProbity можно получить более детализированную информацию о нашей структуре (**Таблица 3**). Важно отметить, что она несколько отличается, т.к сервис убрал три инверсии боковой цепи в модели, речь о которых пойдет дальше. Видим, что в нашей структуре, по мнению сервиса, недостаточно аминокислот, находящихся в предпочитаемой зоне (на самом деле, на мой взгляд, критерий >98% слишком строг в данном случае). Отчасти это должно объясняться наличием в полученной структуре большого числа спиралей 3-10 типа. Кроме того видно, что у нас много аминокислот, для которых углы боковых цепей не находятся в предпочитаемой зоне. Это можно объяснить отчасти тем, что электронная плотность на боковых цепях в нашей структуре в целом ниже, чем на основной цепи (смотри ниже). Сервис считает, что в нашей структуре много "плохих" связей и углов, то есть сильно отличающихся от своих оптимальных параметров. Следует, однако отметить, что большая часть плохих связей и углов дается лигандами, а не самим белком, что вполне приемлемо.

Таблица 3. Общая характеристика структуры 4guw с исправленными инверсиями боковой цепи, сервис MolProbity.

All-Atom	Clashscore, all atoms:	5.72		91 st percentile [*] (N=718, 1.60Å ± 0.25Å)
Clashscore is the number of serious ste				ic overlaps (> 0.4 Å) per 1000 atoms.
	Poor rotamers	3	1.44%	Goal: <0.3%

Protein Geometry	Favored rotamers	197	94.26%	Goal: >98%
	Ramachandran outliers	1	0.41%	Goal: <0.05%
	Ramachandran favored	235	<mark>96.31%</mark>	Goal: >98%
	MolProbity score	1.68		77 th percentile [*] (N=7200, 1.60Å \pm 0.25Å)
	Cβ deviations >0.25Å	3	1.29%	Goal: 0
	Bad bonds:	10 / 1976	0.51%	Goal: 0%
	Bad angles:	17 / 2690	0.63%	Goal: <0.1%
Peptide Omegas	Cis Prolines:	0 / 11	0.00%	Expected: ≤ 1 per chain, or $\leq 5\%$

Так же и validation report, и MolProbity выдает карты Рамачандрана и положение наших остатков в них. В случае второго они явно качественнее, так как учитывают больше фактов из физики белка, к примеру то, что остаток, идущий перед пролином имеет несколько другую карту. Выдача MolProbity представлена на **рисунке 7.**

Также сервис предоставляет возможность визуализировать все ошибки в структуре (**рисунок 8**). Видно, что в нашей структуре ошибок достаточно мало.

РDВ суммирует информацию при помощи графика (**рисунок 9**). Сверху изображено качество покрытия остатков белка электронной плотностью, ниже полоса, показывающая количество параметров, по которым остаток является аутлайером – 0 – зеленый цвет, 1 – зеленый, 2 – оранжевый, 3 и более – красный. Остатки, положение которых плохо описывается электронной плотностью помечены красными точками над ними.



Рисунок 7. Положение аминокислотных остатков 4guw на картах Рамачандрана. Единственный аутлайер – это Val69.



Рисунок 8. Визуализация проблемных мест в структуре 4guw. Изображение получено с помощью KiNG + MolProbity.



Рисунок 9. Визуализация проблемных мест в структуре 4guw в pdb validation report. Особый интерес представляют V69, являющийся единственным аутлайером по картам Рамачандрана, и пара R222V223, где первый является сильным аутлайером, а второй плохо соответствует электронной плотности.

Сервер EDS также предоставляет важную информацию о качестве полученной модели и экспериментальных данных. Из графика Вильсона (**рисунок 10**) видно, что часть рефлексов (особенно в области высокого разрешения) отклоняются от прямой, что ставит под сомнение выбор авторов использовать все рефлексы. Из графика Padilla-Yeates (**рисунок 11**) графика видно, что в кристалле, использовавшемся в эксперименте не было одного из дефектов – perfect twinning (**http://nihserver.mbi.ucla.edu/Twinning/intro.html**) встречающегося у структур с данной симметрий и затрудняющего (хотя и не делающим невозможным) расшифровку рентгенограммы.

Кроме этого сервис предоставляет график Real-space R-value Z-score для остатков в структуре (**рисунок 12**). В целом можно заключить, что в нашей структуре большинство остатков хорошо соответствуют экспериментальной электронной плотности, хотя аутлайеры (z-score>2) присутствуют.

График температурного фактора (**рисунок 13**) для остатков свидетельствует от относительно малой подвижности остатков, средний значение фактора находится в районе 26.3 А для остатков белка. Из покраски структуры белка по значение В-фактора (**рисунок 14**) можно сделать наблюдение, что в основном структура малоподвижна, но у нее имеется подвижные участки. Особенно любопытен подвижный участок в составе домена 4guwA02, возможно, он принимает участие в функционирование фермента.

На **рисунке 15а-в** визуализирована электронная плотность белка при различных уровнях подрезки. Видно, что на уровне подрезки 1 и 2 сигма покрывается все части белка. Однако на уровне подрезке 3 сигма заметно, что упомянутый выше участок электронной плотности почти полностью оголен.

A. Created by O2D at Wed Sep 1807:26:34 2013 for an UNKNOWN mer *** XPS_GRAF 20051205/31.3 *** (c) G1K kyregt, 1992-2006 ***



Рисунок 10. График Вилсона для померенных рефлексов. В идеале экспериментальная часть (синяя кривая) должна соответствовать прямой линии. Получено с помощью сервиса EDS (https://eds.bmc.uu.se/cgi-bin/eds/).

A. Created by DATAMAN at Wed Sep 18 07/26:34 2013 for an UNKNOWN user *** XP5_GRAF 2005/205/8.1.3 *** (c) GI Kleywegi, 1992-2006 **



Рисунок 11. График Padilla-Yeates для померенных рефлексов. Линия нашего кристалла соответствует кристаллу без дефекта perfect twinning. Получено с помощью сервиса EDS (https://eds.bmc.uu.se/cgi-bin/eds/).



Рисунок 12. Real-space R-value Z-score для остатков структуры. Получено с помощью сервиса EDS (<u>https://eds.bmc.uu.se/cgi-bin/eds/</u>)



13. В-фактор для остатков структуры. Получено с помощью сервиса EDS (<u>https://eds.bmc.uu.se/cgi-bin/eds/</u>)



Рисунок 14. Покраска структуры 4guw по величине В-фактора. Области с большим В-фактором покрашены красным, маленьким – синим.



Рисунок 15. Электронная плотность вокруг структуру 4guw при уровне подрезки 1, 2 и 3 сигма.

Маргинальные остатки

В качестве примере маргинальных остатков было 11 аминокислотных остатков, приведенных в таблице 4.

Название остатка	Причина маргинальности
Val69	Не попадает в предпочитаемую зону на
	карте Рамачандрана и имеет нетипичную
	конформацию боковой цепи
Arg222	Нетипичный угол связи NE-CZ-NH,
	нетичная конформация боковой цепи
Val223	Плохо соответствует экспериментальной
	электронной плотности
Asn58	Инверсия боковой цепи(flip)
Gln169	Инверсия боковой цепи(flip)
Arg122	Нетипичные длины связей CG-CD, CB-CG,
Ser42	Плохо соответствует экспериментальной
	электронной плотности
Tyr55	Плохо соответствует экспериментальной
	электронной плотности
Asn245	Инверсия боковой цепи(flip)
Leu77	Clash L77:HD22 и THR78:HG23, Leu77:HA
	и Leu77:HD23, нетипичная конформация
	боковой цепи
Tyr107	имеет нетипичную конформацию боковой
	цепи

Таблица 4. Примеры маргинальных остатков в структуре 4GUW.

Остатки, выделенные зеленым были далее рассмотрены.

Val69

Даже на визуализация электронной плотности на уровне 3 сигма дает изображение (**рисунок 16**), в котором четко определяются центры всех атомов, из чего заключаем, что его нетипичное положении полностью соответствует электронной плотности. Если визуализировать контакты 69 остатка с другими остатками в белке замечаем, что он связан двумя водородными связями с остатками Asp65 и Arg163 (**рисунок 17**), что компенсирует энергетических проигрыш от невыгодной конформации.



Рисунок 16. Визуализация экспериментальной электронной плотности вокруг Val69.



Рисунок 17. Визуализация водородных связей Val69 с остатками Asp65 и Arg163.

Arg222

В случае этого остатка при визуализации электронной плотности даже на уровне 2.5 сигма все не так хорошо (рисунок 18). Логичных объяснений в отхождении от предпочтительных параметров связей найдено не было, потому в данном случае это скорее всего ошибка авторов.



Рисунок 18. Визуализация электронной плотности Arg222.

Val223

У боковой цепи этого остатка нет электронной плотности уже на 1.5 сигма (рисунок 223). Потому тут вставлялась боковая цепь просто из соображений авторов.



Рисунок 19. Визуализация электронной плотности Val223.

Asn58

По утверждению сервиса Molprobity у данного остатка произошла инверсия боковой цепи. Однако никаких свидетельств ни в пользу этого, ни против этого мною найдено не было (**рисунок 20**). Электронная плотность (сигма = 2.5) не позволяет установить положения кислорода и азота. По ходу боковой цепи действительно похоже, что в данном случае наблюдаем инверсию. Перемена местами кислорода и азота к увеличению/уменьшению числа водродных связей не ведет.



Рисунок 20. Визуализация электронной плотности Asn58 и водородных связей кислорода и азота боковой цепи с водой.

Gln169

По утверждению сервиса, у данного остатка также наблюдается инверсия боковой цепи. При визуализации электронной плотности (сигма = 2.5, рисунок 21) видим, что координаты всей боковой цепи Gln169 определить затруднительно По ходу боковой цепи действительно можно сделать подобное утверждение. Кроме того если поменять кислород и азот местами, то получаем возможность образования водяного мостика, который до этого образоваться не мог (рисунок 22)



Рисунок 21. Визуализация электронной плотности Gln169.



Рисунок 22. В при данном положении кислорода и азота в боковой цепи образование водяного мостика между Gln169 и Gln170(невозможная на данный момент часть показана красной связью) невозможно. Однако если поменять их местами, образование мостика становится возможным.

Arg122

Для данного остатка наблюдаем нетипичные длины связей между углеродными атомами в боковой цепи. При визуализации электронной плотности видим уже при сигма равном 2.5 (рисунок 23), что координаты именно этих атомов определить затруднительно. Потому вполне возможно, что они они были расшифрованы неверно. Однако стоит отметить, что при просмотре возможных связей данного остатка с окружающими его остатками (рисунок 24) видно обилие возможностей для него образовать водородную связь с различными остатками. Возможно, именно это делает выгодным в данной ситуации некоторое укорочение длины связи между атомами боковой цепи (она не переходит через минимально допустимое расстояние, то есть, в принципе возможна).



Рисунок 23. Визуализация электронной плотности Arg122



Рисунок 24. Возможные водородные связи боковой цепи Arg122.

PDB_redo

Сервис PDB_redo содержит структуры из банка PDB, на основании экспериментальных данных

переделанные заново или исправленные. Структура 4guw также есть на этом сайте. Два способа оптимизации, примененные на сайте, в основном улучшали характеристики структуры (кроме конформации боковых цепей)(рисунок 25).

	Original PDB entry	Conservatively optimised	Fully optimised
1st generation packing quality ¹	0.270	0.286	0.288
2nd generation packing quality ¹	-0.467	-0.396	-0.143
Ramachandran plot appearance ¹	0.146	0.193	0.098
Chi-1/Chi-2 rotamer normality ¹	-1.165	0.493	0.583
Backbone conformation ¹	0.589	0.442	0.450
Bond length RMS Z-score ²	1.336	0.656	0.673
Bond angle RMS Z-score ²	1.130	0.803	0.783
Total number of bumps ³	22	15	13
Unsatisfied H-bond donors/acceptors ³	12	12	12
Full WHAT_CHECK reports	<u>Link</u>	<u>Link</u>	<u>Link</u>

¹ Higher is better

² Should be lower than 1.000

WHAT CHECK validation

³ Fewer is better

Рисунок 25. PDB_redo – характеристики оригинальной структуры, консервативнооптимизированной и полностью оптимизированной.

Далее мною была скачана полностью оптимизированная структура и подана на вход сервису Molprobity.

Сервис Molprobity считает, что в структуре в Gln93 наличиствует боковая инверсия. На этот раз я проигнорировал это предупреждение. Характеристика оптимизированной структуры представлена в таблице 5. Кроме этой информации стоит отметить, что для оптимизированной структуры R-value и R-free уменьшились, став соответсвенно 0.17159 и 0.19969.

Таблица 5. Общая характеристика структуры полностью оптимизированной 4guw. сервис MolProbity.

All-Atom	Clashscore, all atoms:	3.73		97 th percentile [*] (N=718, 1.60Å \pm 0.25Å)	
Contacts	Clashscore is the number of serious steric overlaps (> 0.4 Å) per 1000 atoms.				
Protein	Poor rotamers	2	0.96%	Goal: <0.3%	
Geometry	Favored rotamers	205	98.09%	Goal: >98%	

	Ramachandran outliers	1	0.41%	Goal: <0.05%
	Ramachandran favored	238	97.54%	Goal: >98%
	MolProbity score [^]	1.26		97 th percentile [*] (N=7200, 1.60Å \pm 0.25Å)
	C β deviations >0.25Å	0	0.00%	Goal: 0
Peptide Omegas	Cis Prolines:	0 / 11	0.00%	Expected: ≤ 1 per chain, or $\leq 5\%$

Видно, что характеристики структуры значительно улучшились. Посмотрим, что случилось с разобранными нами подробно аминокислотами.

Val69 по-прежнему является аутлайером из-за нехарактерных углов вращения в боковой и основной цепях, но, как было показано выше, этому есть биологическое обоснование. Все аминокислоты, у которых были инверсии боковой цепи в исходной структуре их больше не имеют. Arg222 по-прежнему является аутлайером, что объясняется, скорее всего, просто недостатком информации для его качественной реконструкции. Arg122 больше не является аутлайером из-за длины связей, но теперь он является аутлайером из-за clash с одним из водородов (HD11) ILE171. Потому нельзя точно сказать, какой из вариантов был правильнее.

Сравним исходную структуру и оптимизированную.

Из их совмещения (рисунок 26) видно, что в значимо они не отличаются, однако в оптимизированной структуре отсутствуют несколько вызывавших подозрение вторичных структур (одна из которых находилась прямо на конце белка, что довольно странно, а другая достаточно мала). Также видно различие в той области, для которой выше наблюдался высокий температурный фактор и меньшее качество покрытия экспериментальной электронной плотностью. Отличие, впрочем, незначительно и потому, вероятно, его можно проигнорировать.



Рисунок 26. Совмещение исходной структуры с полностью оптимизированной в pdb_redo. Зеленая структура – исходная, голубая – полученная из pdb_redo. Улучшенные различающиеся участки показаны в исходной структуре красным цветом, отличающиеся, но про улучшение которых достоверно что-то сказать затруднительно покарашены в серый и белый цвета.

Рассмотрим в оптимизированной структуре Gln93, для которого сервис утверждает наличие инверсии боковой цепи и остаток, сохранивший положение на карте Рамачндрана не в оптимальной области, но не рассматривавшийся раньше – Leu77, который в обоих структурах является аутлайером.

Gln93

Положение данной аминокислоты (как и соседних) существенно не отличается в обеих структурах (рисунок 27). Каких либо свидетельств в пользу инверсии найдено не было.



Рисунок 27. Совмещение исходной структуры с полностью оптимизированной в pdb_redo. Показан Gln93, который в оптимизированной структуре стал детектироваться MolProbity как остаток с инвертированной боковой цепью.

Leu77

Для данного остатка при визуализации электронной плотности (сигма = 2.5) видно, что местоположение атомов боковой цепи установить практически невозможно (рисунок 28). Геометрия остатка отличается в исходной и оптимизированных структурах, однако он попрежнему остается аутлайером, причина скорее всего просто в недостатке экспериментальных данных.



Рисунок 28. Совмещение исходной структуры с полностью оптимизированной в pdb_redo. Показан Leu77, который в обеих структурах является аутлайером.

В целом же можно заключить, что оптимизация в PDB_redo улучшило качество структуры.

Выводы

- 1) Структура 4guw с заявленным разрешением 1.6 А расшифровано в достаточной мере хорошо.
- 2) В структуре есть нескольких маргинальных остатков, однако в ряде случаев (пример Val66) маргинальность остатков биологически обусловлена.
- Часть структуры (участок цепи 205-229) имеет более низкую по сравнению с остальной структурой величину электронной плотности, по которой сложнее определить положения всех атомов участка, что может быть объяснено увеличенной подвижностью этого участка.
- 4) В структуре отсутствуют явные грубые ошибки, в работе не замечено попытки подогнать значение каких-то оценок качества структуры под оптимальные
- 5) Оптимизация в сервисе PDB_redo улучшает качество структуры, уменьшая число маргинальных остатков и улучшая многие характеристики качества структуры.