

МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

имени М.В. ЛОМОНОСОВА

ФАКУЛЬТЕТ БИОИНЖЕНЕРИИ И БИОИНФОРМАТИКИ

**Отчёт о качестве расшифровки структуры
4НМ, полученной методом
рентгеноструктурного анализа**

**студентки IV курса
Яровенко Светлана**

Москва

2017

Содержание

Аннотация	2
Введение	2
Общая информация о модели 4HJM	2
Результаты и обсуждение	4
Общие показатели качества структуры	4
MolProbity	5
Карта Рамачандрана	6
Пространственный R-фактор (RSR и RSR-Z-score)	6
Комфортность окружения	7
Температурный фактор	8
Показатели качества расшифровки лигандов	9
Анализ маргинальных остатков	10
Перечень маргинальных остатков	10
Лиганд MPD	10
Остаток Lys55	11
Остаток Glu102	11
Остаток Glu6	12
Остаток Asn111	12
PDB_redo	13
Выводы	15
Список использованной литературы и сервисов	16

Аннотация

В данной работе анализируется качество расшифровки модели 4НЖМ глутаредоксина 1 из *Plasmodium falciparum*, которая была получена с помощью рентгеноструктурного анализа. Главной целью было найти неточности в данной расшифровке, а также по возможности ошибки. При этом дополнительно было проведено сравнение исходной модели с улучшенной и сделаны соответствующие выводы.

Введение

Общая информация о модели 4НЖМ

4НЖМ — структура белка глутаредоксина 1 из *Plasmodium falciparum*. Этот белок входит в семейство небольших тиол-дисульфидных оксидоредуктаз и катализирует реакции тиол-дисульфидного обмена, используя глутатион в качестве кофактора. На **рис. 1** представлен общий вид белка, раскрашенный в соответствии с элементами вторичной структуры. Этот белок является мономером, состоит из одной цепи и обладает характерной для глутаредоксинов укладкой: четырёхсоставным β -листом, который окружён несколькими α -спиралями [1]. Было показано, что глутаредоксин 1 *P. falciparum* участвует вместе с белком-антиоксидантом в восстановлении гидропероксидов [2].

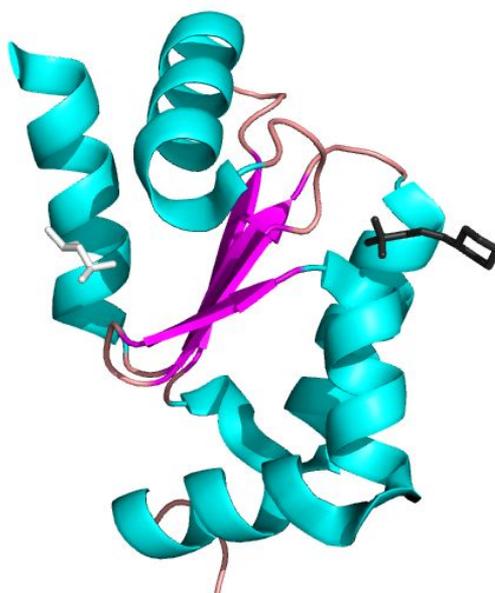


Рис. 1. Общий вид 4НЖМ вместе с двумя лигандами — МРД (белый) и МРО (черный).
Белок окрашен в соответствии с элементами вторичной структуры.

Помимо цепи самого глутаредоксина в состав в комплекса в 4НЖМ входят два лиганда – 3-[N-Морфолино]пропансульфоновая кислота (MOPS или MPO в PDB-файле, чёрная структура на рис. 1) и 2-Метил-2,4-Пентандиол (MPD, белая структура на рис. 1). Последний находится в уникальном для глутаредоксина 1 гидрофобном кармане [1]. Также присутствует довольно много молекул воды (176).

Структура была получена в 2014 году Amit Sharma, Manickam Yogavel и коллегами с разрешением в 1.55 Å. Всего в эксперименте было получено 16735 рефлексов, из которых 15734 использовались при оптимизации модели. В статье [1] заявлено, что полнота данных 97.7%, при этом на сайте EDS указано иное число (99%). Аналогичное несоответствие наблюдается и в диапазоне разрешений: в статье — 50.0 - 1.55 Å, для EDS — 41.65 - 1.55 Å. Возможно это связано с тем, что в базе данных хранится уже улучшенная модель.

Фазовая проблема была решена методом одноволновой аномальной дисперсии серы (Single-wavelength anomalous dispersion of S atoms или S-SAD) с использованием программ SHELX. Оптимизация структуры осуществлялась в PHENIX.

Что касается кристаллографической ячейки, в ней представлено 6 молекул-мономеров, а её параметры следующие: $a = b = 48.09 \text{ \AA}$, $c = 82.59 \text{ \AA}$, $\alpha = \beta = 90^\circ$, $\gamma = 120^\circ$. При этом симметрия внутри белка отсутствует (видно по самой структуре (см. рис.1), а также по низкому значению Z-score в SymD: 2.654).

В структуре 4НЖМ присутствуют 14 аминокислот с альтернативными конформациями, к примеру: Ile21, Lys26, Asn43, Lys55, Arg77, Asn81, Lys82, Asp83, Lys99 и Glu102. Наиболее их частая конформация обозначается, как [A]. Альтернативная — [B].

Результаты и обсуждение

Для анализа качества расшифровки 4НJM были использованы следующие сервисы и базы данных:

1. PDB Validation Report [3]
2. MolProbity
3. EDS
4. WHAT_CHECK [4]

Ниже представлены основные моменты и оценки, основанные на выданной сервисами информации.

Общие показатели качества структуры

Разрешение модели 4НJM (1.55 Å) является хорошим. R-фактор данной модели равен 0.148 (меньше 0.25, что говорит о хорошем соответствии модели и эксперимента), Rfree — 0.179 (меньше 0.20, что также говорит в пользу данной модели). Разница между упомянутыми выше факторами меньше 0.1. Следовательно нет доводов подозревать авторов в переоптимизации данной модели.

При этом сервис WHAT_CHECK обнаружил, что на такие параметры, как длина (Z-score = 0.292) и углы связей (Z-score = 0.538), наложены слишком сильные ограничения при оптимизации структуры с разрешением в 1.55 Å. Программа также предупреждает об отсутствии нескольких атомов в PDB у Glu6 и Asn111.

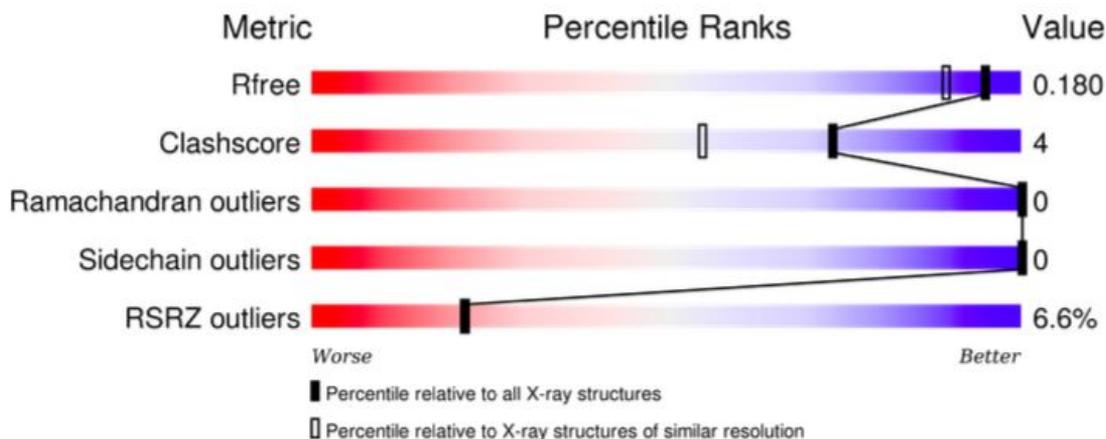


Рис. 2. Общая оценка качества модели 4НJM по сравнению с другими структурами примерно того же (1.55 Å) разрешения.

На **рис. 2** представлено сравнение параметров качества с 4НJM с другими моделями примерно того же разрешения. Данная информация была получена из PDB Validation Report. Видно, что структура сильно хуже других по такому параметру, как пространственный R-фактор. Следовательно большинство ошибок в

расшифровке и неправильные (маргинальные) остатки по идее могут быть определены по этому параметру. Также есть проблемы с пересечением Ван-дер-Ваальсовых радиусов (clashscore).

Помимо представленного выше сравнения в PDB Validation Report схематично изображены части структуры, где присутствуют геометрические затруднения (неправильные углы остова, ротамеры и т.д.). На **рис. 3** показано, что в модели отсутствует 5 аминокислотных остатков (окрашено серым цветом), 7 аминокислот плохо вписаны в электронную плотность (ЭП), а 10 остатков обладают геометрическими затруднениями. В статье [1] авторы указывают, что для первых пяти N-концевых остатков не наблюдалось внятной ЭП. Видимо это послужила причиной того, что их в итоге нет в расшифровке.

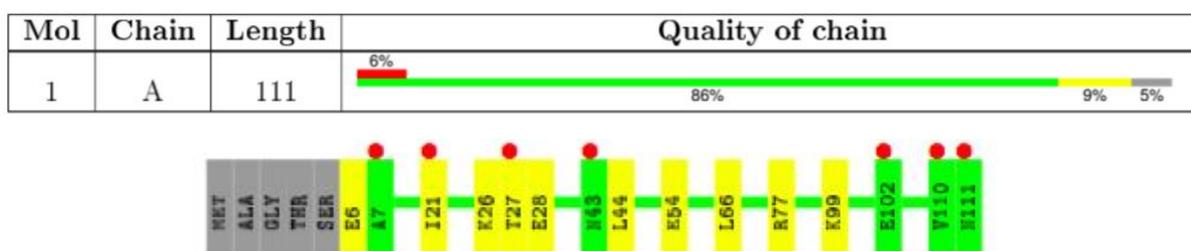


Рис. 3. Геометрические затруднения в 4Н1М из PDB Validation Report.

Сверху приведено процентное соотношение для всей модели, снизу изображены проблемные остатки: жёлтый цвет — несоответствие одному геометрическому критерию, серый — остатка нет в структуре; красная точка — остаток плохо вписывается в ЭП.

MolProbity

Таблица 1. Показатели качества расшифровки, полученные сервисом MolProbity.

Аспект	Показатель	Значение	Процент
Межатомные контакты	Clashscore	4.11	—
Геометрия	“Плохие” ротамеры	2 из 106	1.89%
	Отклонения по углам ψ и ϕ	0	0
	Отклонения по $C\beta$	0	0
	Отклонения по углам связей	0 из 949	0
	Отклонения по связям	0 из 1291	0
Дополнительно	Количество цис-пролинов	1 из 3	33.3%

С помощью сервиса MolProbity можно выяснить, какие именно остатки и атомы являются маргинальными по геометрическим параметрам. В **таблице 1** приведены значения для основных критериев. Видно, что в структуре присутствуют два остатка (Arg77[B], Lys99[B]) с отклонениями по углам χ (плохие ротамеры). При этом совершенно отсутствуют выбросы на карте Рамачандрана и отклонения по нескольким другим параметрам, что говорит в пользу хорошего качества расшифровки. Интересно, что в структуре присутствует один цис-пролин (Pro76), а плохие ротамеры принадлежат альтернативным конформациям аминокислот.

Карта Рамачандрана

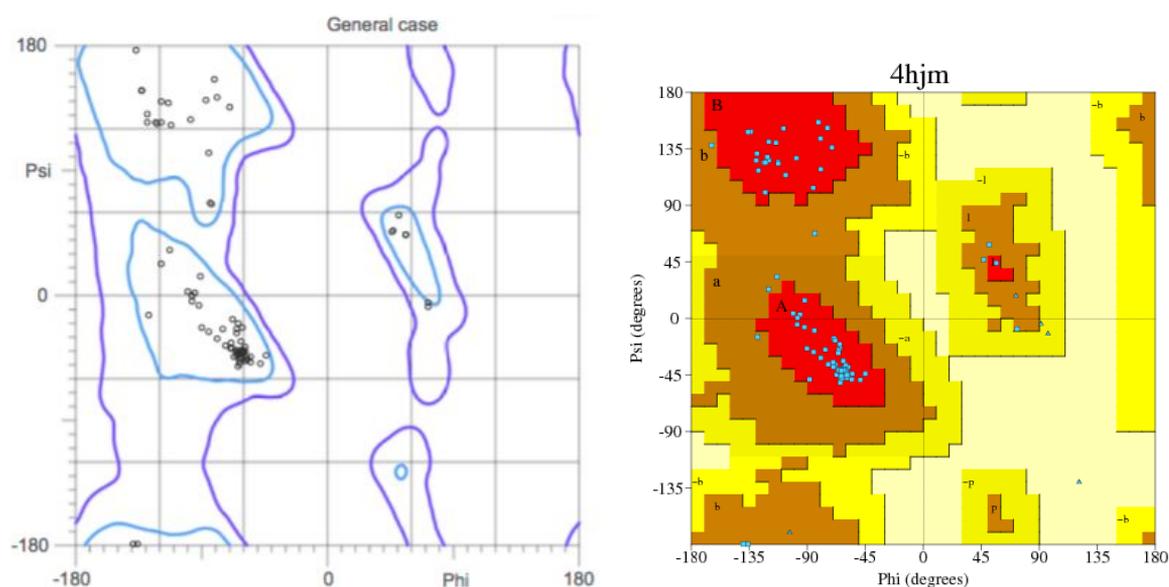


Рис. 4. Карты Рамачандрана, полученные с помощью сервиса MolProbity (слева) и PROCHECK (справа).

На **рис. 4** изображены карты Рамачандрана с отмеченными на них остатками аминокислот. Только 85 остатков находятся в наиболее благоприятной области, что соответствует 89.5% (по данным PROCHECK). У структур хорошего разрешения менее 2 Å процент должен быть больше 90%, в чём наша модель отстаёт.

Пространственный R-фактор (RSR и RSR-Z-score)

Пространственный R-фактор (RSR) позволяет оценить, насколько остаток вписывается в ЭП и рассчитывается по формуле (1) (сумма идёт по углам пространственной решётки в окружении группы атомов). Плохой RSR > 0.2.

$$RSR = \frac{\sum_{A \in L} |p_{\text{эксп}} - p_{\text{модель}}|}{\sum_{A \in L} p_{\text{эксп}}} \times 100\% \quad (1)$$

Помимо RSR также используется его относительная оценка — Z-score, вычисляемая по формуле (2). Она позволяет сравнить, как хорошо вписан данный остаток в данной модели по сравнению с выборкой структур того же разрешения. Z-score больше двух — повод проверить аминокислоту на маргинальность.

$$Z = \frac{RSR - \langle RSR_{1.55 \text{ \AA}} \rangle}{\sigma} \quad (2)$$

На **рис. 5** приведены графики, взятые с сайта EDS. Правый график — Significant regions. Здесь указаны участки с повышенным RSR-фактором и Z-score, которые наиболее вероятно содержат неточности в координатах. Наблюдается два выброса — Val110 (RSR = 0.179) и Asn111 (RSR = 0.222).

На других двух графиках изображены значения Z-score и RSR для каждого остатка. Помимо упомянутого выше Asn111 подозрительным RSR обладает Glu102 (RSR = 0.212) с Z-score = 1.58.

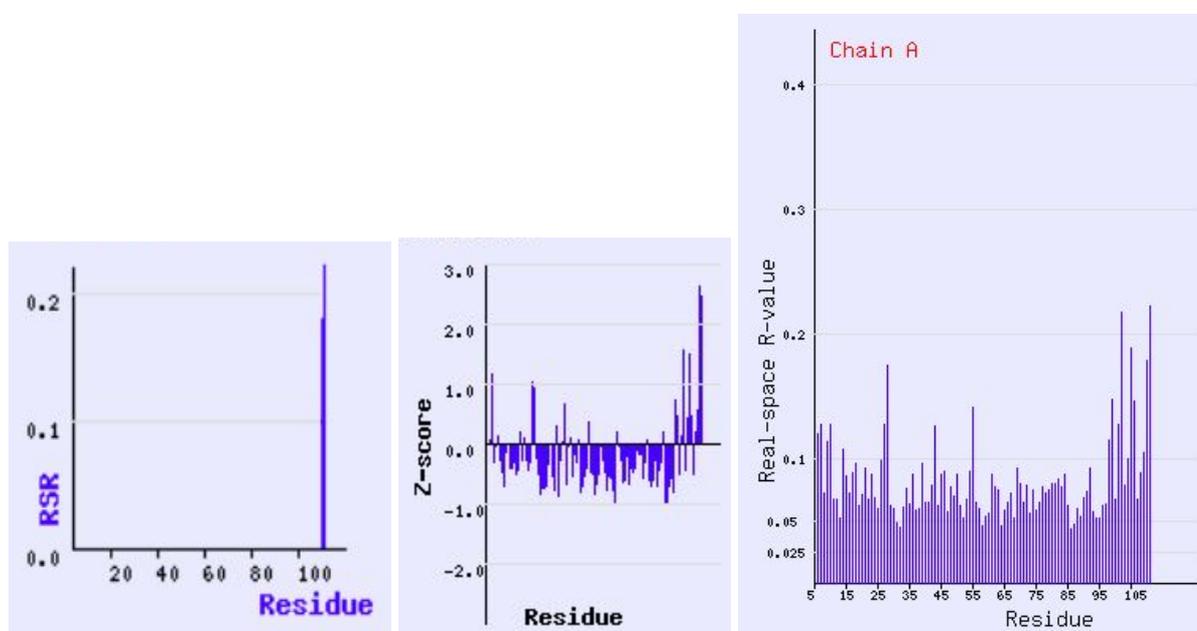


Рис. 5. Графики, характеризующие вписанность остатка в ЭП, взяты с сайта EDS. Правый график — выбросы по Z-score. Средний — Z-score для каждого остатка. Левый — значение RSR для каждой аминокислоты.

Комфортность окружения

Довольно важным показателем качества структуры является “комфортность” окружения атома. Этот параметр не зависит от процедуры оптимизации, поэтому довольно ценен. “Некомфортно” будет атомам (т.е. физически такое невозможно), например, если их Ван-дер-Ваальсовы радиусы пересекаются более чем на 0.4 Å. Такая ситуация называется clash и может быть обнаружена с помощью сервиса MolProbity. В **таблице 2** представлен перечень таких атомов.

Таблица 2. Перечень атомов с clash, выбранные сервисом MolProbability.

№	Первый атом	Второй атом	Межатомное расстояние (Å)	Перекрытие (clash) (Å)
1	99[A]:Lys:NZ	411:HOH:O	2.39	0.56
2	27:Thr:HG22	28:Glu:HG3	1.90	0.53
3	66:Leu:HD13	77[A]:Arg:HG3	1.92	0.52
4	21[A]:Ile:HD13	44:Leu:HB3	1.94	0.49
5	6:Glu:N	327:HOH:O	2.46	0.48
6	6:Glu:N	394:HOH:O	2.49	0.44
7	26[B]:Lys:HB2	54:Glu:HB2	2.01	0.43
8	26[B]:Lys:HE3	398:HOH:O	2.20	0.41

Помимо перекрытия Ван-дер-Ваальсовых радиусов данный показатель рассматривает наличие водородных связей у доноров/акцепторов протонов, скомпенсированы ли заряды на полярном аминокислотном остатке, находятся ли неполярные атомы в гидрофобном окружении. Сервис WHAT_CHECK для оценки использует показатель Z-score. Если он ниже -5, то вероятно что-то не так с окружением. В качестве подозрительных остатков сервис выделил Lys55 ($Z = -5.16$) и Glu97 ($Z = -5.05$).

Также был найден акцептор протонов, который не входит в состав водородных связей в белке: His47:ND1. И доноры: Lys9:N, Cys29:N, Lys82[A]:N, Asp90:N, Lys101:N и лиганд MPO:N1.

Температурный фактор

Температурный фактор (B-фактор) отражает подвижность каждого атома. Значения B-фактора менее 30 говорят о фиксированности позиции атома в то время, как B-фактор больше 60 говорит о высокой подвижности.

На **рис. 6** представлена структура 4HJM, а также значения среднего B-фактора для каждого остатка. Видно, что температурный фактор не превышает значение в 50. Среди всех оснований B-фактор больше 30 наблюдается у Glu102 (31.9), Glu6 (34.1), Lys106 (34.7) и Asn111 (45.7). Температурный фактор для большинства остатков находится в пределах 20, что говорит о хорошем качестве структуры.

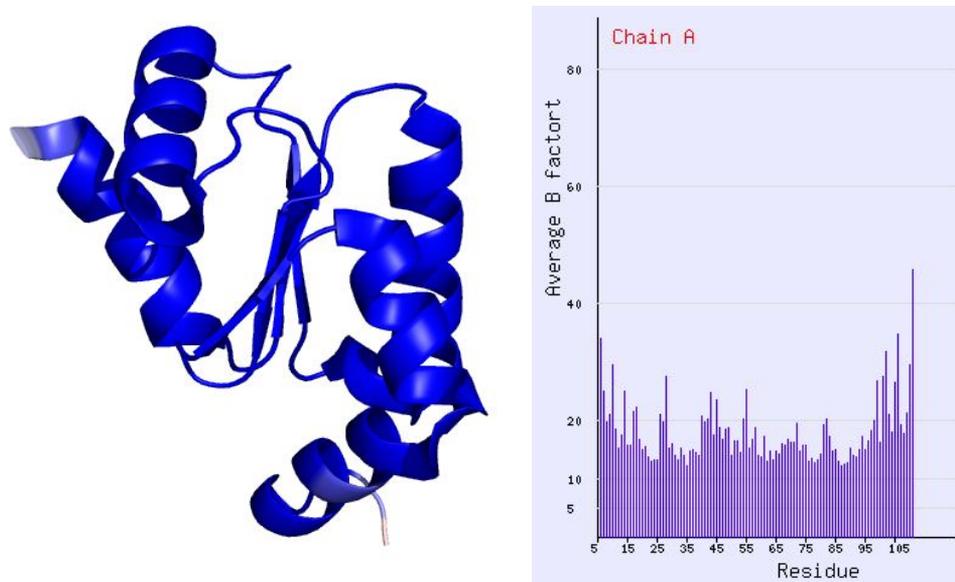


Рис. 6. Распределение температурного фактора по остаткам 4НЖМ. Слева представлена структура белка, окрашенная по значению В-фактора (менее интенсивный синий и красный цвета отражают остатки с высоким значением параметра). Справа — среднее значение В-фактора для каждой аминокислоты структуры.

Показатели качества расшифровки лигандов

В выдаче PDB Validation report также присутствует информация, характеризующая правильность геометрии лигандов. Параметры оцениваются с помощью Z-score. Если величина выходит больше 2, то это повод рассматривать структуру в качестве маргинальной. Средний Z-score для углов связей в МРО равен 5, а в таблице 3 приведены “странные” углы.

Таблица 3. Перечень отклоняющихся от нормы углов в структуре МРО.

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	A	201	MPO	C3-C2-C1	-3.53	106.56	112.37
2	A	201	MPO	O2-S1-O1	-2.06	105.96	113.48
2	A	201	MPO	C3-N1-C4	2.12	116.71	111.27
2	A	201	MPO	O2-S1-C1	3.36	109.78	106.91
2	A	201	MPO	O1-S1-C1	3.72	110.08	106.91

Для характеристики качества ЭП вокруг лиганда помимо RSR используется LLDF (Ligand Density Fit value — насколько хорошо ЭП описывает лиганд по сравнению с соседними аминокислотами). Если этот параметр больше 2, то в расшифровке лиганда могут быть ошибки. В данном случае подозрение в маргинальности вызывает MPD (LLDF = 5.13). К тому же у этого лиганда среднее значение В-фактора превышает 30 (35.4).

Анализ маргинальных остатков

Перечень маргинальных остатков

В **таблице 4** приведены примеры маргинальных остатков, отобранных по критериям, которые рассматривались выше. Остатки, выделенные жирным, будут рассмотрены более подробно.

Таблица 4. Перечень атомов с clash, выбранные сервисом MolProbity.

№	Название остатка	Критерий маргинальности
1	Asn111	Высокое значение RSR, RSR-Z-score и B-фактора; отсутствует часть атомов (CG, OD1, ND2)
2	Glu102	Высокое значение RSR и B-фактора
3	Glu6	Высокое значение B-фактора, clash с водой, отсутствует часть атомов (CG, CD, OE1, OE2)
4	Lys55	Неправильное окружение
5	Glu97	Неправильное окружение
6	Val110	Высокое значение RSR-Z-score
7	MPO	Неправильные торсионные углы, донор протонов не участвует в водородной связи
8	MPD	Высокое значение LLDF и B-фактора

Лиганд MPD

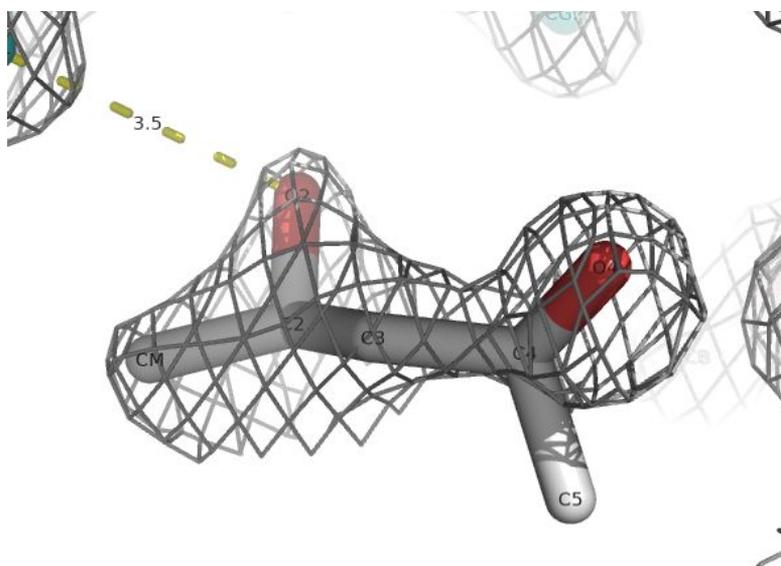


Рис. 7. MPD, вписанный в ЭП ($\sigma = 1$).

На **рис. 7** представлен лиганд MPD вместе с изображением ЭП в виде сетки. Мы видим, что ЭП не полностью описывает данную молекулу (атом C5, например, вообще не покрыт ЭП), хотя в целом о каких-либо ошибках в структуре MPD говорить сложно.

Остаток Lys55

Остаток Lys55 (**рис. 8**) обладает двумя альтернативными конформациями. Особо подвижный участок лизина уже плохо описывается ЭП. Подозрения на маргинальность была вызвано большим значением критерия неправильного окружения. Аминокислота расположена на краю неструктурированной петли, причём направлен наружу, а не внутрь белка. Таким образом проблем в её окружении как таковом я не нашла. Видно, что полярный конец лизина в своей основной конформации может взаимодействовать с водой, т.е. заряд скомпенсирован.

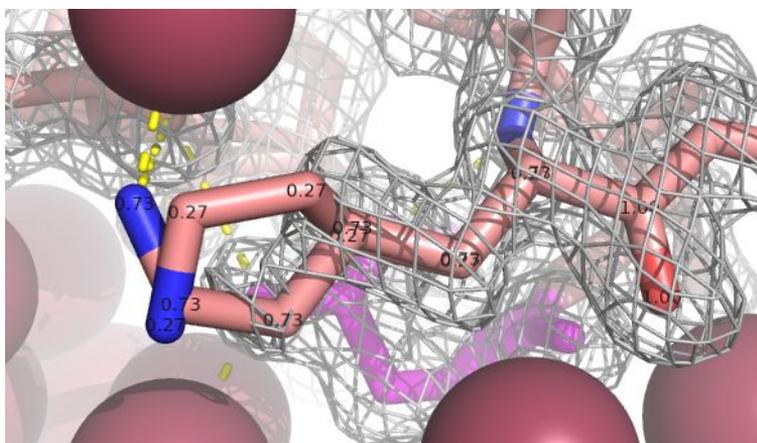


Рис. 8. Lys55, вписанный в ЭП ($\sigma = 1$).

На каждом атоме подписан процент его встречаемости в данном положении. Жёлтым обозначены водородные связи. Молекулы воды представлены в виде шаров.

Остаток Glu102

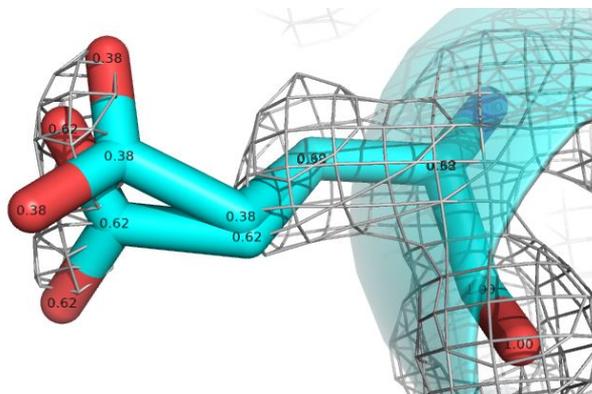


Рис. 9. Glu102, вписанный в ЭП ($\sigma = 1$).

На каждом атоме подписан процент его встречаемости в данном положении.

Глутамат, как и предыдущий остаток, обладает двумя альтернативными конформациями (**рис. 9**). Видимо из-за этого ЭП получилась размазанной в области атомов кислорода (на рис.9 красный цвет). В целом о наличие ошибок судить сложно, как и о том, что координаты данного остатка были определены верно.

Остаток Glu6

Glu6 (**рис. 10**) — один из остатков с наибольшим количеством проблем. Но что больше всего бросается в глаза — отсутствие боковой цепи. В исходной модели действительно есть только координаты остова остатка. Для боковой цепи ЭП отсутствует полностью, что могло быть причиной такой ошибки.

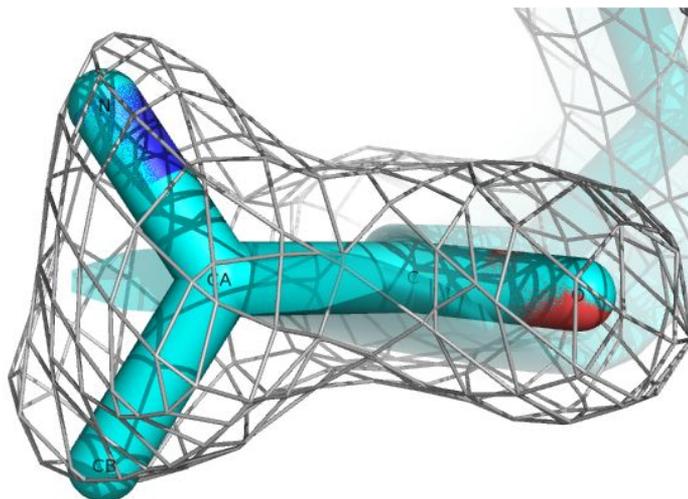


Рис. 10. Glu6, вписанный в ЭП ($\sigma = 1$).

Остаток Asn111

С Asn111 ситуация аналогичная предыдущему остатку: полностью отсутствует боковая цепь (**рис. 11**).

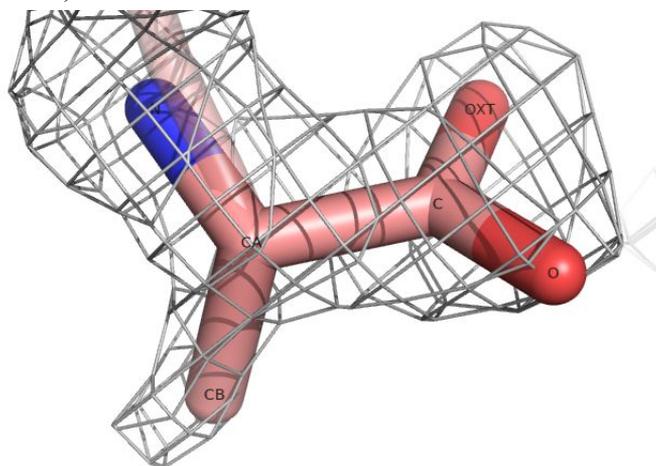


Рис. 11. Asn111, вписанный в ЭП ($\sigma = 1$).

PDB_redo

Сервис PDB_redo позволяет улучшить (уточнить) структуру с помощью автоматизированной процедуры. На выходе пользователь получает информацию о качестве оптимизированной структуры, её координаты, а также данные по исходной модели для сравнения. На **рис. 12** изображено графическое представление, насколько изменились параметры, описывающие правильную геометрию (анализ WHAT_CHECK) и соотношение модели с экспериментальными данными (R-фактор). Видно, что структура особо не изменилась. Rfree у оптимизированной структуры уменьшился (0.1688), а R-фактор немного вырос (0.1517).



Рис. 12. Графическое представление изменения качества модели после оптимизации с сайта PDBe.

Для сравнения исходной и исправленной структуры сервис PDB_redo предлагает использовать изменение параметра RSCC (Real-Space Correlation Coefficient). Этот параметр показывает насколько хорошо остаток вписывается в ЭП.



Рис. 13. График изменения значения RSCC для каждого основания после полной оптимизации структуры.

Красным цветом обозначены участки, где качество значительно ухудшилось, зелёным — значимо возросло.

Из **рис. 13** видно, что сильно качество для структуры не улучшилось. Некоторые молекулы воды стали лучше описываться ЭП, но в целом кардинальных изменений не произошло.

Аналогичную ситуацию можно увидеть и на пространственном выравнивании исходной и полностью оптимизированной структуры (**рис. 14**). Заметно лишь, что слегка поменялся вид и форма α -спиралей.

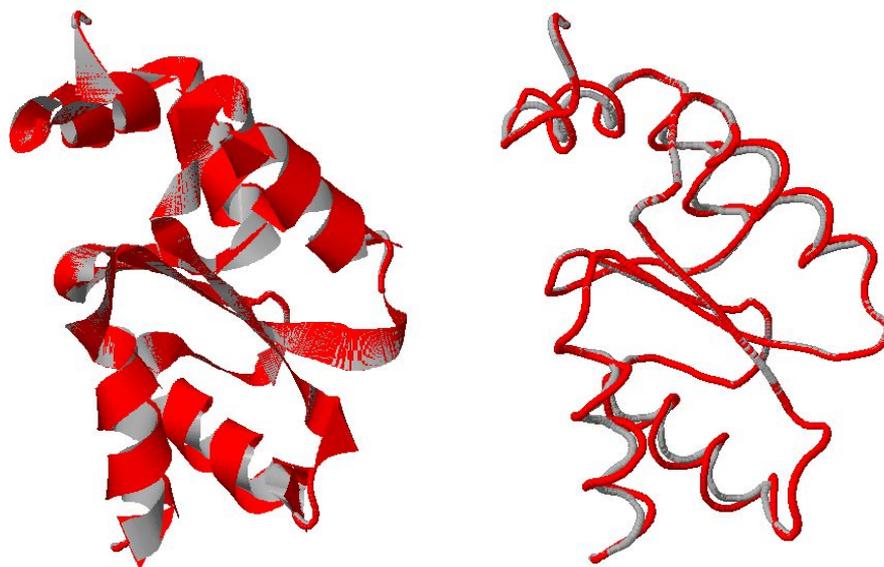


Рис. 14. Пространственное выравнивание исходной и полностью оптимизированной структуры с помощью FATCAT.

Серым цветом обозначена исходная структура, красным — оптимизированная.

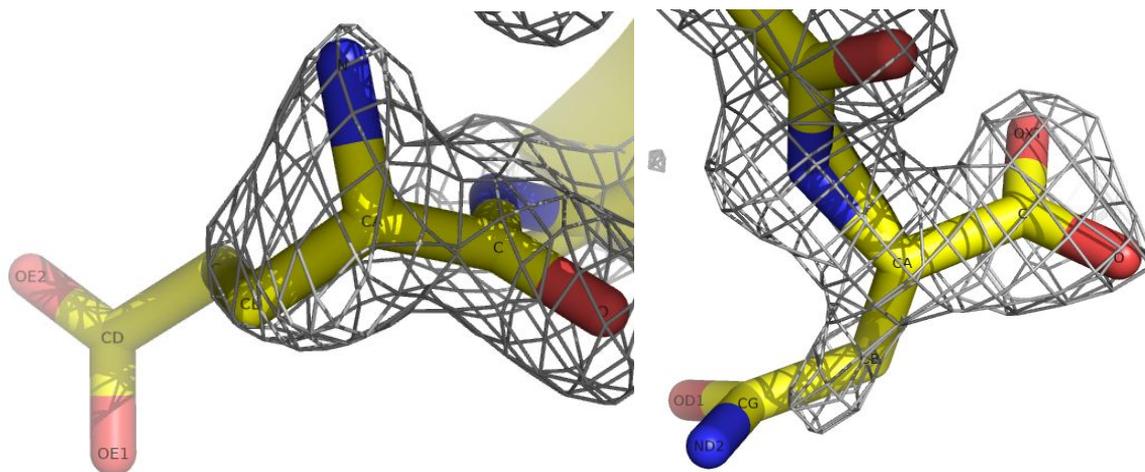


Рис. 15. Исправленные структуры Glu6 (слева) и Asn11 (справа).

Главным отличием оптимизированной структуры является наличие пропущенных атом в остатках Glu6 и Asn11 (**рис. 15**).

Выводы

1. Основная проблема модели 4HJM — плохое соответствие экспериментальным данным.
2. В структуре отсутствуют первые 5 аминокислотных остатков, а также атомы боковых цепей Glu6 и Asn111.
3. При этом по геометрическим параметрам структура вполне качественная и содержит крайне мало маргинальных остатков.
4. Возможно, несмотря на малую разницу Rfree и R, имела место переоптимизация в виду сильного ограничения на длины и углы связей, что в результате дало хорошие показатели в геометрии модели.
5. Полная оптимизация модели с помощью PDB_redo особо не улучшила структуру, но добавила потерянные атомы боковых цепей.

Список использованной литературы и сервисов

Статьи:

- [1] Yogavel, M., Tripathi, T., Gupta, A., Banday, M. M., Rahlfs, S., Becker, K., Belrhali, H. & Sharma, A. “Atomic resolution crystal structure of glutaredoxin 1 from *Plasmodium falciparum* and comparison with other glutaredoxins” (2014). *Acta Cryst. D70*, 91-100.
- [2] Djuika, C. F., Fiedler, S., Schnölzer, M., Sanchez, C., Lanzer, M. & Deponte, M. “Plasmodium falciparum antioxidant protein as a model enzyme for a special class of glutaredoxin/glutathione-dependent peroxiredoxins” (2013). *Biochim. Biophys. Acta*, 1830, 4073–4090.

Сервисы:

- [3] PDB Validation report:
http://ftp.wwpdb.org/pub/pdb/validation_reports/hj/4hjm/4hjm_full_validation.pdf
- [4] WHAT_CHECK:
http://www.cmbi.ru.nl/pdb_redo/hj/4hjm/wo/pdbout.txt
- [5] EDS:
<https://eds.bmc.uu.se/cgi-bin/eds/uusfs?pdbCode=4hjm>
- [6] MolProbity:
Chen et al. (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallographica D66*:12-21.
- [7] PROCHECK
- [8] FATCAT:
<http://fatcat.burnham.org>
- [9] PDB_redo:
http://www.cmbi.ru.nl/pdb_redo/hj/4hjm/index.html
- [10] PDBe:
<http://www.ebi.ac.uk/pdbe/entry/pdb/4hjm>

Программы:

- [11] PyMOL