

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

ФАКУЛЬТЕТ БИОИНЖЕНЕРИИ И БИОИНФОРМАТИКИ

**Отчет по качеству расшифровки РСА структуры
зависимого от совместной локализации
белка-переключателя (Co-LOCKR)**

Выполнила студентка 4-го курса
Смородина Ева Геннадьевна
в рамках курса по структурной
биоинформатике

Москва

2020 г.

Содержание

Содержание	2
Аннотация	2
Введение	2
Результаты	4
Общая информация о модели	4
Значения индикаторов качества модели в целом	6
Маргинальные остатки	7
Анализ 5 маргинальных остатков и/или гетеромолекул	8
Сравнение модели из PDB с моделью из PDB_redo	10
Обсуждения	11
Заключение	12
Благодарности	12
Список литературы	12

Аннотация

Данный отчет представляет собой анализ расшифровки PCA структуры зависящего от совместной локализации белка-переключателя (Co-LOCKR) (PDB ID: 7JH5). Во “Введении” кратко рассказано о самой белке и связанных с ней публикациях. Систематизированная информация о структуре и ее качестве представлены в разделах “Результаты”. Там же приводится описание нескольких маргинальных остатков и варианты их замены или исправления, а также сравнение структуры из PDB с моделью из PDB_redo. В “Обсуждении” приводятся показатели, на основе которых качество структуры может подвергаться сомнению. Итоговая оценка качества структуры дается в “Заключении”. В завершении отчета идут “Благодарности” и “Список используемой литературы”.

Введение

Аллостерически регулируемые белки широко распространены в природе, но их *de novo* дизайн все еще остается сложной задачей вычислительной структурной биологии, поскольку требует явного моделирования множественных состояний со сравнимыми свободными энергиями. Однако в последние пару лет это направление стремительно развивается. Так появилась “Latching, Orthogonal Cage/Key pRotein” (LOCKR) –

созданная *de novo* белковая система, которая меняет свою конформацию (переключается) путем модуляции конкурирующих меж- и внутримолекулярных взаимодействий. В основе LOCKR лежит сконструированная статическая пятиспиральная структура, состоящая из “клетки” с единым интерфейсом, которая может взаимодействовать либо внутримолекулярно с концевой “защелкивающейся” спиралью, либо межмолекулярно с пептидным “ключом”. На защелке закодированы функциональные мотивы связывания, деградации или ядерного экспорта, которые действуют только тогда, когда ключ перемещает защелку из клетки. Клетка изолирует биоактивный пептид, но связывание ключа с клеткой меняет конформацию структуры и “обнажает” этот пептид. Путем замены идентификатора заключенного в клетку пептида и настройки аффинности связывания можно создать различные переключатели LOCKR для широкого диапазона выходных сигналов (рисунок 1) [1, 2].

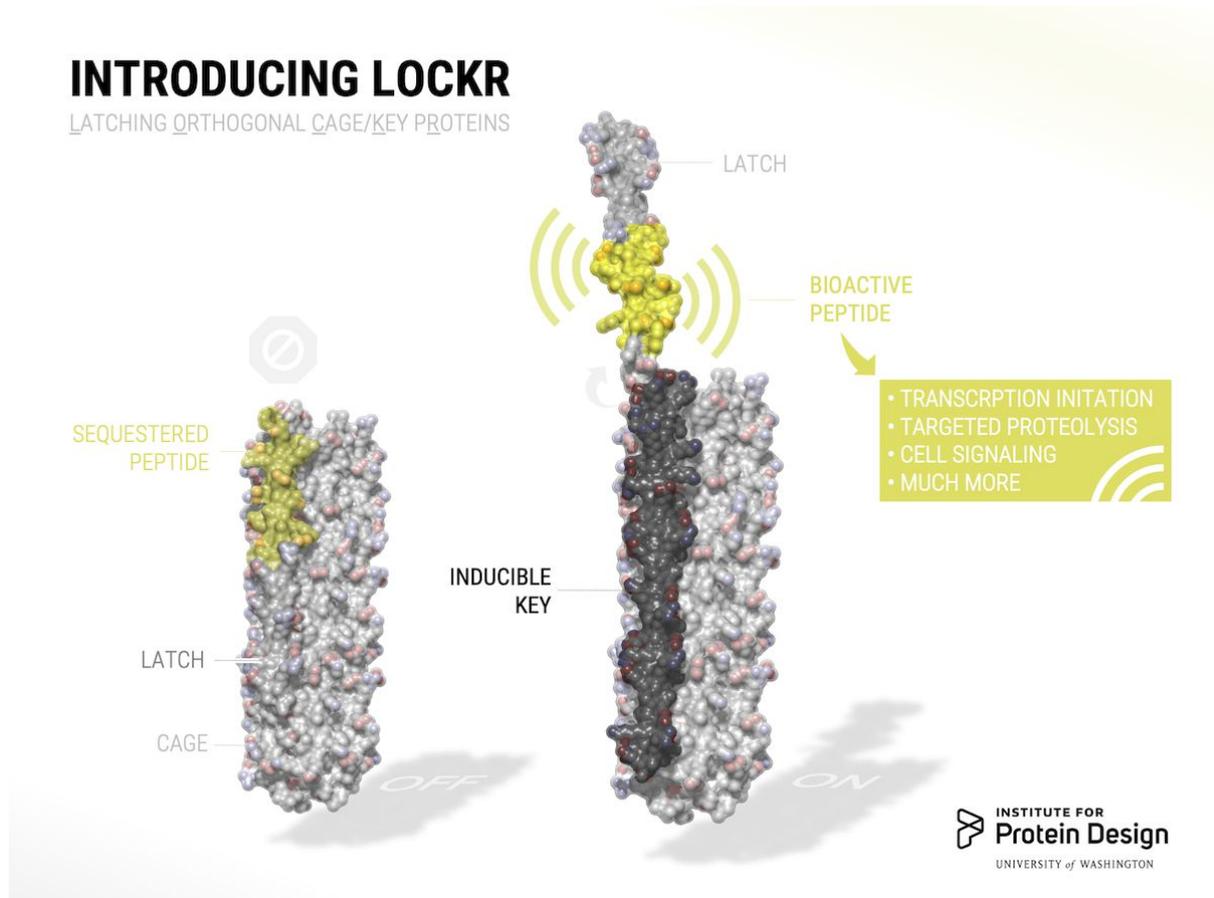


Рисунок 1. Latching, Orthogonal Cage/Key pRotein (LOCKR) система.

LOCKR системы открыли совершенно новые возможности для “программирования” работы клеток. Так с помощью системы, базирующейся на LOCKR, была решена проблема “недораспознавания” раковых клеток – точного нацеливание на клетки при отсутствии единственного поверхностного маркера, отличающего бы их от нормальных клеток. Белковые переключатели, называемые Co-LOCKR (“colocalization-dependent protein switches”), связываются с антигенами на поверхности клетки и активируются посредством конформационного изменения только при наличии точной комбинации антигенов. Другими словами. они воздействуют на

клетку, используя определенные комбинации белков, присутствующих на поверхности раковой клетки.

Чтобы проверить работу этих систем, Co-LOCKR использовали для направления Т-лимфоцитов рецептора химерного антигена к опухолевым клеткам, экспрессирующим определенные антигены. Co-LOCKR выполняли только 3 логические операции: AND перенаправлял специфичность Т-клеток против опухолевых клеток, экспрессирующих два поверхностных антигена, при этом избегая нецелевого распознавания одноантигенных клеток, а “переключатели” с тремя входами, добавляющие логику NOT или OR, избегал или включал клетки, экспрессирующие третий антиген. Таким образом, белки, созданные *de novo*, могут выполнять “вычисления” на поверхности клеток, объединяя несколько различных взаимодействий связывания в один выходной сигнал (рисунок 2) [3].

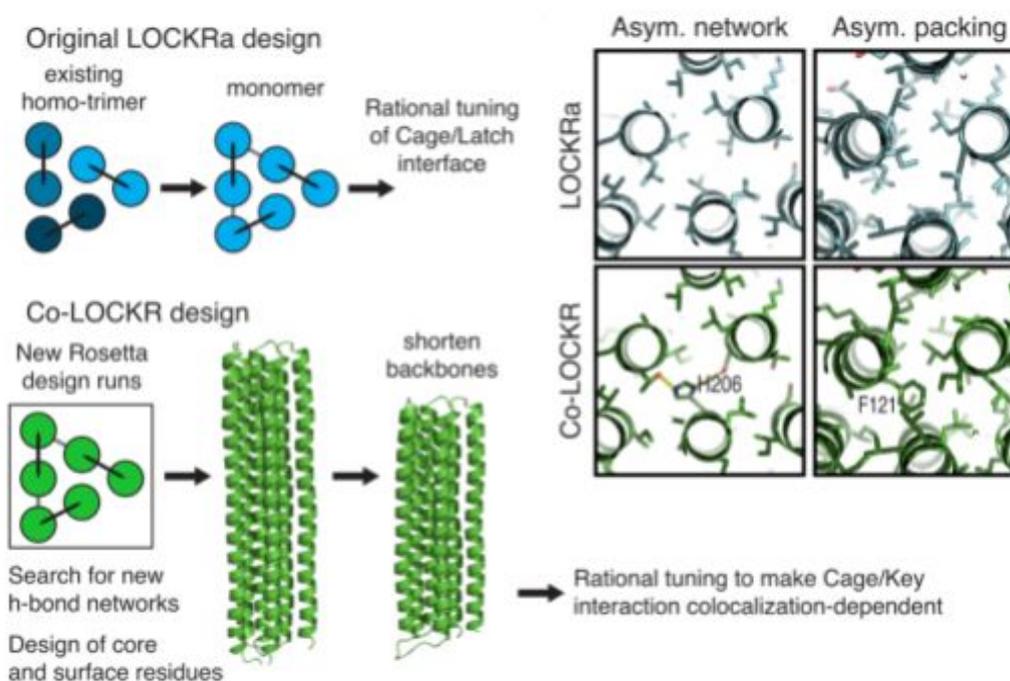


Рисунок 2. Вычислительный дизайн LOCKR и Co-LOCKR [3].

О структуре Co-LOCKR пойдет речь в данном отчете.

Результаты

Общая информация о модели

Модель конструкции клетки, используемой для создания Co-LOCKR, была депонирована на сайт RCSB PDB (PDB ID: 7JH5) 20 июля 2020 года и выпущена 9 сентября 2020 года. Ее выложили Vick, M.J., Lajoie, M.J., Boyken, S.E., Sankaran, V., Baker, D. Структура была получена методом PCA, чтобы подтвердить результаты *de*

novо дизайна данного белка. Соответствие вычислительной и экспериментальной структуры показано на рисунке 3. На модель ссылаются в соответствующей статье [3], вышедшей 25 сентября 2020 года в журнале “Science”. Ген белка был экспрессирован в *Escherichia coli*.

Модель обладает следующими характеристиками (общая информация о структуре была взята из базы данных PDB [4]):

- состав комплекса: Co-LOCKR (цепи A и B)
- решение фазовой проблемы: метод молекулярного замещения
- число измеренных рефлексов: 23830 рефлексов
- разрешение: 2.10 Å
- полнота набора структурных факторов: 85.2%
- диапазон разрешений структурных факторов: 2.10 - 42.38 Å
- параметры кристаллографической ячейки: $a = 61.84$, $\alpha = 90$, $b = 52.914$, $\beta = 102.64$, $c = 75.591$, $\gamma = 90$
- кристаллографическая группа: P 1 21 1
- количество кристаллографических симметрий в асимметрической ячейке: 2
- количество некристаллографических симметрий в асимметрической ячейке: 0

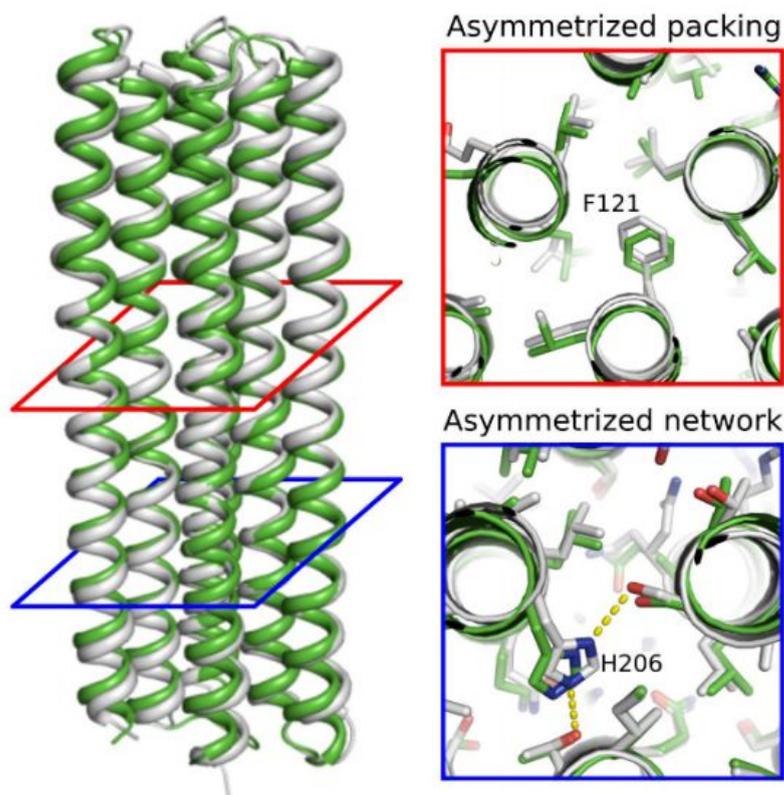


Рисунок 3. Рентгеновский кристалл Co-LOCKR (показан белым) соответствует модели, сделанной с помощью *de novo* дизайна (показана зеленым) со среднеквадратичным отклонением 1.1 Å по всем атомам основной цепи. Поперечные сечения иллюстрируют асимметричную упаковку гидрофобных остатков (красный квадрат) и асимметричную сеть водородных связей (синий квадрат).

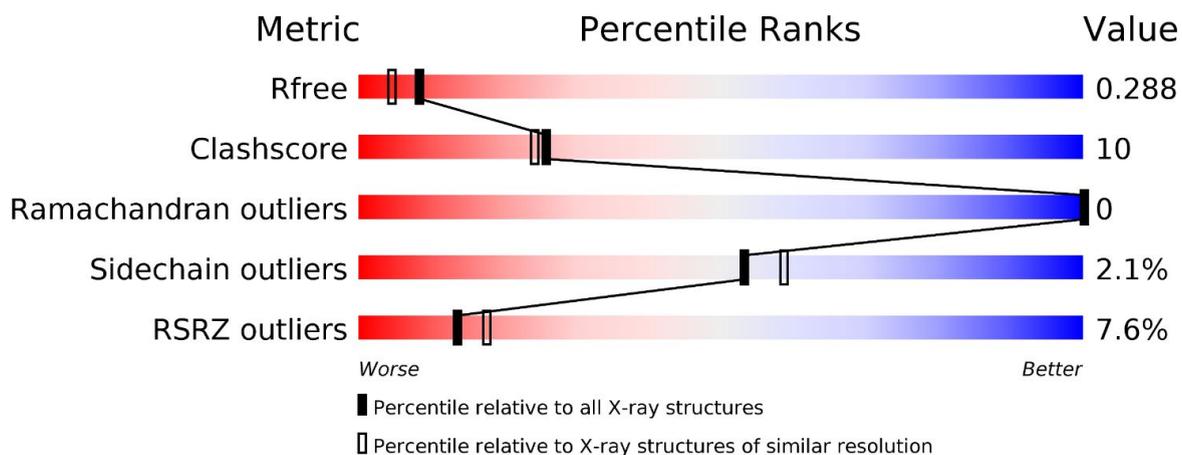


Рисунок 4. Основные показатели качества модели.

Значения индикаторов качества модели в целом

Для поиска маргинальных остатков были использованы данные Full Report с сайта PDB и сервисы PDB_redo [5] и MolProbity [6, 7].

Модель обладает следующими параметрами, характеризующими ее качество в целом:

- R-value (весь набор рефлексов) : 0.260
- R-value (тренировочная выборка): 0.257
- R-free (тестовая выборка): 0.288
- размер R-free: 1736 (7,28 % т всего набора рефлексов)
- карта Рамачандрана: 0 (0%) маргинальных остатков (рисунки 5-7)

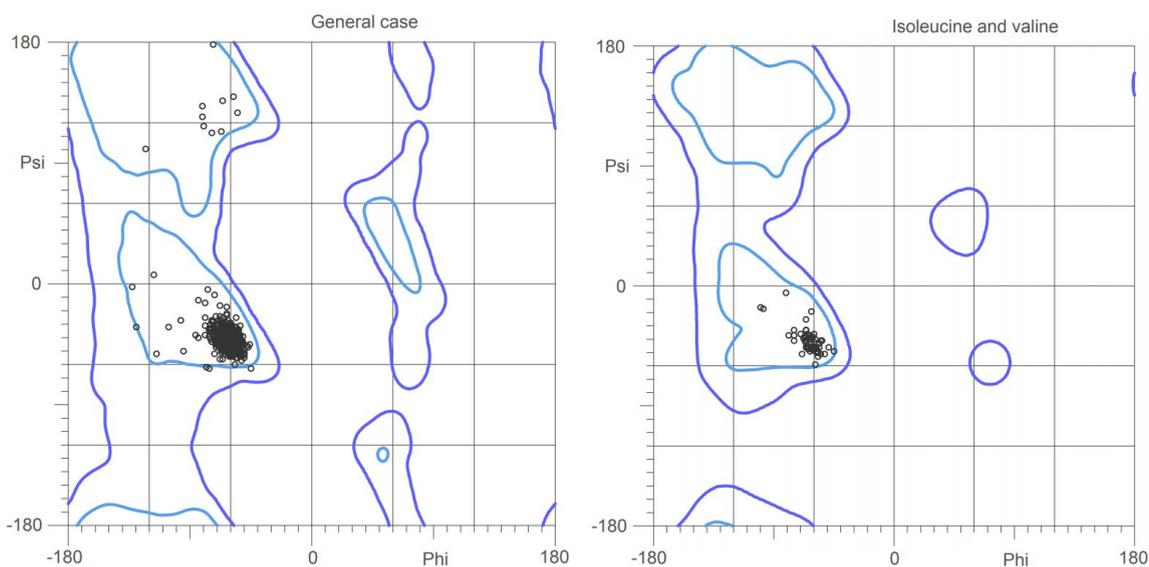


Рисунок 5. Карты Рамачандрана для общего случая (слева) и остатков изолейцина и валина (справа).

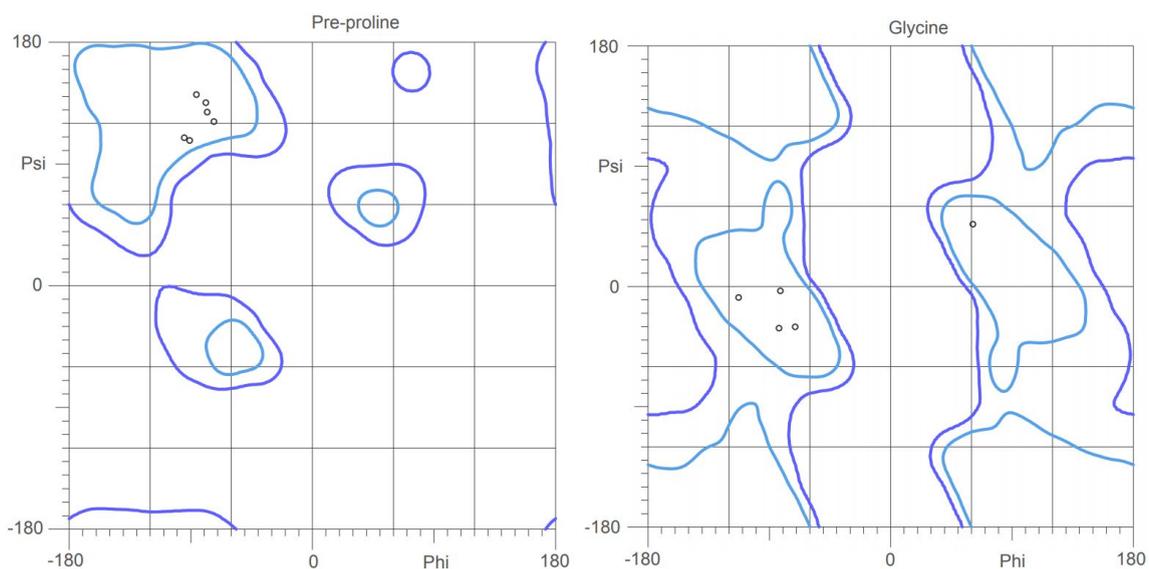


Рисунок 6. Карты Рамачандрана для остатков перед пролином (слева) и глицина (справа).

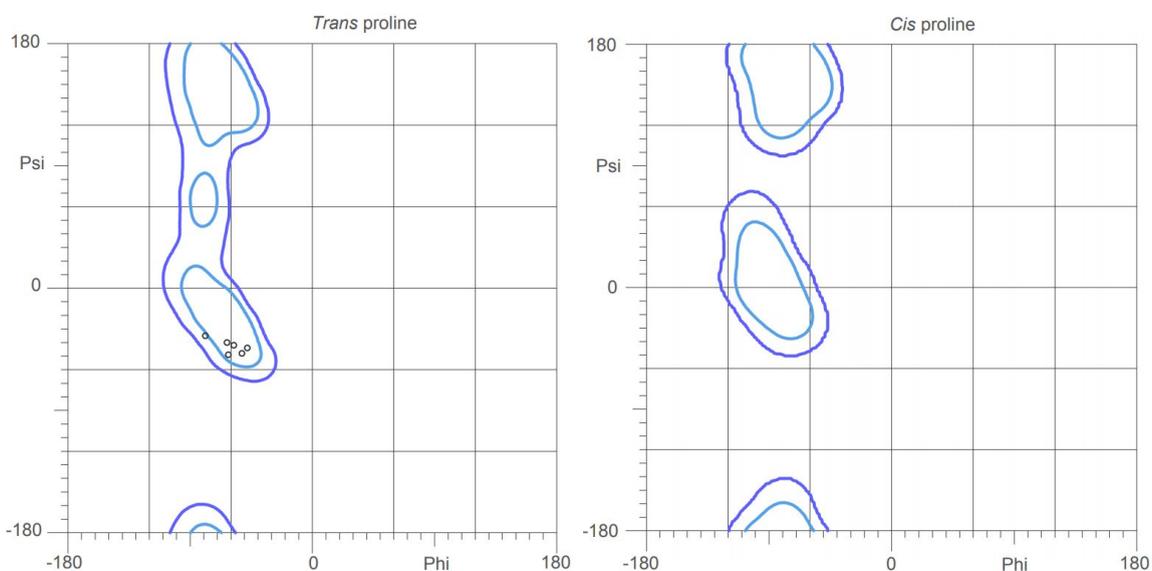


Рисунок 7. Карты Рамачандрана для тран- (слева) и цис-пролина (справа).

Сервис MolProbity выявил 5 маргинальных остатков по энергетически невыгодным ротамерам (A/LYS/6, A/ILE/181, B/ASP/48, B/LYS/101 и B/GLU/182) и 1 – по плохим длинам связей (A/ILE/181). Ни один остаток не располагался в запрещенных областях карты Рамачандрана, также не было отклонения в $\Sigma\beta$ и углах связей, а все пептидные связи имели транс-конформацию.

Однако PDB redo исправил PDB-модель так, что в структуре появился 1 остаток в энергетически невыгодном участке общей карты Рамачандрана, а 4 – в разрешенной, но не идеальной области.

Маргинальные остатки

Таблица 1. Список некоторых маргинальных остатков структуры Co-LOCKR с указанием параметров, по которым они были отобраны. Название формируется следующим образом: имя цепь/имя остатка/номер остатка.

Название	Показатель
A/SER/-1	Отсутствует OG атом (WHAT_CHECK), RSRZ > 2 (Full Report)
A/SER/1	Отсутствует OG атом (WHAT_CHECK), RSRZ > 2 (Full Report)
A/GLU/2	Отсутствуют атомы CG, CD, OE1 и OE2 (WHAT_CHECK)
A/ARG/5	Отсутствуют атомы CD, NE, CZ, NH1 и NH2 (WHAT_CHECK)
A/LYS/50	Отсутствуют атомы CG, CD, CE и NZ (WHAT_CHECK)
A/ASP/243	Отсутствуют атомы CG, OD1 и OD2 (WHAT_CHECK)
A/THR/242	Отсутствуют атомы CG1 и OG1 (Full Report)
A/LEU/3	RSRZ > 2 (Full Report)
A/LEU/42	RSRZ > 2 (Full Report)
A/LEU/46	RSRZ > 2 (Full Report)
A/GLN/129	Подозрение на необходимость инверсии (WHAT_CHECK)
A/GLU/86	Отсутствуют атомы CG, CD, OE1 и OE2 (WHAT_CHECK)

Анализ пяти маргинальных остатков и/или гетеромолекул

Неполные остатки

На рисунке 8 видны два маргинальных по неполноте состава остатка из таблицы 1: в A/ARG/5 отсутствуют пять атомов (CD, NE, CZ, NH1 и NH2), в в A/SER/1 – один (OG). Как мне кажется, это вызвано тем, что на этих атомах электронной плотности не было и авторы не захотели их “придумывать”. Такое решение проблемы недостаточной расшифровки сейчас считается более корректным, чем вписывание в структуру несуществующих остатков или атомов.

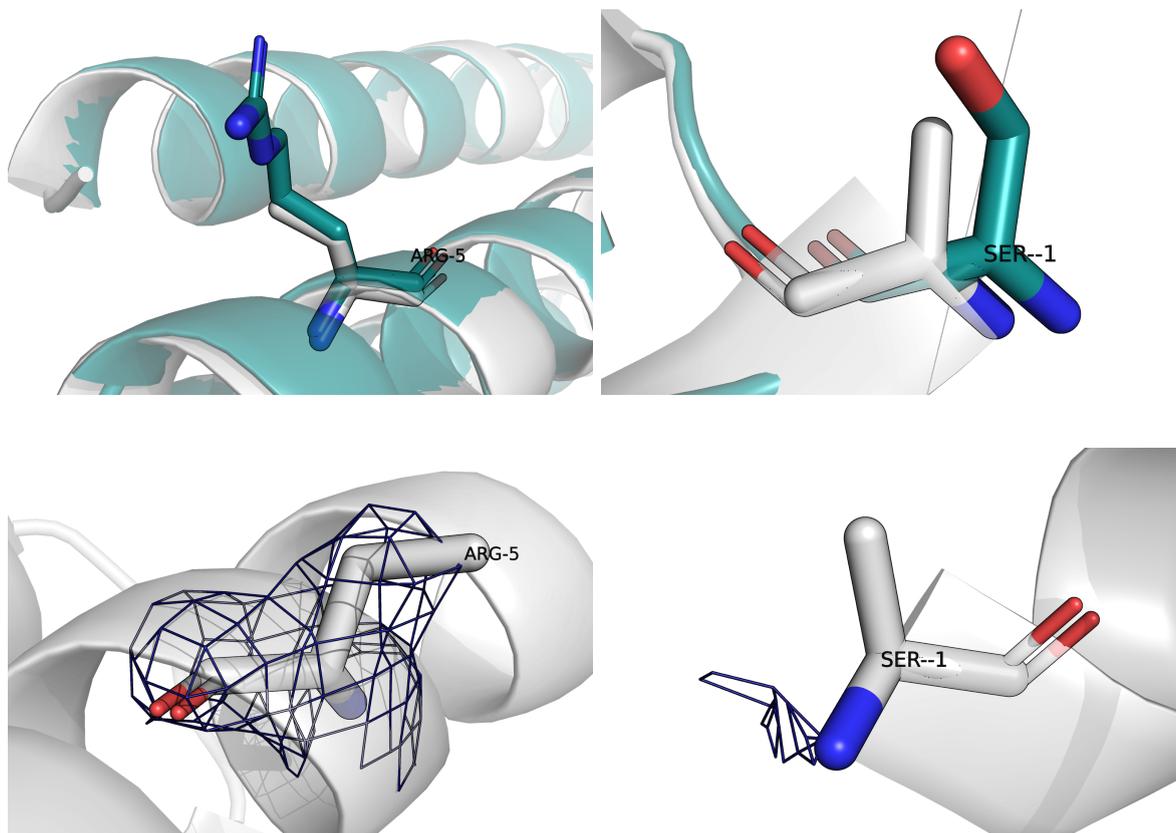


Рисунок 8. Часть структуры Co-LOCKR (PDB ID: 7JH5) из PDB (светло-серая) и PDB_redo (сине-зеленая). У A/ARG/5 (слева) отсутствуют атомы CD, NE, CZ, NH1 и NH2, у A/SER/1 (справа) – атом OG в PDB-структуре.

Инвертированные боковые цепи

По данным Full Report A/GLN/210 имел $RSRZ > 2$. Остаток в модели изначально был повернут так, как показано на рисунке 9 в левой нижней части. Сервис PDB_redo поменял местами атомы кислорода и азота в боковой цепи этого остатка. Благодаря этому образовалась водородная связь между кислородом A/GLN/210 и азотом A/ASN/114, которой не было в PDB-структуре.

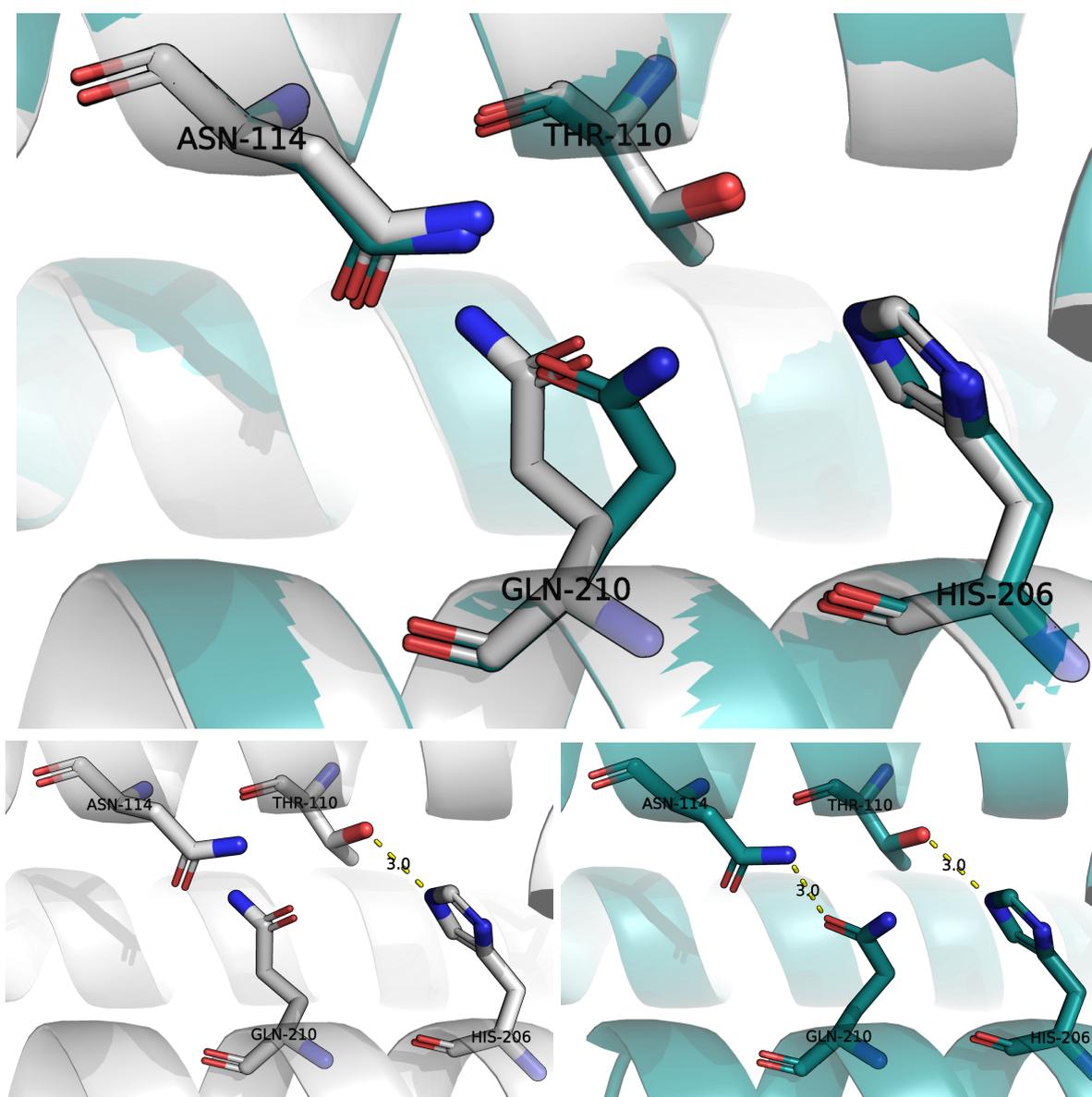


Рисунок 9. Часть структуры Co-LOCKR (PDB ID: 7JH5) из PDB (светло-серая) и PDB_redo (сине-зеленая). Остатки A/ASN/114 и A/GLN/210 формируют водородную связь только в структуре из PDB_redo, т. к. в ней боковая цепь A/GLN/210 инвертирована.

Аналогичная ситуация наблюдается и с A/ASN/114. WHAT_CHECK для этого остатка предсказал скрытые неудовлетворенные акцепторы водородной связи, а также поставил его в ряд проблемных. Если повернуть карбоксамидную группу A/ASN/114 на 90° , образование водородных связей становится более благоприятным (рисунок 11). При этом поворот никак не мешает соседним атомам, а электронная плотность боковой цепи разрешена плохо.

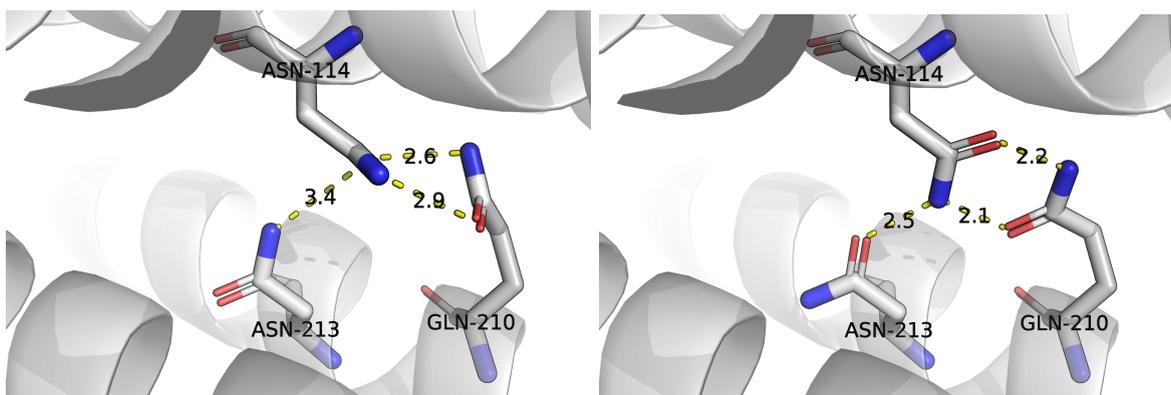


Рисунок 10. Часть структуры Co-LOCKR (PDB ID: 7JH5) из PDB. У A/ASN/114 отсутствует атом OG в PDB-структуре. Водородные связи A/ASN/114 с A/GLN/210 сильнее при повороте карбоксамидных групп A/ASN/114 и A/GLN/210. Водородная связь между A/ASN/114 и A/ASN/213 спорна, но после инверсии ее образование более вероятно.

Нефиксированные молекулы воды

Еще один неполный по составу остаток из таблицы 1 (A/GLU/86) представлен на рисунке 11 (верхний ряд). Помимо этого молекулы растворителя (воды) в области A/GLU/86, A/GLU/87 и A/GLU/90 не образуют водородных связей ни с аминокислотами из этой ячейки, ни из соседней. Такая же ситуация наблюдается и с молекулами воды A/НОН/378 и A/НОН/80 (рисунок 12). Они были определены WHAT_CHECK как не имеющие водородных связей. Скорее всего, их наличие связано с ошибкой программы для расшифровки электронной плотности.

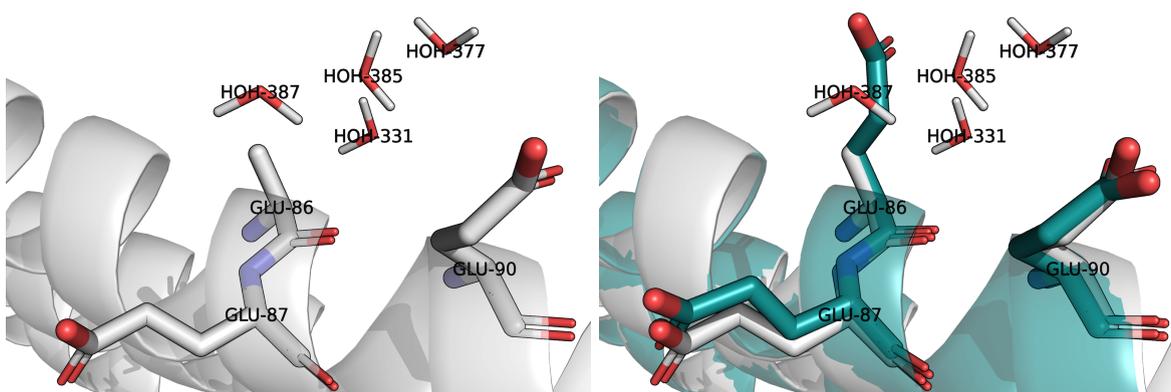


Рисунок 11. Часть структуры Co-LOCKR (PDB ID: 7JH5) из PDB (светло-серая) и PDB_redo (сине-зеленая). У A/GLU/86 отсутствуют атомы CG, CD, OE1 и OE2 в PDB-структуре. Молекулы воды не образуют водородных связей ни с показанной молекулой, ни с молекулами из соседних ячеек.



HOH-380



HOH-378

Рисунок 12. Часть структуры Co-LOCKR (PDB ID: 7JH5) из PDB. Молекулы воды не образуют водородных связей ни с показанной молекулой, ни с молекулами из соседних ячеек.

Сравнение модели из PDB с моделью из PDB_redo

В PDB_redo-структуре все отсутствующие атомы дополнены, т. е. в ней исправлены все неполные остатки. Метрики, по которым валидировалась структура, показаны на рисунке 12.

Validation metrics from PDB-REDO		
	PDB	PDB-REDO
Crystallographic refinement		
<i>R</i>	0.2628	0.2584
<i>R-free</i>	0.2904	0.3277
<i>Bond length RMS Z-score</i>	0.240	0.514
<i>Bond angle RMS Z-score</i>	0.519	0.694
Model quality (raw scores percentiles)		
<i>Ramachandran plot appearance</i>	27	36
<i>Rotamer normality</i>	18	27
<i>Coarse packing</i>	N/A	100
<i>Fine packing</i>	N/A	99
<i>Bump severity</i>	37	37
<i>Hydrogen bond satisfaction</i>	100	100

Рисунок 12. Метрики валидации структуры Co-LOCKR из PDB-REDO.

В целом, показатели структуры из PDB_redo превосходят таковые в PDB-структуре (рисунок 13). Суммарно в модели из PDB_redo по сравнению со структурой из PDB изменились 6 ротамеров, у 81 боковой цепи была изменена конформация и 19 молекул воды были удалены.

Однако в целом кажется, что структура молекулы изменилась не сильно после ее преобразования PDB_redo.

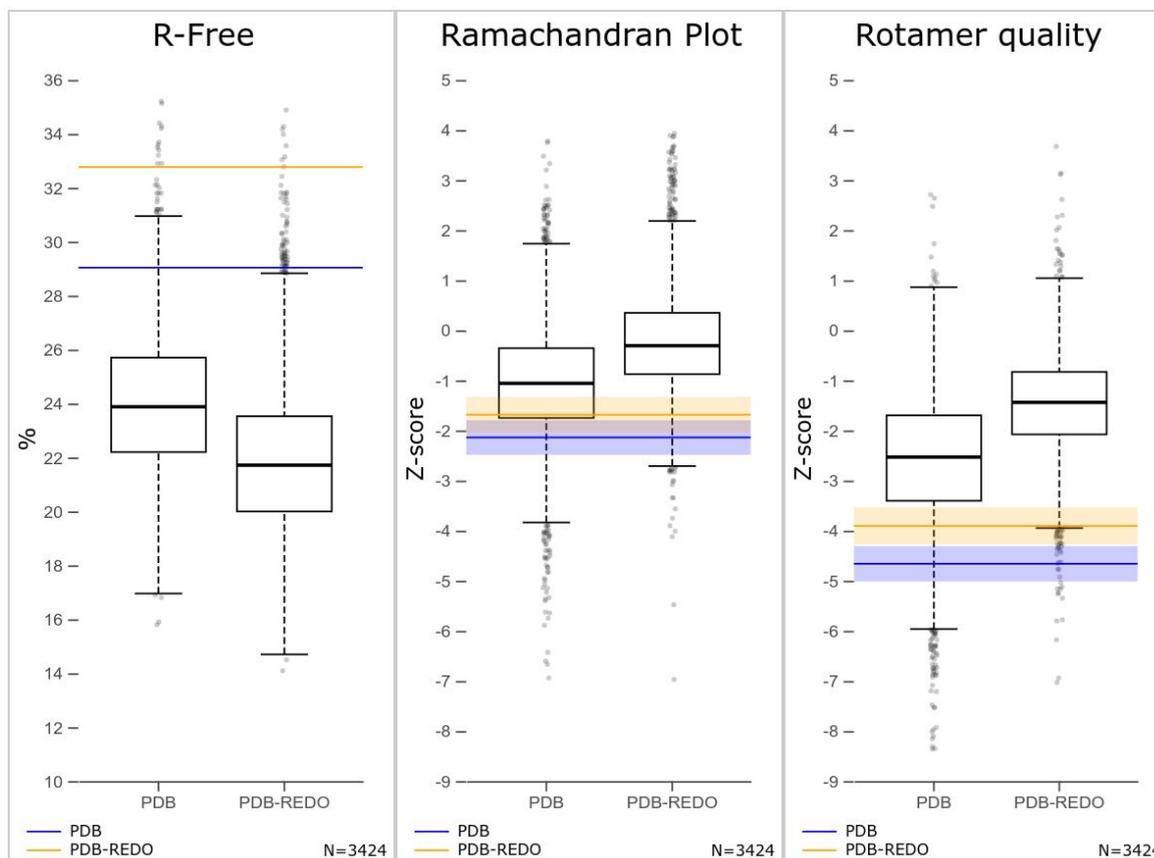


Рисунок 13. Сравнение основных показателей качества расшифровки структуры Co-LOCKR из PDB и PDB_redo. Модель из PDB_redo имеет лучшие средние значения по всем характеристикам.

Обсуждения

С учетом недостаточной полноты набора гармоник (85.2%) можно утверждать, что настоящее разрешение структуры хуже того, которое пишут авторы (2.10 Å), т. к. достоверна полнота для “честного” определения разрешения должна быть больше 90%. Помимо этого за разрешения просто взяли самую низкую величину (RESOLUTION RANGE HIGH) из всего диапазона.

R-free немного больше 0.2, но отличается от R-value тренировочной выборки меньше, чем на 0.1. По этим данным можно сказать, что структура не идеальна, но достаточно удовлетворительная.

Тем не менее модель отлично соответствует картам Рамачандрана и к ротамерам нет практически никаких замечаний. Такое идеальное соответствие может быть связано с тем, что структура изначально была получена с помощью *de novo* дизайна в Rosetta,

алгоритм подбора аминокислот которой построен на подборе остатков по наиболее энергетически выгодным участкам карт Рамачандрана и оптимальным значениями ψ углов.

Было довольно сложно искать маргинальные остатки, поскольку даже те остатки, которые как-то выбивались по параметрам, отвечающим на качество разрешения структуры, все равно хорошо соответствует физико-химическим и геометрическим характеристикам

Заключение

Оценка структуры по картам Рамачандра и ротамером говорит о практически идеальном качестве структуры. Тем не менее другие характеристики (недостаточная полнота гармоник, много остатков, у которых $RSRZ > 2$, немного больший 0.2 R-free) портят эту оценку. Таким образом, качество структуры назвать отличным затруднительно, но хорошим – вполне возможно.

Благодарности

Я бы хотела поблагодарить преподавателей курса структурной биоинформатике на ФББ за интересные и полезные лекции и практикумы, которые помогли лучше понять устройство биологических систем в 3D.

Список литературы

1. Langan, Robert & Boyken, Scott & Ng, Andrew & Samson, Jennifer & Dods, Galen & Westbrook, Alexandra & Nguyen, Taylor & Lajoie, Marc & Chen, Zibo & Berger, Stephanie & Mulligan, Vikram & Dueber, John & Novak, Walter & El-Samad, Hana & Baker, David. (2019). De novo design of bioactive protein switches. *Nature*. 572. 10.1038/s41586-019-1432-8.
2. Ng, Andrew & Nguyen, Taylor & Gomez Schiavon, Mariana & Dods, Galen & Langan, Robert & Boyken, Scott & Samson, Jennifer & Waldburger, Lucas & Dueber, John & Baker, David & El-Samad, Hana. (2019). Modular and tunable biological feedback control using a de novo protein switch. *Nature*. 572. 1. 10.1038/s41586-019-1425-7.
3. Lajoie, Marc & Boyken, Scott & Salter, Alexander & Bruffey, Jilliane & Rajan, Anusha & Langan, Robert & Olshefsky, Audrey & Muhunthan, Vishaka & Bick, Matthew & Gewe, Mesfin & Quijano-Rubio, Alfredo & Johnson, JayLee & Lenz, Garreck & Nguyen, Alisha & Pun, Suzie & Correnti, Colin & Riddell, Stanley &

Baker, David. (2020). Designed protein logic to target cells with precise combinations of surface antigens. *Science*. 369. eaba6527. 10.1126/science.aba6527.

4. <http://www.rcsb.org/>

5. <https://pdb-redo.eu/>

6. Chen et al. (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallographica D*66:12-21.

7. Davis et al. (2007) MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Research* 35:W375-W383.