

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

ФАКУЛЬТЕТ БИОИНЖЕНЕРИИ И БИОИНФОРМАТИКИ

**Качество расшифровки структуры белка вирусной флавин-
зависимой тимидилатсинтазы Thyx (PDB ID: 2CFA) методом
рентгеноструктурного анализа**

*Отчет по структурной биоинформатике студентки 4-го курса
Николаевой Дарьи Дмитриевны*

Преподаватели:

**А. В. Алексеевский
С. А. Спирин
В. Ю. Лунин**

**Москва
2017**

Аннотация

В данной работе была произведена оценка качества расшифровки структуры белка вирусной флавин-зависимой тимидилатсинтазы Thyx (PDB ID: 2CFA) методом рентгеноструктурного анализа с помощью информации на сайте PDB, а также сервисов MolProbity, EDS и других. Производилось как оценивание параметров качества структуры в целом, так и для отдельных остатков. Особое внимание было уделено поиску и анализу маргинальных остатков по следующим основным критериям: расположение на карте Рамачандрана, значение RSR, ротамеры. Итогом работы является вывод о качестве расшифровки данной структуры.

ВВЕДЕНИЕ

Осуществление процесса репликации ДНК требует присутствия в клетках всех живых организмов нуклеотидов – строительных блоков для синтеза нуклеиновых кислот. Тимидилат (2' – дезокситимидин – 5' – монофосфат; dTMP) – обязательный компонент биосинтеза одного из четырех мономеров ДНК – тимина (dTTP). В клетках тимидилат может быть образован двумя способами: путем фосфорилирования тимидина тимидинкиназой либо синтезом *de novo* из уридинмонофосфата (dUMP) с помощью фермента тимидилатсинтазы. Тимидилатсинтаза осуществляет реакцию метилирования уридинмонофосфата по 5 положению пиримидинового кольца.

Тимидилатсинтазы представлены в природе двумя классами: ThyA и ThyX, отличающимися по структуре белка и механизму катализа. Ферменты класса ThyA (EC 2.1.1.45) изучены очень хорошо: они катализируют реакцию восстановительного метилирования dUMP, используя метилентетрагидрофолат одновременно в качестве донора одноуглеродного фрагмента и восстановителя. Семейство ферментов класса ThyX (EC 2.1.1.148) было открыто сравнительно недавно и сразу привлекло внимание исследователей: в структуре тимидилатсинтаз этого класса присутствует новая укладка, которая прочно удерживает кофермент ФАД. Это приобретение привело к изменениям в механизме катализа: так как ФАД опосредует перенос водорода от НАДФН в ходе катализа, метилентетрагидрофолат выступает в этой реакции только в роли донора углерода. С использованием высокоэффективной жидкостной хроматографии у простейшего *Chlamydia trachomatis* был идентифицирован продукт этой реакции – тетрагидрофолат, тогда как катализ ферментом класса ThyA приводит к образованию дигидрофолата.

Детальное выяснение механизма катализа тимидилатсинтазы ThyX представляет значительный интерес для разработки методов детекции и дизайна новых соединений с антимикробными свойствами. С целью прояснения неразрешенных вопросов, касающихся подробностей каталитического механизма тимидилатсинтаз ThyX, коллективом авторов под руководством Sébastien Graziani было проведено исследование высокоактивного фермента из вируса *Paramecium bursaria chlorella virus-1* (PBCV-1). Помимо биохимических экспериментов, позволивших выявить аминокислотные остатки, необходимые для связывания субстрата и протекания каталитического превращения, авторы также получили кристаллическую структуру фермента. Конформация ключевых остатков активного центра в ней отличалась от таковой в структурах, полученных ранее.

Это позволило авторам сделать вывод о том, что активный центр тимидилатсинтаз ThyX претерпевает структурные изменения в ходе катализа.

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

Общая информация о модели

Структура белка вирусной флавин-зависимой тимидинсинтазы Thyx (PDB ID: 2CFA) была получена методом рентгеноструктурного анализа и впервые загружена на сервер PDB 17 февраля 2006 года авторами Graziani, S., Bernauer, J., Skouloubris, S., Graille, M., Zhou, C.-Z., Marchand, C., Decottignies, P., van Tilbeurgh, H., Myllykallio, H., Liebl, U. Изначально структура была опубликована на сайте PDB 16 мая 2006 года, а 17 июля 2013 года была обновлена и добавлена информация о структуре. Расшифровка этой структуры была сопряжена с исследованиями, описанными в статье “Catalytic Mechanism and Structure of Viral Flavin-dependent Thymidylate Synthase ThyX” [1], опубликованной 25 января 2006 года в Journal of Biological Chemistry.

В файле структурных факторов содержится экспериментальная информация обо всех измеренных рефлексах – всего их 22506. Согласно информации на сервере EDS, полнота данных составляет 98.6 %. Общее разрешение структуры составляет 2.30 Å, а разрешения структурных факторов находятся в диапазоне 33.65 - 2.30 Å (а по данным авторов, диапазон составляет 20.00 - 2.30 Å; в остальном данные на сайте PDB, в статье и в выдаче сервисов будут примерно совпадать).

Фазовая проблема была решена методом молекулярного замещения, в качестве гомолога использовалась структура из организма *Thermotoga maritima* TM0449, имеющая разрешение 2.3 Å.

Информация о параметрах кристаллографической ячейки представлена в файле PDB, а также, в наиболее наглядной форме, прямо на сайте PDB (Таблица 1). В левом столбце перечислены длины направляющих векторов кристаллографической ячейки, а в правом – значения углов между ними. Видим, что все углы между векторами прямые. Также в поле CRYST файла PDB содержится информация о группе симметрии: данная структура принадлежит группе симметрии P 21 21 2. В элементарной ячейке модели содержится 4 молекулы (последняя колонка поля CRYST).

Согласно информации в полях BIOMT и MTRX, при образовании олигомера используются как кристаллографические, так и некристаллографические симметрии.

Таблица 1. Информация о параметрах кристаллографической ячейки с сайта PDB.

Length (Å)	Angle (°)
a = 69.26	α = 90.00
b = 76.99	β = 90.00
c = 93.44	γ = 90.00

Асимметрическая единица содержит 2 цепи: А и В, а биологическая единица представляет собой тетрамер, таким образом, для данной структуры биологическая единица больше асимметрической. Интересно, что так как в В цепи есть мутация, то на сайте PDB структура биологической единицы называется гетеротетрамером, хотя в статье белок называют гомотетрамером, что, по всей видимости, определяет его природное состояние.

Значения идентификаторов качества модели в целом

Согласно сервису EDS, значения идентификаторов качества модели в целом для структуры 2CFA таковы:

- R value (work) = 0.213
- R value (free) = 0.262 (1121 рефлексов были отобраны случайным образом для теста; их перечисление нигде не приведено) (плохие значения превышают 0.4, а хорошие – меньше 0.25, так что здесь почти хорошее)
- процент маргинальных остатков на карте Рамачандрана: 1.2% (тогда как норма – 0.5%)
- R value (free) - R value (work) = 0.049 – разница меньше 0.1 говорит о хорошем качестве оптимизации и о том, что «переподгонки» не произошло

Поиск маргинальных остатков

В качестве стартового ресурса для оценки качества структуры белка использовался сервис MolProbity [2]. Данный сервис позволяет добавить атомы водорода к выбранной пользователем структуре, чтобы определить недопустимые наложения атомов. При этом можно выполнить инверсию боковых цепей остатков Asn, Gln, His в случае, если возможны более выгодные конформации этих остатков, нежели имеющиеся в модели. Однако для того чтобы максимально критически подойти к оценке качества рассматриваемой структуры, инверсия остатков в первом запуске MolProbity не проводилась.

В Таблице 2 представлена выдача программы MolProbity. Для удобства оценки полученные значения параметров окрашены: зеленым цветом обозначены параметры, значения которых попали в границы, приемлемые для структуры хорошего качества, желтым и красным цветами выделены незначительные и крупные отклонения от нормы, соответственно.

В данной структуре превышен порог по количеству неверно определенных длин связей (0.32%) и значений торсионных углов (0.57%). Но хуже всего обстоит дело с остатками, находящимися в «плохой» конформации с точки зрения ротамеров: это значение более чем на 10% выходит за пределы нормы.

Зато только один остаток в структуре расположен в запрещенной области карты Рамачандрана, поэтому параметр выделен желтым цветом. Общая доля остатков в разрешенной области превышает 98%, что свидетельствует о хорошем качестве структуры по этому параметру. Желтым цветом также отмечено отклонение C β атомов, встретившееся всего в двух остатках. А вот по положению C α атомов модель осталась в «зеленой зоне»: программой CaBLAM предсказан лишь один остаток с неправильной конфигурацией, а встроенная программа, оценивающая геометрию, не обнаружила ни одного.

Параметр ClashScore, отражающий число недопустимых наложений атомов на 1000, соответствует в данном случае 96 перцентилю. Это означает, что 96% структур с похожим разрешением имеют ClashScore хуже, чем данная. Интегральная оценка MolProbity score, объединяющая значения оценок ClashScore, ротамеров и областей по

карте Рамачандрана, соответствует 77 перцентилю, что свидетельствует о хорошем качестве данной структуры среди структур со сходным разрешением.

Таблица 2. Выдача программы MolProbity

All-Atom Contacts	Clashscore, all atoms:	8.84	96 th percentile* (N=355, 2.3Å ± 0.25Å)	
	Clashscore is the number of serious steric overlaps (> 0.4 Å) per 1000 atoms.			
Protein Geometry	Poor rotamers	32	10.39%	Goal: <0.3%
	Favored rotamers	253	82.14%	Goal: >98%
	Ramachandran outliers	1	0.30%	Goal: <0.05%
	Ramachandran favored	333	98.52%	Goal: >98%
	MolProbity score [^]	2.25	78 th percentile* (N=8909, 2.3Å ± 0.25Å)	
	Cβ deviations >0.25Å	2	0.58%	Goal: 0
	Bad bonds:	10 / 3094	0.32%	Goal: 0%
	Bad angles:	24 / 4194	0.57%	Goal: <0.1%
Peptide Omegas	Cis Prolines:	0 / 9	0.00%	Expected: ≤1 per chain, or ≤5%
Low-resolution Criteria	CaBLAM outliers	1	0.30%	Goal: <1.0%
	CA Geometry outliers	0	0.00%	Goal: <0.5%

Интересно, что сервис предлагает улучшить положение дел: под выдачей есть сноска, где говорится, что внесение инверсий по остаткам Asn/Gln/His позволит избавиться от одного «плохого» ротамера, что улучшит (= уменьшит) ClashScore на 0.17.

Проверим, на какие еще параметры может повлиять внесение инверсий остатков аспарагина, глутамина и гистидина. Сервис предлагает разрешить инверсию 11 остаткам (Рис. 1), при этом только для трех остатков (A/GLN85, B/ASN34 и B/HIS177) есть явные свидетельства того, что инверсия необходима, для оставшихся 8 остатков инверсия лишь незначительно улучшит положение.

Chain	Res#	Alt	Res ID	Orig	Flip	Flip-Orig	Code	Explanation
A	17		ASN	0.87	1.4	0.53	FLIP	Some evidence recommending flip.
A	34		ASN	-0.61	-0.0068	0.6032	FLIP	Some evidence recommending flip.
A	35		GLN	-1.4	-0.19	1.21	FLIP	Some evidence recommending flip.
A	85		GLN	0.49	4.2	3.71	FLIP	Clear evidence for flip.
A	196		ASN	0.37	1.6	1.23	FLIP	Some evidence recommending flip.
A	217		HIS	-0.94	-0.23	0.71	FLIP	Some evidence recommending flip.
B	17		ASN	-0.29	1.1	1.39	FLIP	Some evidence recommending flip.
B	34		ASN	-2.4	-0.17	2.23	FLIP	Clear evidence for flip.
B	136		GLN	-1.1	-0.017	1.083	FLIP	Some evidence recommending flip.
B	177		HIS	-2.3	0.59	2.89	FLIP	Clear evidence for flip.
B	185		ASN	-0.78	-0.063	0.717	FLIP	Some evidence recommending flip.

Рис. 1. Остатки, которые предлагается инвертировать по данным сервиса MolProbity.

Выдача программы после выполнения инверсии 11 остатков представлена в Таблице 3. Действительно, за счет перевода одного ротамера в более выгодную конформацию ClashScore и MolProbity score немного улучшаются. Интересно, что «плохих» ротамеров стало на один меньше, при этом «хороших» стало больше на два.

Таблица 3. Выдача программы MolProbity с выполнением инверсий

All-Atom Contacts	Clashscore, all atoms:	8.84		96 th percentile* (N=355, 2.3Å ± 0.25Å)
	Clashscore is the number of serious steric overlaps (> 0.4 Å) per 1000 atoms.			
Protein Geometry	Poor rotamers	32	10.39%	Goal: <0.3%
	Favored rotamers	253	82.14%	Goal: >98%
	Ramachandran outliers	1	0.30%	Goal: <0.05%
	Ramachandran favored	333	98.52%	Goal: >98%
	MolProbity score [^]	2.25		78 th percentile* (N=8909, 2.3Å ± 0.25Å)
	Cβ deviations >0.25Å	2	0.58%	Goal: 0
	Bad bonds:	10 / 3094	0.32%	Goal: 0%
Bad angles:	24 / 4194	0.57%	Goal: <0.1%	
Peptide Omegas	Cis Prolines:	0 / 9	0.00%	Expected: ≤1 per chain, or ≤5%

Далее будут отдельно рассмотрены маргинальные остатки по каждому из параметров.

Остатки из запрещенной области карты Рамачандрана

Сервис MolProbity позволяет получить набор карт Рамачандрана для шести разных случаев:

- общая карта;
- изолейцины и валины;
- остатки перед пролином;
- глицины;
- *транс*-пролины;
- *цис*-пролины.

В данной структуре имеется только один остаток из запрещенной области – VAL29 из цепи А. Он показан на карте для изолейцинов и валинов (в правом верхнем углу на Рис. 2). Он расположен между двумя разрешенными областями, от которых отклоняется по значению угла ψ .

Карту Рамачандрана для данной структуры строит также сервер EDS [3]. По этим данным, уже четыре остатка из структуры находятся в запрещенной области. Помимо обнаруженного ранее VAL24, за пределами разрешенной области находятся следующие остатки:

- SER7 (цепи A и B)
- LYS151 (цепь B)

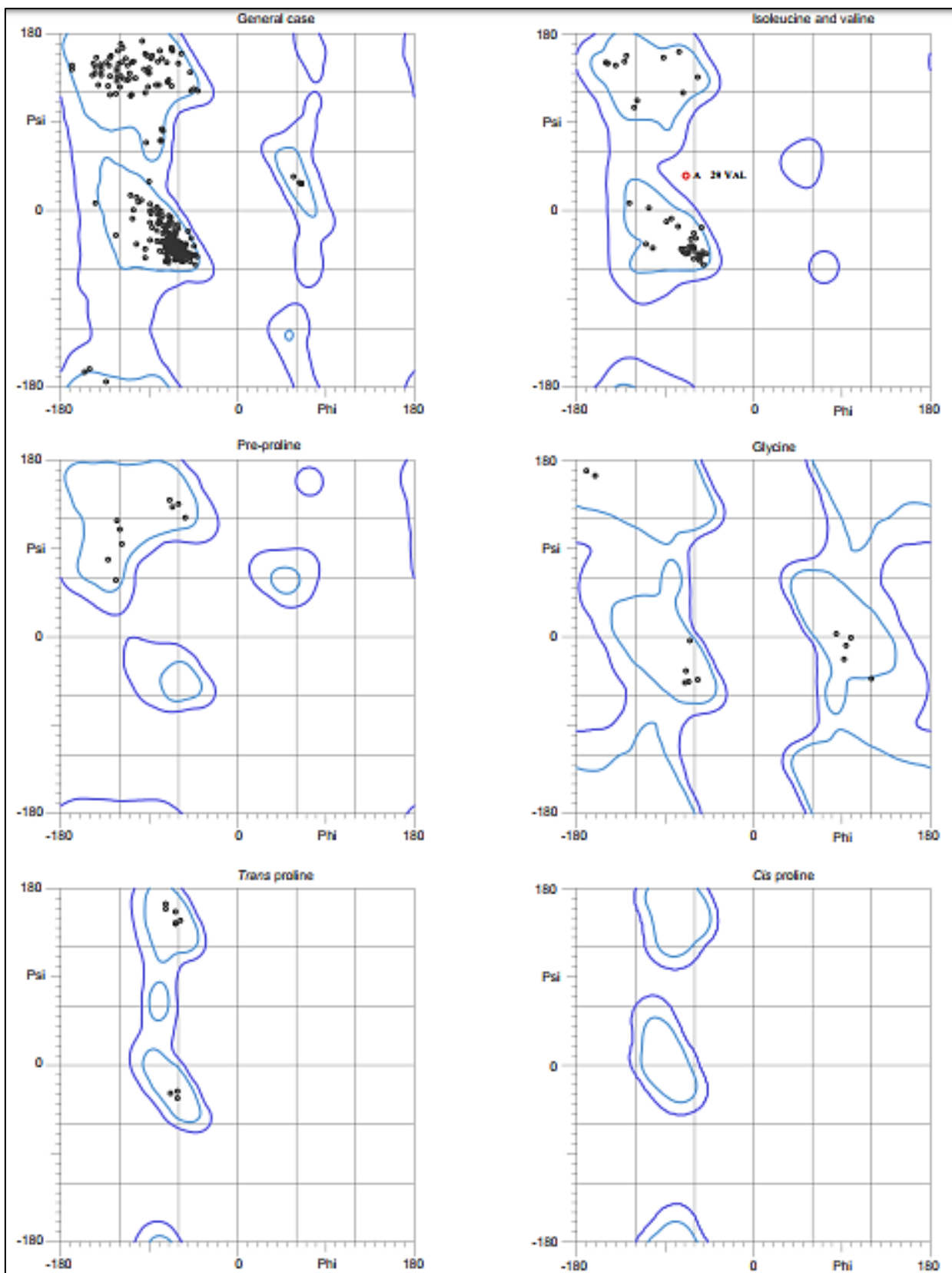


Рис. 2. Карты Рамачандрана с сервиса MolProbity. Голубыми линиями выделены области, содержащие 98% остатков, темно-синими - > 99.8% остатков.

Остатки, маргинальные по RSR

Пространственный R-фактор (RSR) – параметр, который характеризует, насколько модель группы атомов (или даже отдельного атома) соответствует “экспериментальной” электронной плотности. Этот параметр позволяет как оценивать структуру в целом, так и осуществлять поиск отдельных маргинальных остатков, чье положение боковой цепи не соответствует действительности. Считается, что $RSR > 20\%$ является показателем маргинального остатка.

На сервере EDS представлена информация не только о значениях RSR для каждого остатка, но и более удобная относительная оценка RSR – Z-score. Она показывает, насколько RSR для остатка отличается от среднего RSR для такого же типа остатков в структурах PDB со сходным разрешением. Z-score – более простая и информативная оценка, так как, по всей видимости, тип и расположение остатков влияют на качество их расшифровки: бывают случаи, когда RSR имеет большое значение, а Z-score укладывается в норму, значит, остаток расшифрован плохо, но не хуже, чем в среднем по структурам для этого типа остатков.

На Рис. 3 представлены значения Z-score для каждого остатка по каждой цепи. Высокие пики, где $Z\text{-score} > 2$ (то есть RSR для конкретного остатка отклоняется от среднего значения RSR для таких же остатков по всем структурам более чем на два стандартных отклонения), соответствуют маргинальным остаткам по RSR. Информация о них отдельно доступна на сервере EDS и приведена на Рис. 4.

Цепь А разрешена несколько хуже: в ней 26 остатков с $Z\text{-score} > 2$ (14%), тогда как в цепи В их 19 (11%). Можно заметить, что маргинальные остатки по RSR имеют тенденцию кластеризоваться: в обеих цепях много таких остатков в диапазоне номеров 124-156, а в цепи А есть также кластер остатков с номерами 30-40. Интересно, что первая группа маргинальных остатков расположена в длинной альфа-спирали (остатки 126-146), что довольно странно, так как обычно остатки в регулярных структурах разрешаются лучше, чем в свободных петлях. К тому же, данная альфа-спираль находится не на самом конце молекулы: после нее еще около 30 остатков по ходу цепи разрешены хорошо. Заметим также, что единственный маргинал по карте Рамачандрана (VAL26 в цепи А) не является маргиналом по RSR, то есть хорошо вписывается в «экспериментальную» электронную плотность.

Плохие RSR: 1+30-35+89+124-127+139-141+143-148+156+185+217 (А)

1+125-128+134+138+139+143+145+146+148-150+152+153+155+187+189 (В)

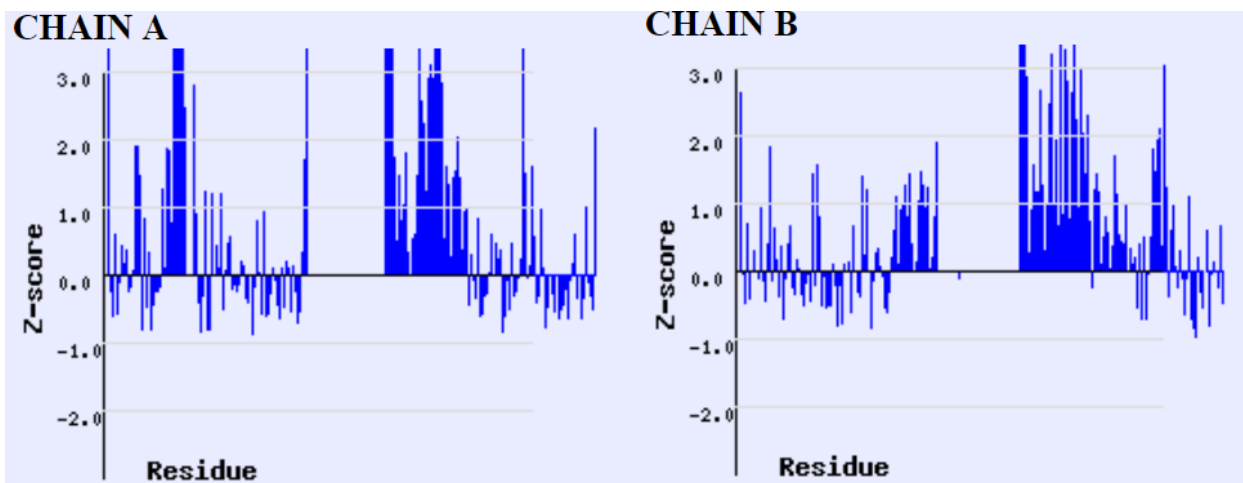


Рис. 3. Значения Z-score для каждого остатка цепи А (слева) и цепи В (справа). $Z\text{-score} = (RSR - \langle RSR_{\text{resolution}} \rangle) / \text{Sigma}_{\text{resolution}}$, где $\langle RSR_{\text{resolution}} \rangle$ - среднее значение RSR для остатков данного типа по всем структурам с похожим разрешением, а $\text{Sigma}_{\text{resolution}}$ – стандартное отклонение для набора разрешений.

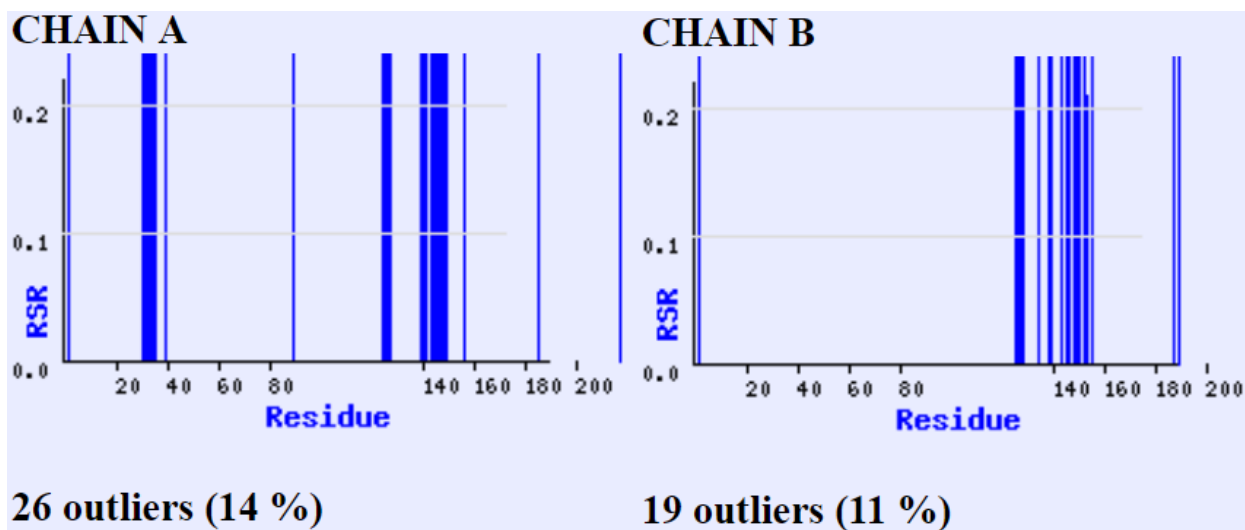


Рис. 4. Положение остатков, маргинальных по RSR ($Z\text{-score} > 2$).

Маргинальные ротамеры

Сервис MolProbity выдает информацию об остатках, являющихся «плохими» ротамерами. Ротамерами называются боковые цепи в типичных для данного остатка конформациях, которые определяются значениями углов между атомами в боковой цепи.

Всего структура содержит 32 «плохих» ротамера, таким образом, это основной фактор, понижающий качество расшифровки структуры. Так как две цепи практически идентичны, за исключением мутации L77I в цепи В, то можно проследить, какие из маргинальных остатков присутствуют в обеих цепях (Таблица 4).

Таблица 4. Информация о плохих ротамерах с сервиса MolProbity

Остаток	Chi angles	Остаток	Chi angles
A 1 MET	chi angles: 77.7,284.2,171.2		
		B13 VAL	chi angles: 331.9
		B36 ILE	chi angles: 210.1,101.1
A44 LEU	chi angles: 211.3,216	B44 LEU	chi angles: 217.1,215.1
		B45 LEU	chi angles: 272.5,8.2
		B63 LEU	chi angles: 133.9
A64 LEU	chi angles: 260.3,40.1		
A77 LEU	chi angles: 266.4,15		
		B131 GLY	chi angles: 252.2,52.3,234.2
A132 LYS	chi angles: 208,206.3,268.1,83.6	B132 LYS	chi angles: 262.7,104.3,99.1,299.8
A143 LYS	chi angles: 279.8,269.7,261.4,65.1		
		B152 GLU	chi angles: 46.3,104.7,300.5
A155 ARG	chi angles: 243.9,178.4,43.9,86.5	B155 ARG	chi angles: 245.1,165.5,17.3,92.2
A157 ILE	chi angles: 212.4,144.1		
A158 LEU	chi angles: 268.1,27.9	B158 LEU	chi angles: 268.1,27.9
A160 LEU	chi angles: 277.8,342.2		
A162 THR	chi angles: 161		
A165 THR	chi angles: 152.1		
		B166 ILE	chi angles: 67.8,265.8
A172 ILE	chi angles: 269.7,29.2		
A187 THR	chi angles: 98.6	B187 THR	chi angles: 153.9
A189 ARG	chi angles: 272.7,20.7,141.9,62.3	B189 ARG	chi angles: 180.1,97.7,252.3,118.2
		B190 GLU	chi angles: 29.5,280.3,315.8

A194 LEU	chi angles: 226.3,205.3	B194 LEU	chi angles: 221.9,207.1
		B200 GLU	chi angles: 154,132.9,117.2

Не так уж много маргинальных остатков встречается в обеих цепях, всего 7: LEU44, LYS132, ARG155, LEU158, THR187, ARG189, LEU194. Надо будет позже обратить на эти остатки внимание, так как то, что они встретились в обеих цепях, снижает вероятность того, что это случайные проблемы расшифровки. То есть это либо действительно уникальные конформации, либо надо думать, почему дефект расшифровки произошел в одном и том же месте.

Среди «плохих» ротамеров встречаются только определенные типы остатков: в цепи А – ILE, LEU, LYS, ARG, THR, в цепи В к ним добавляются GLU, VAL и GLY. За исключением глицина, боковые цепи этих остатков большие, содержат много атомов и, следовательно, много χ -углов, поэтому неудивительно, что именно у них могут возникнуть проблемы по трактовке этих углов.

Может показаться на первый взгляд, что «плохие» ротамеры тоже кластеризованы. Однако у данной структуры просто исходно нет информации о координатах остатков 89-122 (в цепи А и немного меньше в цепи В) по данным на сайте PDB [3].

В Таблице 5 перечислена информация о 15 маргинальных остатках, которые, как правило, маргинальны по нескольким признакам. Здесь, помимо прочего, присутствуют 3 остатка, которые предлагалось инвертировать, а также 6 из 7 остатков, которые являются плохими ротамерами в обеих цепях.

Таблица 5. Информация о маргинальных остатках в структуре 2CFA

Остаток	Критерии маргинальности остатка							Другое
	Карта Рамачандрана	Ротамеры	C β -конформация (>0.25 Å)	RSR (Z-score)	Clash	Bad bonds	Bad angles	
A/LYS143		+ (0%)		+ (2.89)				
B/ALA150				+ (B)		+ (S)	+ (S)	

						worst is C-- N: 86.6 σ	worst is CA-C- N: 29.1 σ	
B/THR63		+ (0%)	+ (0.33Å)		0.56Å HB with B 169 SER OG			
B/LYS151			+ (0.70Å)				+ (S) worst is N-CA- CB: 8.4 σ	
A/VAL29	+ (0.01 %) Ile or Val / - 71.0,3 5.4				0.58Å O with A 29 VAL HG22			CaBL AM Outlier (0.077 %)
A/GLN85								Clear flip
B/ASN34					0.91Å HB3 with B 36 ILE HG22			Clear flip
B/HIS177					0.57Å HD2 with B2040 HOH O			Clear flip
ARG155		+ (0.2%)		+ (B)				

		A, (0%) B						
LEU44		+ (0%) (A+B)						
LYS132		+ (0.1%) A, (0%) B						
LEU158		+ (0%) (A+B)			0.42Å HB3 with A 162 THR HG21			
THR187		+ (0%) A, (0.1%) B		+ (B)	A: 0.60Å HG21 with A 182 ARG HA, B: 0.65Å HG21 with B 182 ARG HA			
ARG189		+ (0.2%) B, (0%) A		+ (B)	A: 0.41Å HD3 with A 189 ARG N			
LEU194		+ (0%) (A+B)						

Подробное обсуждение 5 маргинальных остатков

Для более пристального рассмотрения были выбраны остатки, которые являются маргинальными либо по совокупности признаков, либо по одному, но тогда имеют по нему наихудшие значения среди всех маргиналов.

Пример 1. A/LYS143

Этот остаток является маргинальным по конформации ротамеров и значению RSR. Последний факт уже сам по себе означает то, что остаток недостаточно хорошо вписывается в «экспериментальную» функцию электронной плотности. Однако Z-score по RSR для этого остатка еще не самый плохой по структуре, поэтому на Рис. 5 остаток частично вписан в функцию электронной плотности. Данный остаток находится в той самой длинной альфа-спирали, которая в целом расшифрована хуже, чем остальная структура (Рис. 6), поэтому неудивительно, что Рис. 5 получен при уровне подрезки всего 0.4 σ . Удивительно другое: на Рис. 5 функция электронной плотности образует как бы два альтернативных варианта расположения боковой цепи лизина. Причем среди полученных с помощью сервиса MolProbity углов χ (279.8, 269.7, 261.4, 65.1) у лизинов не наблюдаются такие значения для углов χ_2 и χ_3 (по данным сервиса Rotamers revisited [4] – Рис. 7), а вращение именно по этим углам как раз привело бы боковую цепь LYS143 в другую, альтернативную конформацию. По всей видимости, плохое качество расшифровки или дефекты при кристаллизации привели к возникновению дополнительной, более значимой, чем правильная, области, где определена функция электронной плотности, в которую и оказалась вписанной боковая цепь лизина.

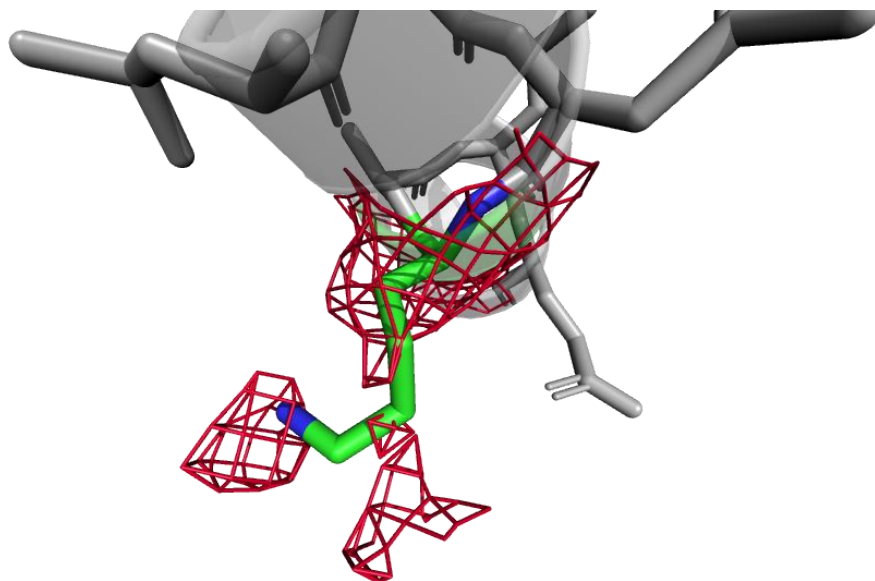


Рис. 5. Функция электронной плотности для остатка A/LYS143 на уровне подрезки 0.4 σ .

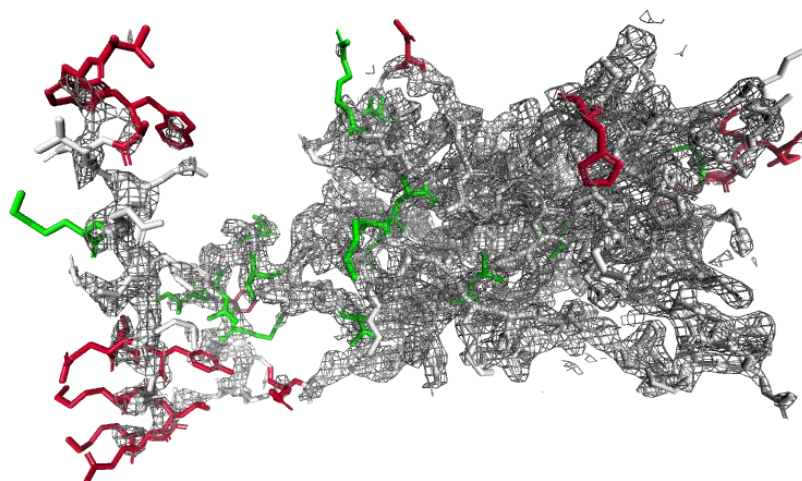


Рис. 6. Функция электронной плотности для цепи A на уровне подрезки 1 σ . Длинная альфа-спираль очень плохо разрешена. Остатки, покрытые электронной плотностью, почти все являются плохими ротамерами (красные) или имеют плохое значение RSR (зеленые).

Так как у нас есть вторая почти идентичная цепь, я решила найти этот остаток в ней, тем более что про него неизвестно, что он плохой ротамер. На Рис. 8 изображена функция электронной плотности вокруг B/LYS143 с тем же уровнем подрезки 0.4 σ . Электронная плотность здесь определена еще хуже, чем для лизина в цепи A, – это согласуется с более плохим значением Z-score по RSR для B/LYS143 (3.52 против 2.89 в A/LYS143). Однако измерение торсионных углов показало, что они все соответствуют распространенным вариантам для Lys (надо смотреть на Рис. 7, и если значения

отрицательные, прибавлять к ним 360). Это наблюдение еще больше подтверждает то, что положение «плохого» ротamera A/LYS143 является результатом ошибки в расшифровке, а не уникальной конформацией.

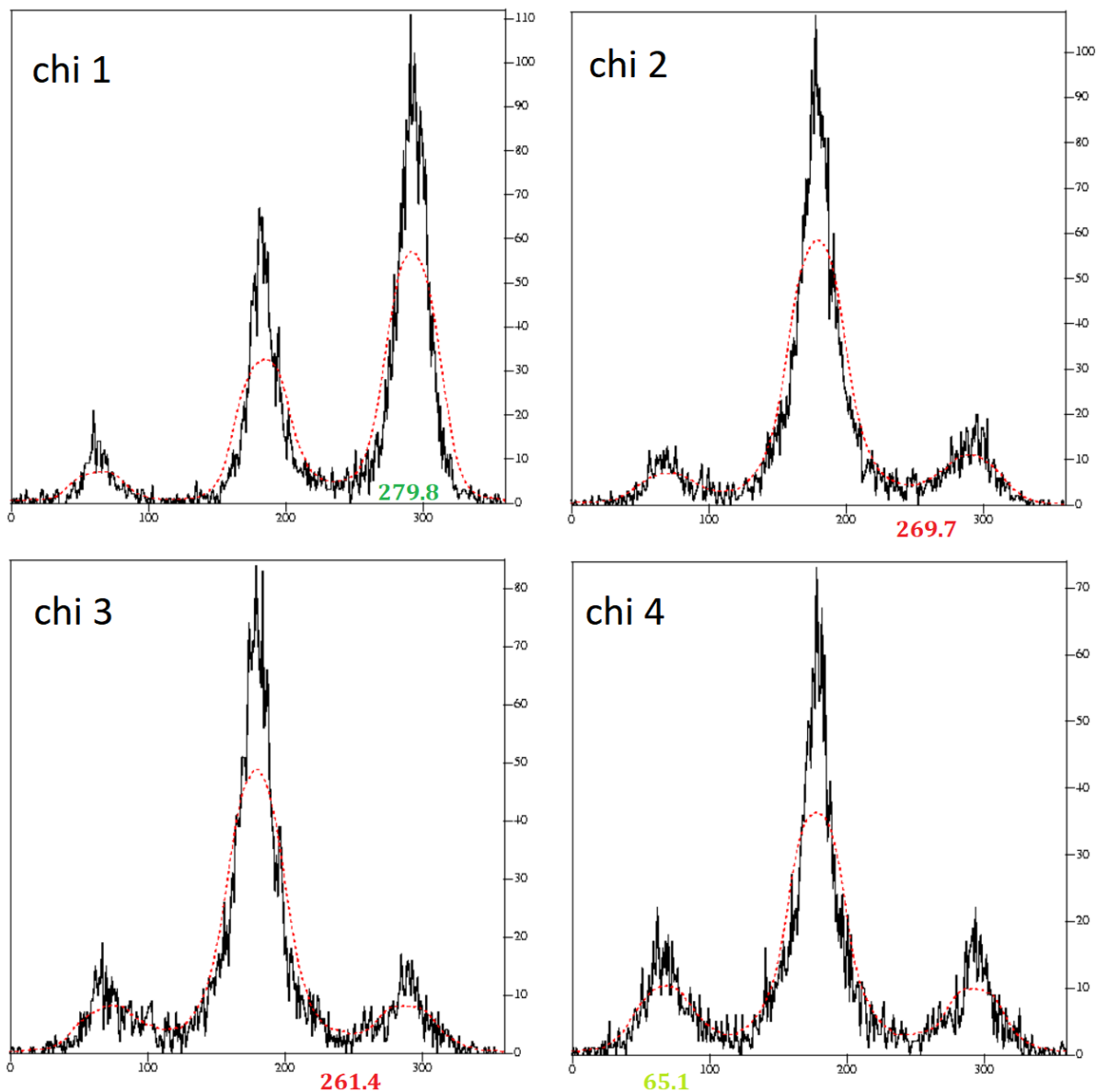


Рис. 7. Распределения конформаций χ углов для лизинов. На рисунке для каждого из углов отмечены значения углов для остатка A/LYS143. Цвет чисел отражает, насколько характерно полученное значение для лизина: темно-зеленый – наиболее характерное, салатный – не максимальный из пиков, красный – наименее вероятное значение.

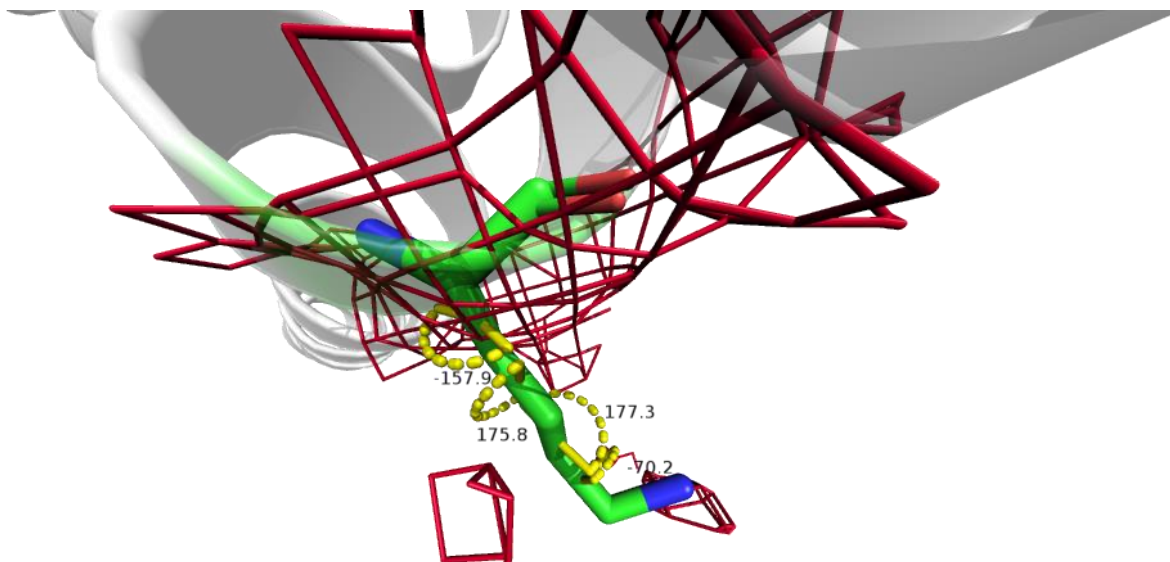


Рис. 8. Функция электронной плотности для остатка В/LYS143 на уровне подрезки 0.4 σ . Желтыми пунктирными дугами обозначены измеренные торсионные углы χ .

Я также попыталась определить причину возникновения «фантомной» области электронной плотности, в которую ошибочно вписалась боковая цепь А/LYS143. Длинная альфа-спираль, в которой расположен этот остаток, участвует в тетрамеризации [1], а асимметрическая ячейка представлена только димером. Чтобы проверить, не «приходит» ли эта электронная плотность от остатков из соседних ячеек, я восстановила ближайшие к этому остатку соседние ячейки (Рис. 9). Ближайшие соседние остатки расположены довольно далеко, чтобы повлиять на наш лизин. Никаких молекул воды, на которые тоже можно было бы списать эту область функции электронной плотности, в этом промежутке не обнаружено, и вообще, известно, что тетрамер удерживается за счет стеккинг-взаимодействий между этими длинными альфа-спиралями [1]. Так что можно только сказать, что все дело в плохом качестве расшифровки этого участка белка.

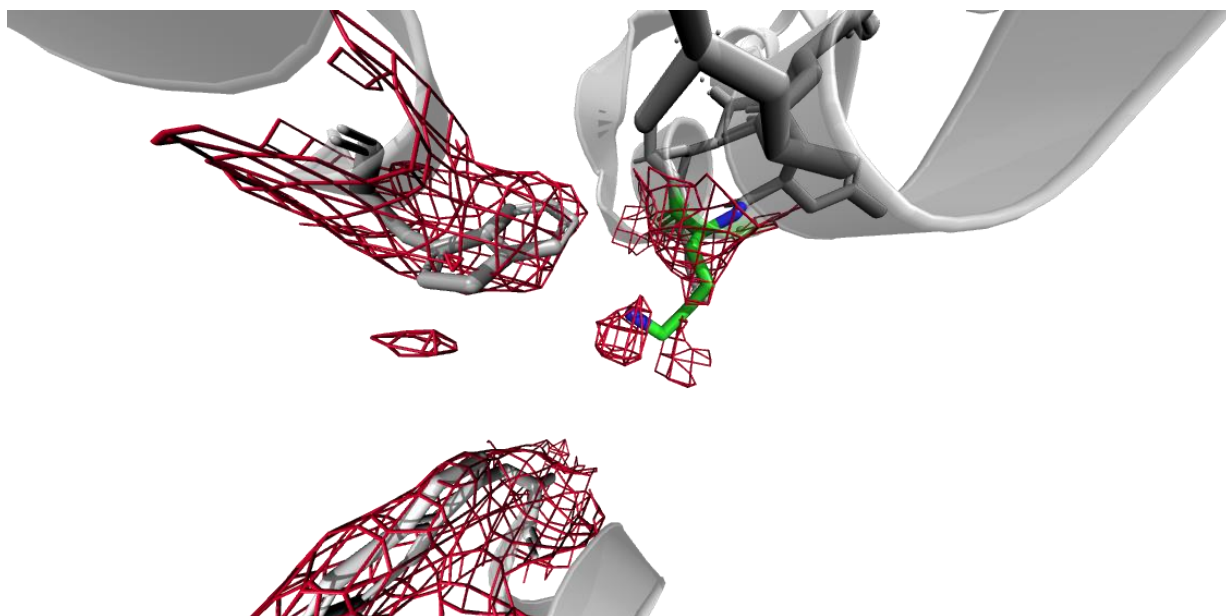


Рис. 9. Восстановление ячеек, соседних с A/LYS143 (показан зеленым).

Пример 2. LEU44 – «плохой» ротамер

При обзоре ротамеров были найдены примеры остатков с плохими конформациями боковых цепей в обеих цепях белка. Интересно, что все «плохие» ротамеры в цепи А, кроме A/LYS143, имеют хороший Z-score по RSR (в цепи В совпадений побольше, но все равно основная тенденция – эти множества мало пересекаются).

Остаток LEU44 является «плохим» ротамером в обеих цепях, при этом в обеих же цепях у него хороший Z-score по RSR, то есть, казалось бы, нельзя списать неправильную конформацию боковой цепи только лишь на несоответствие «экспериментальным» данным. Однако на Рис. 10 видно, что на уровне подрезки 1.5σ LEU44 (зеленый) гораздо хуже описан функцией электронной плотности, чем соседний с ним остаток LEU45 (голубой). При этом ситуация в цепи В (Рис. 10 справа) несколько лучше, чем в А (Рис. 10 слева): значение Z-score по RSR для этого остатка в цепи В также лучше, чем в цепи А.

На Рис. 11 показана функция электронной плотности на более низком уровне подрезки (0.8σ) и только для LEU44. Видно, что боковая цепь вписана в функцию электронной плотности неидеально: возможно, если немного поварьировать положение боковой цепи, можно добиться лучшего результата и попробовать избавиться от плохого ротамера.

Также может оказывать влияние расположение остатка в месте структурного изгиба, на границе альфа-спирали и свободной петли. Более того, рядом в петле есть

нерасшифрованный участок (показан пунктиром), что должно повышать подозрения к качеству расшифровки и LEU44.

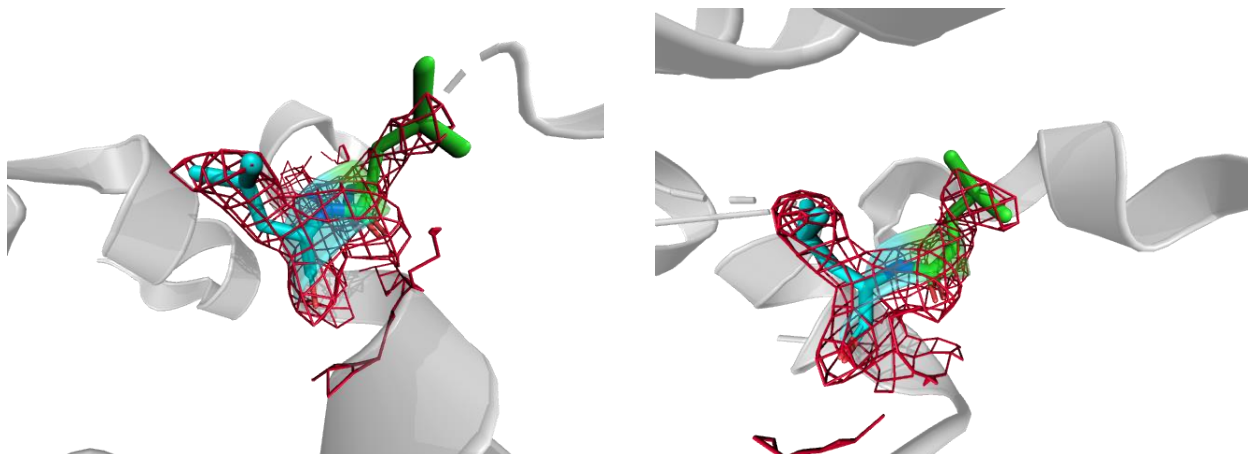


Рис. 10. Функция электронной плотности для остатков LEU44 (зеленый) и LEU45 (голубой) из цепей А (слева) и В (справа) и на уровне подрезки 1.5 σ .

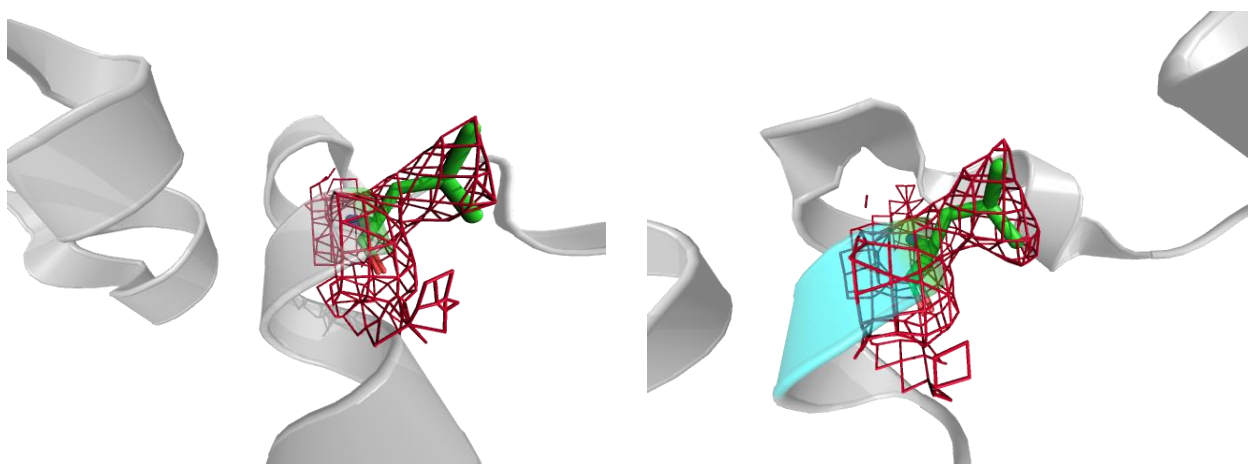


Рис. 11. Функция электронной плотности для остатка LEU44 (зеленый) из цепей А (слева) и В (справа) и на уровне подрезки 0.8 σ .

Пример 3. A/VAL29

Этот остаток собрал больше всего маргинальных оценок: он единственный маргинал по карте Рамачандрана, программа CaBLAM назвала его маргиналом по положению С-альфа атома (0.077%), а еще он находится в участке белка (остатки 28-31), в котором перекрываются атомы разных остатков (clashes). На Рис. 12 показана функция электронной плотности для остатка A/VAL29 на уровне подрезки 1.2 σ . Остаток является пограничным между альфа-спиралью и свободной петлей, в этом месте происходит изгиб, ввиду чего электронная плотность от соседних остатков перекрывается (Рис. 13 – получен из файла *H.pdb, содержащего водороды, добавленные программой MolProbity). Думаю,

что этот факт осложнил вписание остатков в электронную плотность, поэтому на этом участке там много стерических затруднений и перекрытий, а A/VAL29 оказался плохо вписанным в ФЭП и маргинальным по карте Рамачандрана.

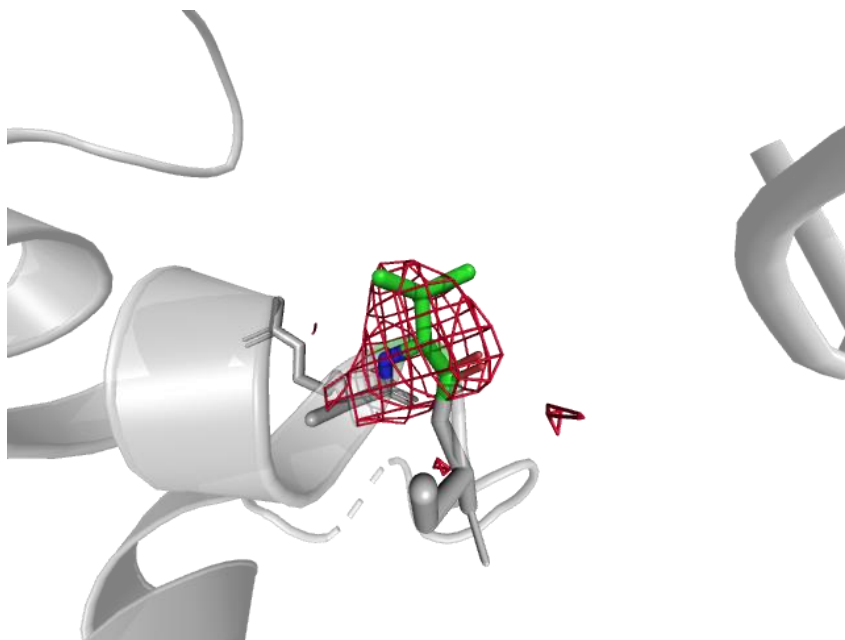


Рис. 12. Функция электронной плотности для остатка A/VAL29 на уровне подрезки 1.2 σ .

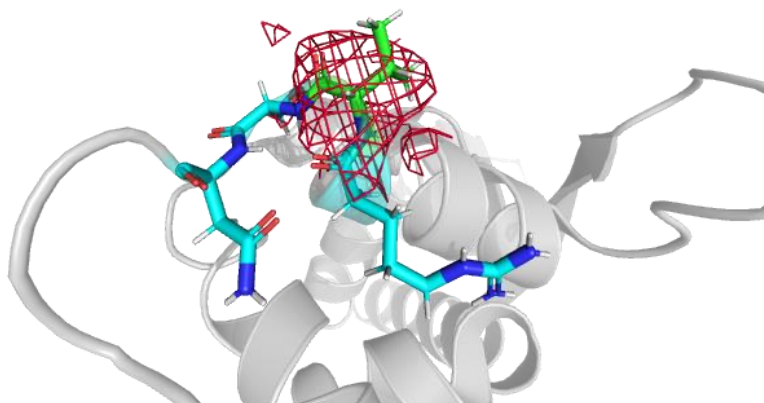


Рис. 13. Функция электронной плотности для остатка A/VAL29 (зеленый) на уровне подрезки 1 σ . Голубым показаны остатки, имеющие перекрытия по атомам.

Пример 4. Под кодовым названием «ошибка на ошибке» (B/ALA150)

У этого остатка сервис MolProbity обнаружил очень неестественные связь (worst is C--N: 86.6 σ) и угол (worst is CA-C-N: 29.1 σ). Оказалось, что между аланином и следующим лизином есть небольшой участок, который обозначен как место проблемы с расшифровкой (бледно-зеленый толстый пунктир). Но программа все равно распознает аланин и лизин как соседние остатки, поэтому и находит между ними странный угол и слишком длинную пептидную связь. При этом где на самом деле должен находиться аланин, тоже неясно: в электронную плотность он не попадает совсем (Рис. 14). Таким образом, мы имеем здесь дело не только с неточностью расшифровки, но и с локальной нехваткой информации.

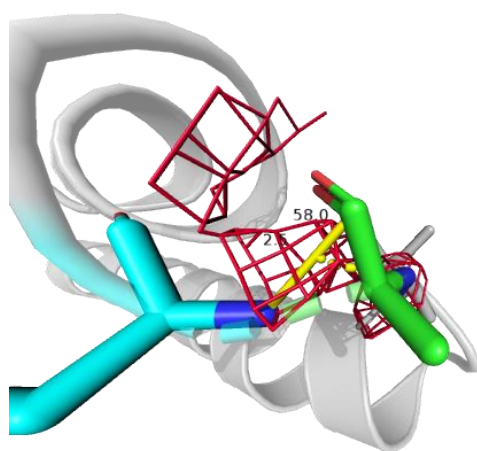


Рис. 14. Функция электронной плотности для остатка B/ALA150 (зеленый) на уровне подрезки 1.5 σ . Голубым показан соседний остаток B/LYS151. Измерена длина пептидной связи (желтая линия) и угол (желтая пунктирная дуга).

Пример 5. The importance of being inverted (B/ASN34)

На Рис. 15 показаны функции электронной плотности для остатка B/ASN34 до (слева) и после (справа) выполнения инверсии с помощью сервиса MolProbity. Хорошо видно, что исходное расположение ошибочно, а после инверсии боковая цепь гораздо лучше вписывается в электронную плотность.

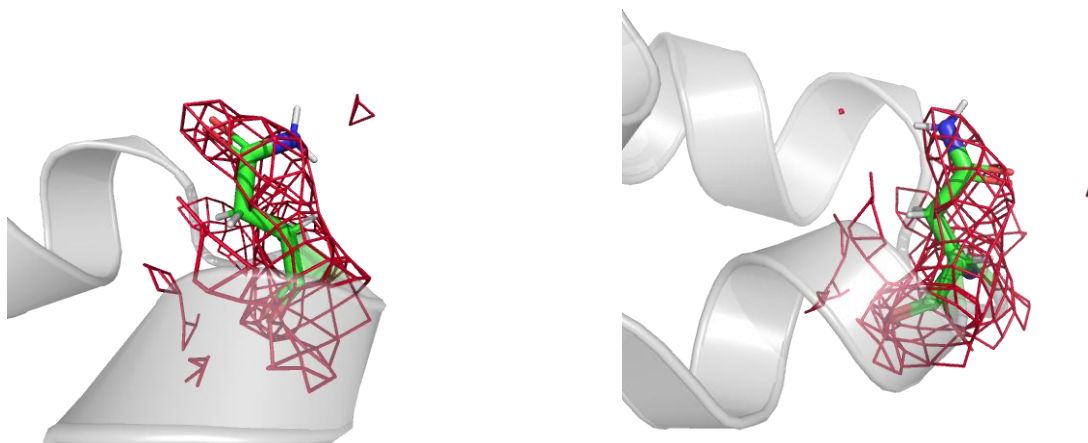


Рис. 15. Функция электронной плотности для остатка V/ASN34 до (слева) и после (справа) выполнения инверсии с помощью сервиса MolProbity (на уровне подрезки 1σ).

Сравнение модели из PDB с моделью из PDB-REDO

Сервис PDB-REDO позволяет выполнить улучшение модели структуры. В самой выдаче данной программы приводится сравнение параметров качества расшифровки для исходной модели и для модели из PDB-REDO. Зеленым цветом обозначены параметры, значения которых улучшились в модели PDB-REDO (Таблица 6).

Таблица 6. Сравнение параметров оценки качества расшифровки для моделей из PDB и PDB-REDO

Validation metrics from PDB-REDO		
	PDB	PDB-REDO
<i>Crystallographic refinement</i>		
<i>R</i>	0.2147	0.1704
<i>R-free</i>	0.2574	0.2162
<i>Bond length RMS Z-score</i>	0.469	0.661
<i>Bond angle RMS Z-score</i>	0.651	0.850
<i>Model quality (raw scores percentiles)</i>		
<i>Ramachandran plot appearance</i>	16	37
<i>Rotamer normality</i>	3	14

Видим, что у нас улучшились Rfree, процент нахождения остатков на карте Рамачандрана и процент ротамеров. Однако мне такое представление информации

показалось не слишком наглядным и доказательным, поэтому я подключила сервис MolProbity к оценке качества модели из PDB-REDO (Таблица 7).

Таблица 7. Оценка качества расшифровки структуры для модели из PDB-REDO

All-Atom Contacts	Clashscore, all atoms:	2.5		100 th 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	16	5.18%	Goal: <0.3%
	Favored rotamers	267	86.41%	Goal: >98%
	Ramachandran outliers	1	0.30%	Goal: <0.05%
	Ramachandran favored	331	97.93%	Goal: >98%
	MolProbity score [^]	1.59		98 th percentile* (N=8909, 2.30Å ± 0.25Å)
	Cβ deviations >0.25Å	4	1.15%	Goal: 0
	Bad bonds:	18 / 3097	0.58%	Goal: 0%
	Bad angles:	23 / 4198	0.55%	Goal: <0.1%
Peptide Omegas	Cis Prolines:	0 / 9	0.00%	Expected: ≤1 per chain, or ≤5%
	Twisted Peptides:	2 / 352	0.57%	Goal: 0
Low-resolution Criteria	CaBLAM outliers	0	0.00%	Goal: <1.0%
	CA Geometry outliers	0	0.00%	Goal: <0.5%

Действительно, сильно улучшилась статистика по ротамерам (с 33 до 16). А вот ситуация с картой Рамачандрана несколько ухудшилась: уменьшилось количество остатков в благоприятной зоне, хотя маргинал по-прежнему только один. Количество «плохих» связей увеличилось на целых 8, а «плохих» углов – уменьшилось на 1. Также увеличилось количество остатков, отклоняющихся по С-бета атому. Что наиболее удивительно, появилось 2 неких перекрученных пептида. А отклонение одного остатка по положению С-альфа атома исчезло. Несмотря на кажущуюся спорность улучшения, параметры Clashscore и MolProbity score выросли очень сильно: до 100 и 98 перцентилей, соответственно (а были 96 и 78). Получается, что отрицательный вклад плохих ротамеров настолько велик, что программа улучшения пытается в первую очередь избавиться от них, даже ценой появления некоторых других проблем.

ЗАКЛЮЧЕНИЕ

В целом впечатления об оценке качества расшифровки структуры 2CFA соответствуют положению и цвету бегунков на отчете по валидации с сайта PDB (Рис. 16). В принципе, структура имеет атомарное разрешение и по большинству параметров качества находится еще в пределах нормы. Однако большое количество плохих ротамеров и остатков с плохим RSR сильно портят качество структуры. Более того, многие из этих маргиналов так или иначе связаны с участками структуры, информация о которых вообще отсутствует (~90-122), что тоже сильно вредит качеству структуры. Что касается связи структуры и биологического смысла, вновь напомню, что участок длинной альфа-спирали (остатки ~130-150) разрешен настолько плохо, что не стоит доверять структурной информации о нем, и, тем не менее, именно он отвечает за образование биологически активного тетрамера. В итоге, можно сказать, что структурой для исследований пользоваться можно, но придется обращаться к функции электронной плотности, так как модель не вполне соответствует «экспериментальным» данным.

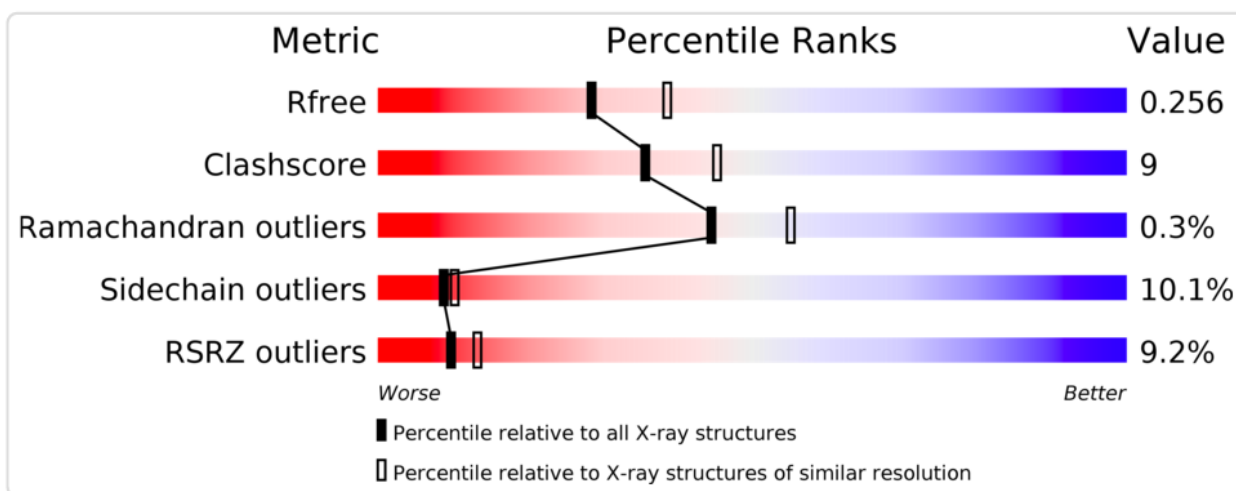


Рис. 16. Информация о валидации с сайта PDB.

Список литературы и ссылок

1. Graziani S. et al. Catalytic Mechanism and Structure of Viral Flavin-dependent Thymidylate Synthase ThyX // J. Biol. Chem. 2006. Vol. 281, № 33. P. 24048–24057.
2. Main page - MolProbity [Electronic resource]. URL: <http://molprobity.biochem.duke.edu/> (accessed: 28.11.2017).
3. EDS - Uppsala Electron Density Server [Electronic resource]. URL: <http://eds.bmc.uu.se/eds/> (accessed: 03.12.2017).
4. Rotamers revisited [Electronic resource]. URL: <http://xray.bmc.uu.se/gerard/supmat/chi.html> (accessed: 05.12.2017).