

Отчёт по качеству РСА расшифровки
структуры
полинуклеотидкиназы (PDB ID: 4GP7)
организма *Clostridium thermocellum*

Валяева Анна
4 курс

Москва, 2016

ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ.....	3
РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ.....	4
ОБЩАЯ ИНФОРМАЦИЯ О МОДЕЛИ.....	4
ИНДИКАТОРЫ КАЧЕСТВА МОДЕЛИ В ЦЕЛОМ.....	5
ПРИМЕРЫ МАРГИНАЛЬНЫХ ОСТАТКОВ	10
СРАВНЕНИЕ МОДЕЛИ ИЗ PDB С МОДЕЛЬЮ ИЗ PDB_REDO.....	13
ЗАКЛЮЧЕНИЕ.....	16
ЛИТЕРАТУРА	16

Аннотация

Отчет написан по результатам анализа качества расшифровки структуры 4GP7 - гомодимера киназного домена полинуклеотидкиназы/фосфатазы Pnkp организма *Clostridium thermocellum*. В отчете содержится общая информация о структуре, приведены основные показатели качества построенной модели, полученные с помощью ряда сервисов по оценке структур, и рассмотрены маргинальные по некоторым показателям остатки.

Введение

Бактериальная система репарации РНК Pnkp-Hen1 необходима для сшивания разрывов в РНК, возникающих в результате процессинга или под воздействием различных стрессов. При этом происходит метилирование по сайту разрыва, вследствие чего этот сайт становится менее подвержен повторному разрыву.

Компонентами вышеназванной системы репарации являются два белка, которые физически взаимодействуют друг с другом во время реакции, образуя гетеродимерный комплекс: Hen1 и Pnkp. Первый осуществляет метилирование, второй фосфорилирует 5'-конец РНК и зашивает разрыв (рис. 1). Pnkp - полинуклеотидкиназа/фосфатаза – состоит из трех функциональных доменов: N-концевого киназного домена, фосфатазного домена и С-концевого лигазного домена.[1]

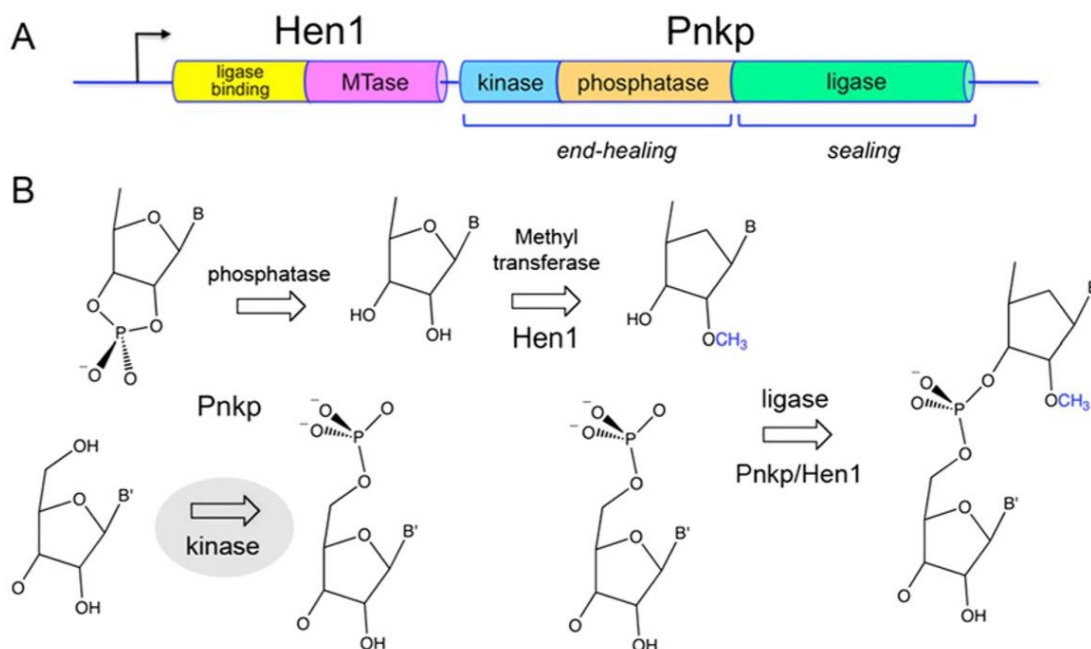


Рис. 1. Система репарации РНК Pnkp-Hen1.[1]

К 2012 году были уже известны структуры метилтрансферазного домена белка Hen1, лигазного домена Pnkp и гетеродимера, состоящего из С-концевого лигазного домена Pnkp и N-концевого домена Hen1. Знания о структуре этих доменов и межбелковых взаимодействиях позволили глубже понять механизм

работы данной системы репарации РНК. Расшифровка структуры киназного домена еще больше подтвердила уникальность системы репарации Pnkr-Hen1 и помогла спланировать изучение механизма катализируемой этим доменом киназной реакции путем направленного мутагенеза.

Результаты и обсуждение

Общая информация о модели

Структура киназного домена полинуклеотидкиназы Pnkr была расшифрована в 2012 году. Домен был закристаллизован в виде гомодимера в комплексе с АТФ·Mg²⁺, связанным в активном центре, цитратом и ионами Na⁺. Авторы этой работы - Wang LK, Das U, Smith P, Shuman S., статья называется «Structure and mechanism of the polynucleotide kinase component of the bacterial Pnkr-Hen1 RNA repair system». Помимо определения структуры авторы статьи еще описали активный центр полинуклеотидкиназы и предложили механизм реакции.

Фазовая проблема была решена методом аномального рассеяния SeMet-SAD. Поскольку нативная структура полинуклеотидкиназы содержит только один остаток метионина (на N-конце), белок был модифицирован и были искусственно введены 3 дополнительных метионина вместо Val44, Ile93, Leu137. Также были попытки решить фазовую проблему методом молекулярного замещения, используя гомологичные структуры киназных доменов Pnkr бактериофага T4 и млекопитающих, однако они ни к чему не привели.[1]

В РСА эксперименте были измерены 25245 структурных факторов, из них 23021 были использованы для построения и оптимизации модели, 1977 – для ее оценки, остальные оказались недостаточно хорошо измерены. Разрешение структуры считается хорошим и составляет 2.00 Å, полнота данных - 98.9 %. Диапазон разрешений структурных факторов равен 44.54 - 2.00 Å.[2]

Параметры кристаллографической ячейки, содержащиеся в поле CRYST1 в PDB файле, представлены в таблице 1.

Таблица 1. Параметры кристаллографической ячейки.[3]

Длины направляющих векторов кристалла	a	45.306 Å
	b	66.781 Å
	c	119.546 Å
	α	90°

Углы между направляющими векторами кристалла	β	90°
	γ	90°
Кристаллографическая группа		P 21 21 21
Число молекул в ячейке		8

Для пространственной группы P 21 21 21 возможны следующие операторы симметрии:

X,Y,Z; -X+1/2,-Y,Z+1/2; -X,Y+1/2,-Z+1/2; X+1/2,-Y+1/2,-Z.

Индикаторы качества модели в целом

Значения основных индикаторов качества модели, то есть насколько хорошо модель соответствует экспериментальным данным, приведены на сайте PDB[3] в виде графика (рис. 2)[4]. На графике сравниваются значения метрик для конкретной структуры (4GP7) со значениями метрик всех доступных в PDB структур (черный прямоугольник) или структур того же разрешения (незакрашенный прямоугольник). Абсолютные значения характеристик качества модели приведены справа в столбце Value. Синяя область на графике соответствует хорошим значениям, красная – плохим.

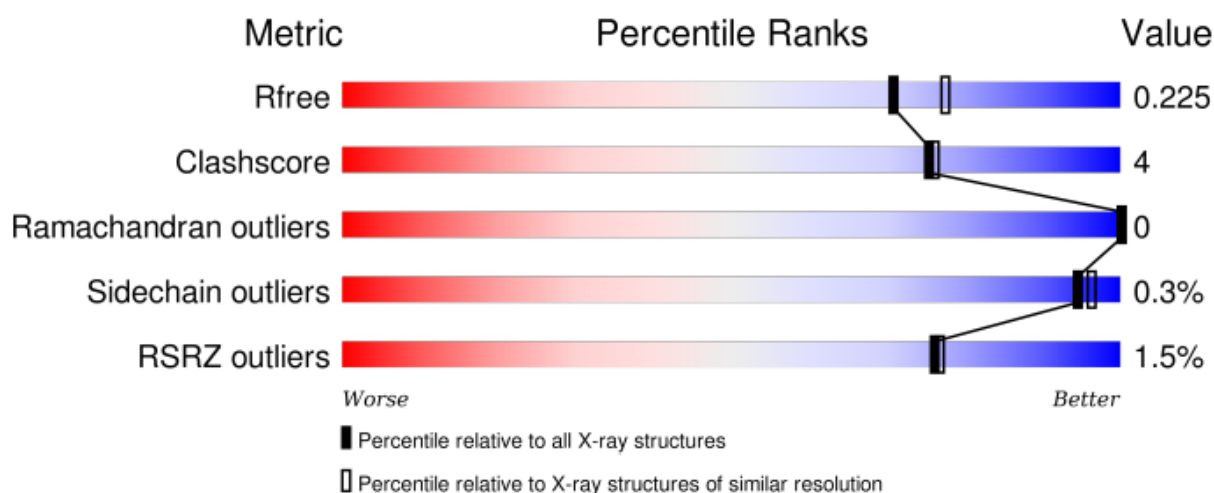


Рис. 2. График, иллюстрирующий сравнение параметров структуры 4GP7 со структурами, доступными в PDB.

Поскольку все значения характеристик качества модели лежат в синей области, можно сделать вывод, что модель достаточно хорошего качества. Далее рассмотрим все характеристики по-отдельности.

1. R-фактор (R-factor)

R-фактор говорит о соответствии модели экспериментальным данным, то есть на сколько отличаются рассчитанные по модели структурные факторы (F_{calc}) от полученных в ходе эксперимента (F_{obs}):

$$R = \frac{\sum ||F_{obs}| - |F_{calc}||}{\sum |F_{obs}|}$$

Соответственно чем меньше значение R-фактора, тем лучше построенная модель согласуется с экспериментальными данными. Для структуры полинуклеотидкиназы 4GP7 значение R-фактора равно 0.175. Однако поскольку при оптимизации модели происходит минимизация именно этого параметра (в качестве слагаемого в составном R-факторе) и поэтому есть опасность переобучения (оверфиттинга), для оценки качества модели следует иметь контрольную выборку рефлексов, по которым будет рассчитан свободный R-фактор.

2. Свободный R-фактор (R-free)

Свободный R-фактор рассчитывается так же как и обычный R-фактор, но по отдельной выборке контрольных рефлексов, которые не использовались для оптимизации модели. Этот фактор рассчитывается по конечной модели. Для 4GP7 R-free равен 0.225. Такое значение нельзя назвать очень хорошим (<0.2), но оно не плохое. Разница между R-фактором и свободным R-фактором составляет 0.05 и говорит о том, что, возможно, произошла переоптимизация модели.

3. Clashscore

Параметр Clashscore показывает число слишком близко расположенных атомов (неблагоприятно перекрывающихся более чем на 4 Å) на 1000 атомов структуры. Для структуры 4GP7 Clashscore равен 4; всего таких перекрываний в ассиметрической единице 22, из них 1 связано с преобразованиями симметрии. Как показывает график на рис. 2, это значение является достаточно посредственным.

4. Маргинальные остатки по карте Рамачандрана (Ramachandran outliers)

Доля маргинальных остатков, у которых значения торсионных углов ϕ и ψ остова лежат в неблагоприятных зонах на карте Рамачандрана, является хорошим показателем качества модели, который не зависит от оптимизации. На рисунке 3 приведены карты Рамачандрана, построенные для модели 4GP7, для всех аминокислотных остатков (3A) и для отдельных аминокислот, имеющих отличные от общей, «усредненной» карты (3B-F). Как видно по этим картам все остатки (100%) лежат в разрешенных областях, а 99.1% остатков – в наиболее предпочтительных. Выбросов нет. Это очень хорошие показатели.

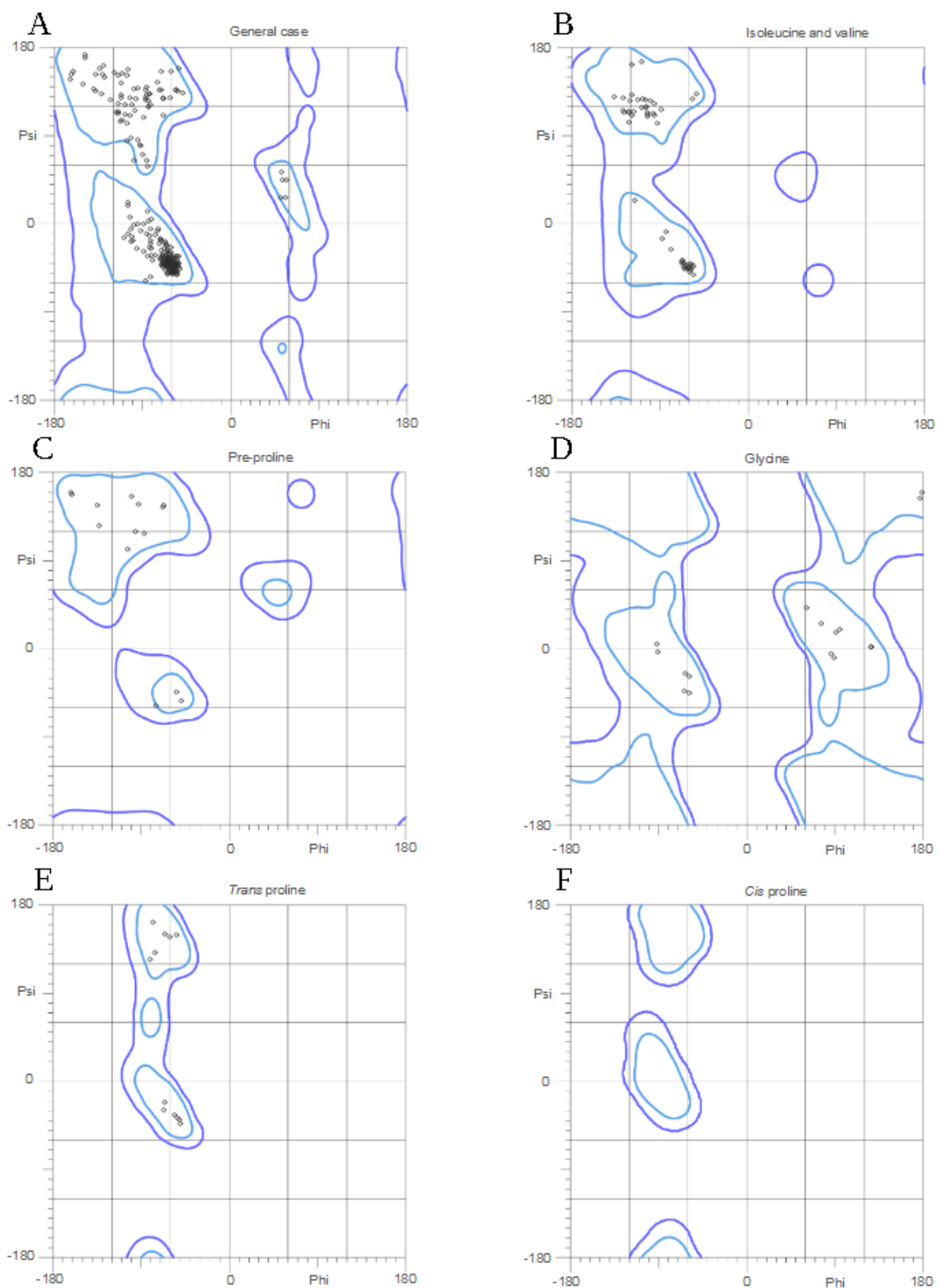


Рис. 3. Карты Рамачандрана для всех остатков (А), для изолейцинов и валинов (В), для глицинов (D) и для транс- и цис-пролинов (Е и F, соответственно).[2]

5. Sidechain outliers

Аналогично торсионным углам ϕ и ψ торсионные углы χ в боковых цепях могут принимать только некоторые области значений. Также подсчитывается процент «хороших» и «плохих» ротамеров. Для структуры 4GP7 процент «хороших» ротамеров равен 93.35% (в идеале >98%), число и процент «плохих» - 7 (2.22%, в идеале <0.3%). Список последних приведен в таблице 2.

Таблица 2. Список «плохих» ротамеров в структуре 4GP7.

Цепь	Остаток и его номер	Значения углов χ
A	Mse1	253.5, 197.2, 27.7
A	Asp46	31, 18.7
A	Glu162	245.4, 297.5, 307.8
B	Lys26	168.1, 189.6, 275, 70.4
B	Val53	145.4
B	Pe91	278.5, 115.1
B	Glu159	22.6, 272.3, 285

6. Пространственный R-фактор (RSR)

Пространственный R-фактор показывает, насколько согласуется «экспериментальная» ($\rho_{\text{эксп}}$ построенная по полученным из эксперимента структурным факторам и рассчитанным фазам) и модельная электронная плотность ($\rho_{\text{модель}}$), и рассчитывается по формуле:

$$RSR = \frac{\sum_{A \in L} |\rho_{\text{эксп}} - \rho_{\text{модель}}|}{\sum_{A \in L} \rho_{\text{эксп}}} [\cdot 100\%]$$

Большие значения RSR указывают на то, что данный остаток в модели плохо вписан в электронную плотность. Это значит, что модель на данном участке плохо согласуется с экспериментальными данными. На рисунке 4 приведены графики распределения значений RSR по цепям А и В.

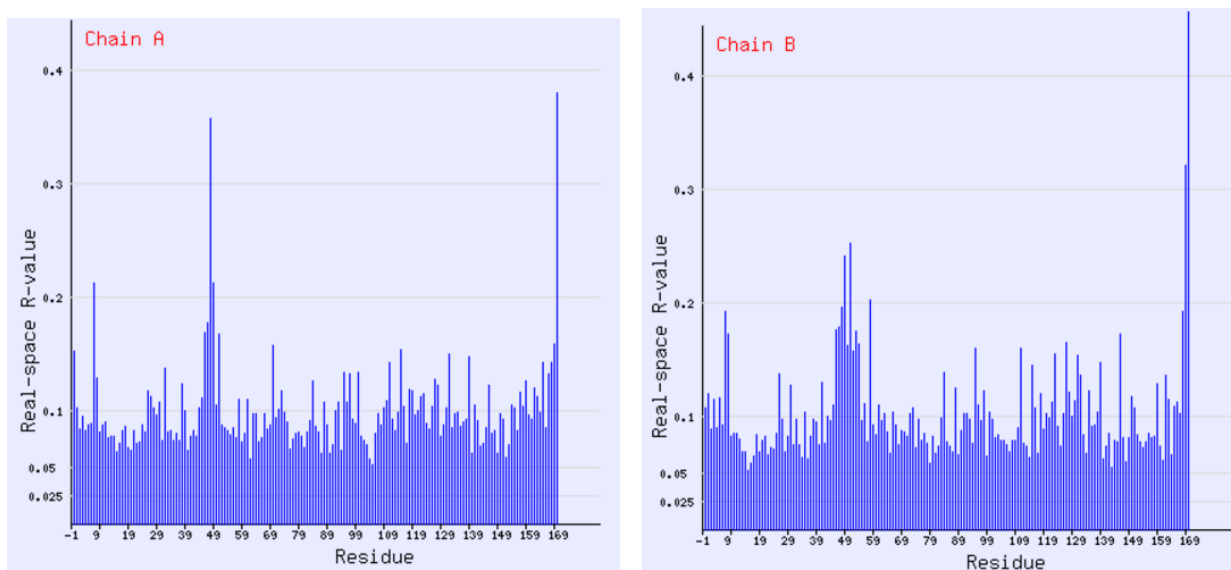


Рис. 4. Значения RSR для остатков в цепи А (слева) и цепи В (справа) структуры 4GP7.[2]

Как видно из графиков, в модели есть несколько остатков с «плохими» значениями RSR (>0.2), но большинство значений лежит в зоне «хороших» значений ($RSR < 0.1$). Список остатков с $RSR > 0.2$ приведен в таблице 3. Из перечисленных 9 остатков 6 можно назвать аутлайерами – они значительно хуже вписаны в электронную плотность, по сравнению с остатками того же типа в структурах с тем же разрешением, что и 4GP7 (в таблице выделены жирным).

Таблица 3. Список аминокислотных остатков с «плохими» значениями RSR (>0.2).

Цепь	Остаток и его номер	Значение RSR
A	Glu7	0.212
A	Glu48	0.357
A	Asn49	0.213
A	Pro170	0.380
B	Asn49	0.241
B	Gln51	0.252
B	Phe58	0.203
B	Gln169	0.351
B	Pro170	0.519

Оценить, насколько значения RSR для каждого из остатков отличаются относительно среднего по цепи, можно с помощью величины RSRZ, которая является Z-score для RSR. График значений RSRZ приведен на рисунке 5.

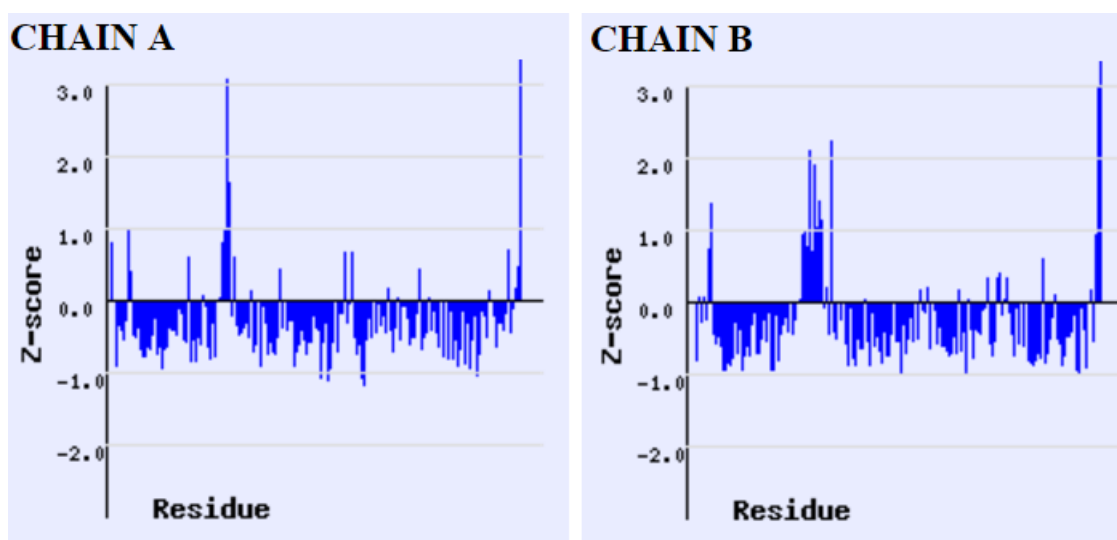


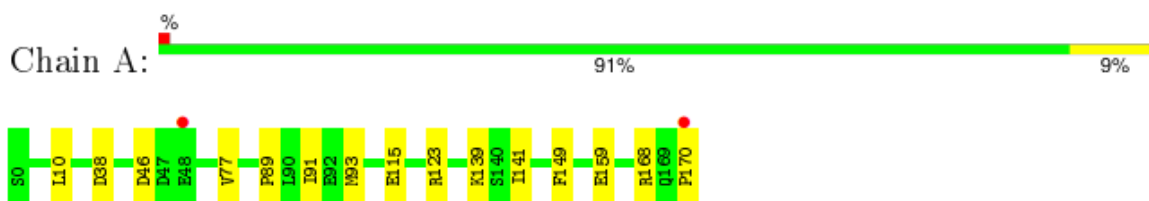
Рис. 5. Значения RSRZ для остатков в цепи А (слева) и цепи В (справа) структуры 4GP7.[2]

Большие значения RSRZ (>2) у остатка цепи свидетельствуют о том, что он плохо вписан в электронную плотность по сравнению с другими структурами с тем же разрешением, поэтому его можно рассматривать как маргинальный. Такими маргинальными остатками в структуре 4GP7 являются остатки, выделенные жирным в таблице 3, - всего 6 штук. Остальные аминокислоты, судя по значениям RSR и RSRZ, хорошо вписаны в электронную плотность, что говорит в пользу относительно высокого качества модели.

Примеры маргинальных остатков

Примеры маргинальных по значениям углов χ в радикалах и RSR аминокислот структуры 4GP7 приведены в таблицах 2 и 3, соответственно. Также о качестве отдельных остатков можно узнать из отчета по валидации структуры на сайте PDB (рис. 6).[3] Информация о них содержится в таблице 4.

- Molecule 1: Metallophosphoesterase



- Molecule 1: Metallophosphoesterase

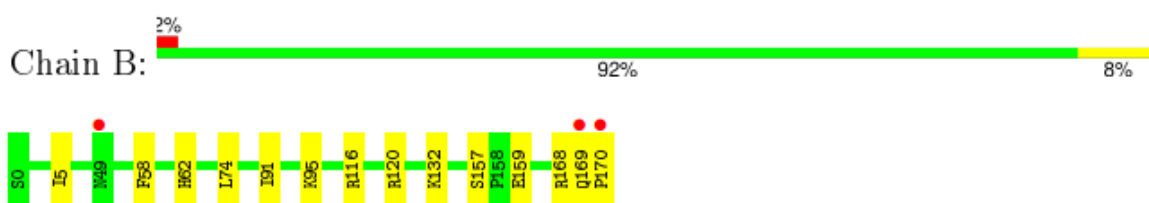


Рис. 6. Качество отдельных остатков в структуре 4GP7.[4] Маргиналы по одному геометрическому параметру жёлтые, по двум оранжевые, по трем и более красные. Красная точка над остатком означает то, что остаток плохо вписан в электронную плотность.

Таблица 4. Список маргинальных аминокислотных остатков структуры 4GP7.

Цепь	Остаток и его номер	Почему считается маргиналом
A	Asp46	«Плохой» ротамер, высокий температурный фактор (94.09) – положение остатка определено ненадежно
A	Glu48	Высокие значения RSR и RSRZ – остаток плохо вписан в электронную плотность, высокий температурный фактор (99.84) – положение остатка определено ненадежно
A	Glu162	«Плохой» ротамер
A	Pro170	Высокие значения RSR и RSRZ – остаток плохо вписан в электронную плотность
B	Asn49	Высокие значения RSR и RSRZ – остаток плохо вписан в электронную плотность, высокий температурный фактор (99.75) – положение остатка определено ненадежно
B	Phe58	Высокие значения RSR и RSRZ – остаток плохо вписан в электронную плотность, высокий температурный фактор (102.1) – положение остатка определено ненадежно
B	Glu159	«Плохой» ротамер, высокий температурный фактор (86.45) – положение остатка определено ненадежно
B	Gln169	Высокие значения RSR и RSRZ – остаток плохо вписан в электронную плотность, высокий температурный фактор (114.47) – положение остатка определено ненадежно
B	Pro170	Высокие значения RSR и RSRZ – остаток плохо вписан в электронную плотность

Далее детально рассмотрим несколько маргиналов.

1. Pro170 в цепи В. Это С-концевая аминокислота и для нее плохо смоделирована электронная плотность, поэтому при создании модели, вероятно, возникли трудности с тем, как бы расположить этот пролин (рис. 7).

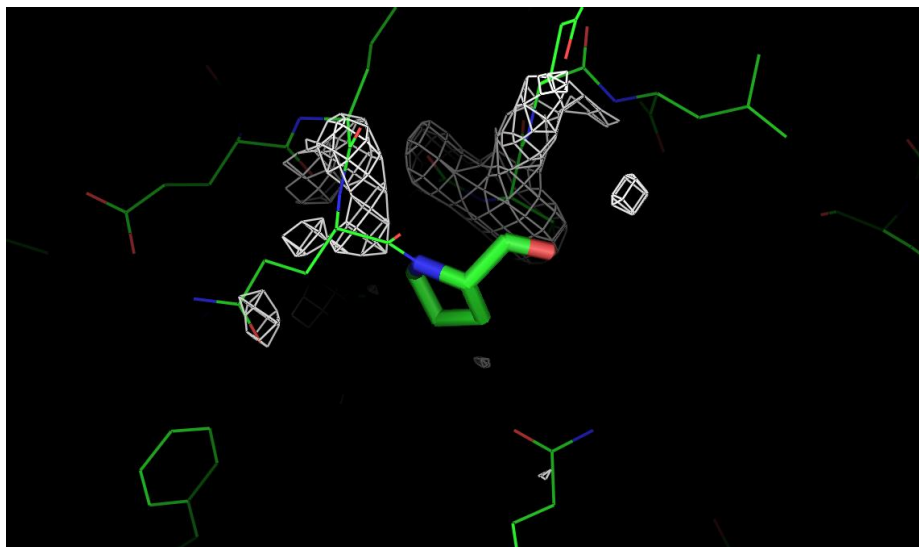


Рис. 7. Pro170 в цепи В, плохо вписанный в свою электронную плотность, которой нет.

2. Phe58 в цепи В. (рис. 8) Этот остаток также плохо вписан в электронную плотность, этим объясняется высокие значения RSR и RSRZ.

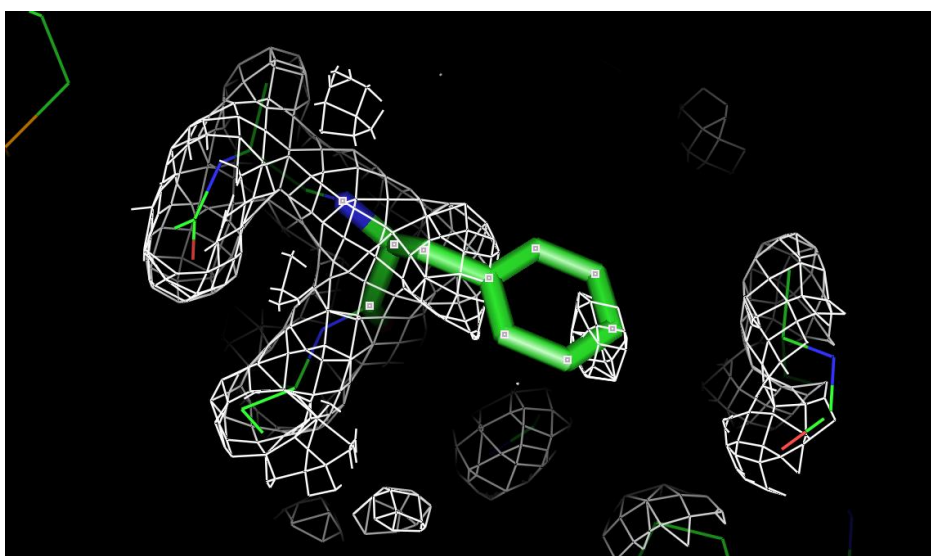


Рис. 8. Phe58 в цепи В, плохо вписанный в свою электронную плотность.

3. Asn49 в цепи В. Электронная плотность, соответствующая области, в которой располагается этот остаток, плохо восстановлена – на рисунке 9 видно, что соответствующее аспарагину «облако» электронной плотности как бы пропущено.

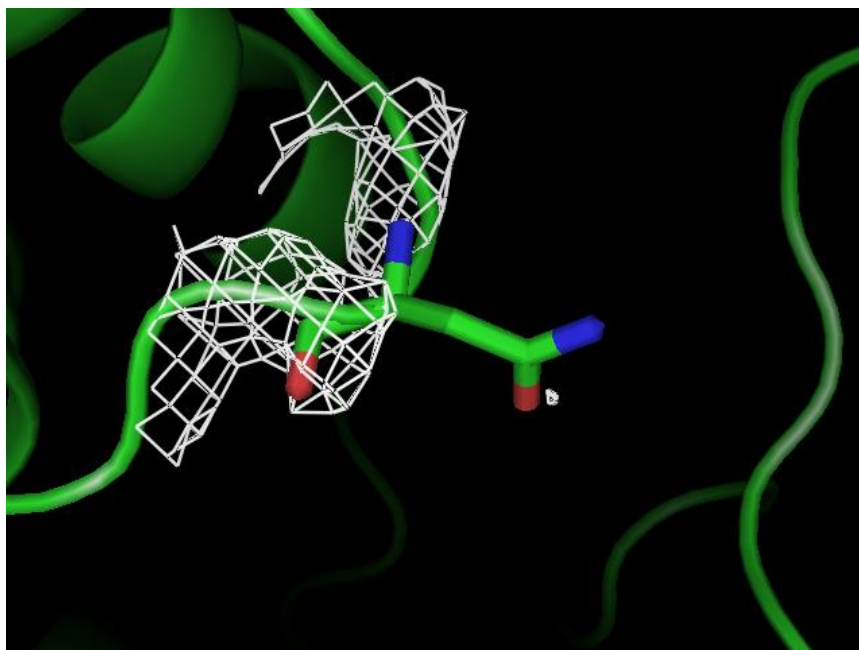


Рис. 9. Asn49 в цепи В, соответствующая ему электронная плотность не восстановлена.

4. Asp46 в цепи А. Этот остаток является маргиналом, поскольку его конформация не является оптимальной. Однако этот аспарат не образует никаких функциональных связей – он располагается в петле и обращен в растворитель, поэтому можно предположить, что необычная конформация данного остатка возникла вследствие ошибки расшифровки. Как видно по рисунку 10, рассматриваемому остатку соответствует только маленькое «облачко» электронной плотности, не вмещающее в себя весь радикал, вероятно, из-за этого торсионные углы χ для данной аминокислоты были определены неправильно.

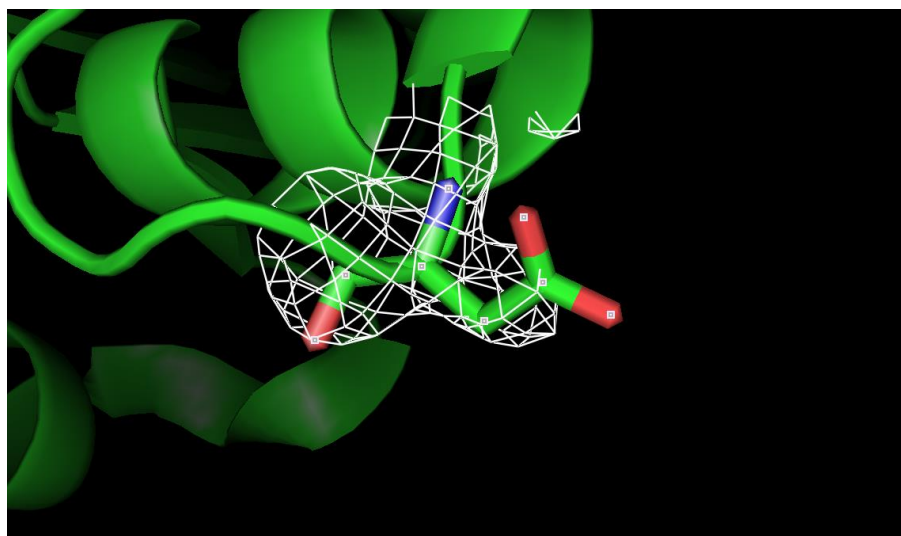


Рис. 10. Asp46 в цепи А.

5. Перекрытие остатков Asp38 и Arg123 в цепи А. Эти два остатка образуют солевой мостик между двумя удаленными по последовательности, но сближенными в пространстве участками одной цепи белка. Однако связь между атомом Н –NH₂ группы аргинина и O⁻ аспартата оказывается очень короткой (2.18 Å) в модели. Как видно по рисунку 11, проблема возникает

из-за того, что для аспартата предлагается два положения радикала, при одном из них возникает аномальное сближение с аргинином. Эти две конформации аспартата согласуются с картой электронной плотности, так что, видимо, данный остаток принимает разные положения в кристалле. Но если присмотреться на совмещение структуры и электронной плотности, то можно заметить, что один из атомов О карбоксильной группы аспартата в «сближенной» конформации обделен электронной плотностью, и если немного изменить угол χ_2 , то он войдет в ее «облако» (участок сетки под обделенным кислородом) и перекрытие скорее всего исчезнет

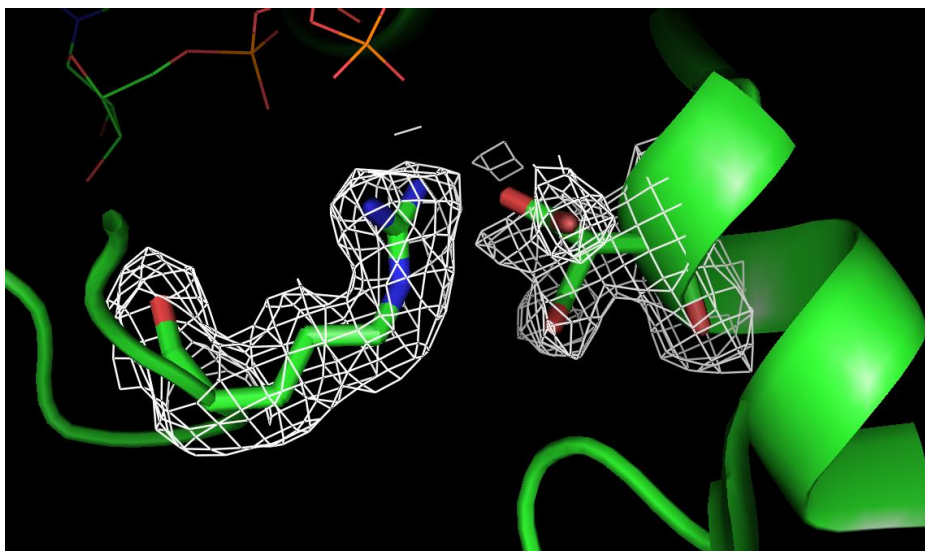


Рис. 11. Asp38 (справа) и Arg123(слева) в цепи А.

Таким образом, большинство проблем с маргинальными остатками возникает из-за низкого качества восстановления электронной плотности и ошибок, появляющихся при вписывании остатков в нее.

Сравнение модели из PDB с моделью из PDB_redo

Сервис PDB_redo служит для оптимизации и валидации моделей кристаллографических структур.[5] Сервис использует различные параметры уточнения структур, подбирает оптимальные конформации ротамеров, перестраивает сети водородных связей и так далее, а дальше запускает валидацию WhatCheck по основным параметрам.

Для структуры 4GP7 на сервере PDB_redo уже хранится оптимизированная структура.[6] На рисунке 12 приведены значения R-факторов и свободных R-факторов исходной и оптимизированных моделей. Видно, что после оптимизации их значения снизились и немного сократился разрыв между R-фактором и R-free, что свидетельствует об успехе оптимизации структуры.

R-values etc.

	From PDB header	Calculated from data	After re-refinement	After re-refinement and rebuilding
R	0.1750	0.1895	0.1704	0.1681
R-free	0.2300	0.2360	0.2035	0.2036
σ R-free		0.0053	0.0046	0.0046
Z(R-free)		0.36	2.26	1.61

Рис. 12. Значения R-факторов для исходной и оптимизированных моделей структуры 4GP7.[6]

В результате оптимизации снизилось число аминокислотных остатков, попадающих в сравнительно неблагоприятные области карты Рамачандрана, была улучшена конформация остова белка, оптимизированы длины и углы связей, снижено количество перекрываний атомов. Значительно лучше стали вписываться в электронную плотность 13 аминокислотных остатков, хуже – 2.[6], [7]

На рисунке 13 изображено совмещение исходной (желтая) и оптимизированной (красная) моделей. Они практически совпадают: есть небольшие расхождения $C\alpha$ -атомов на концах цепей и в неструктурированных петлях.

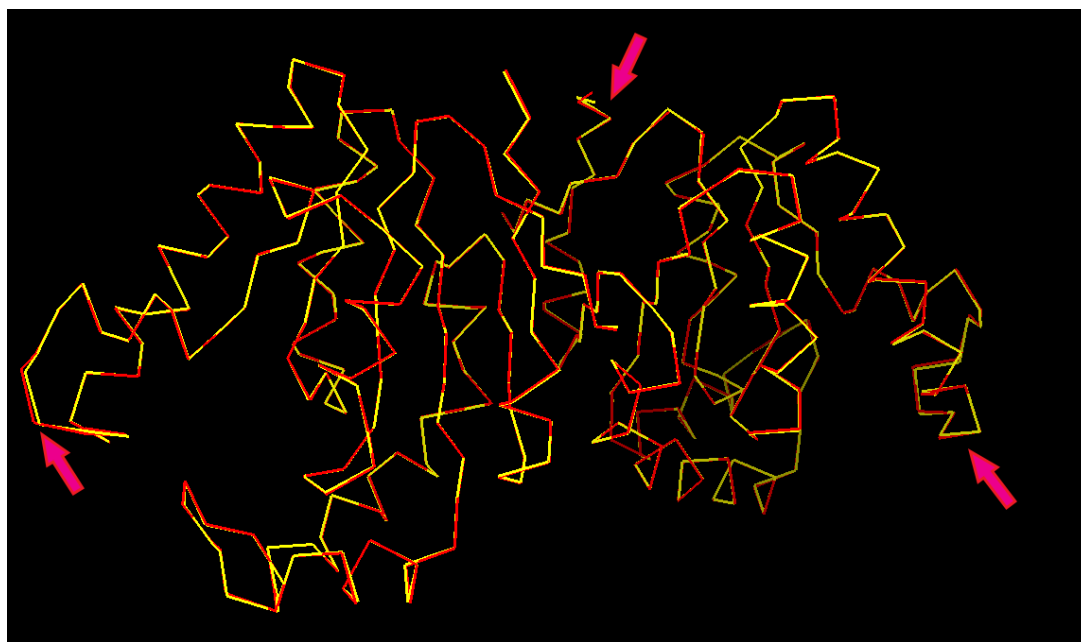


Рис. 13. Совмещение исходной (желтая) и оптимизированной модели (красная) структуры 4GP7. Стрелки указывают на расхождения между моделями.

В оптимизированной модели немного изменены границы некоторых элементов вторичных структур, в основном спиралей. Некоторые α -спирали укорочены на 1 остаток, на нескольких участках спирали заменены на неструктурированные петли – получились гибкие линкеры между спиральями (рис. 14). Эти изменения выглядят правдоподобно.

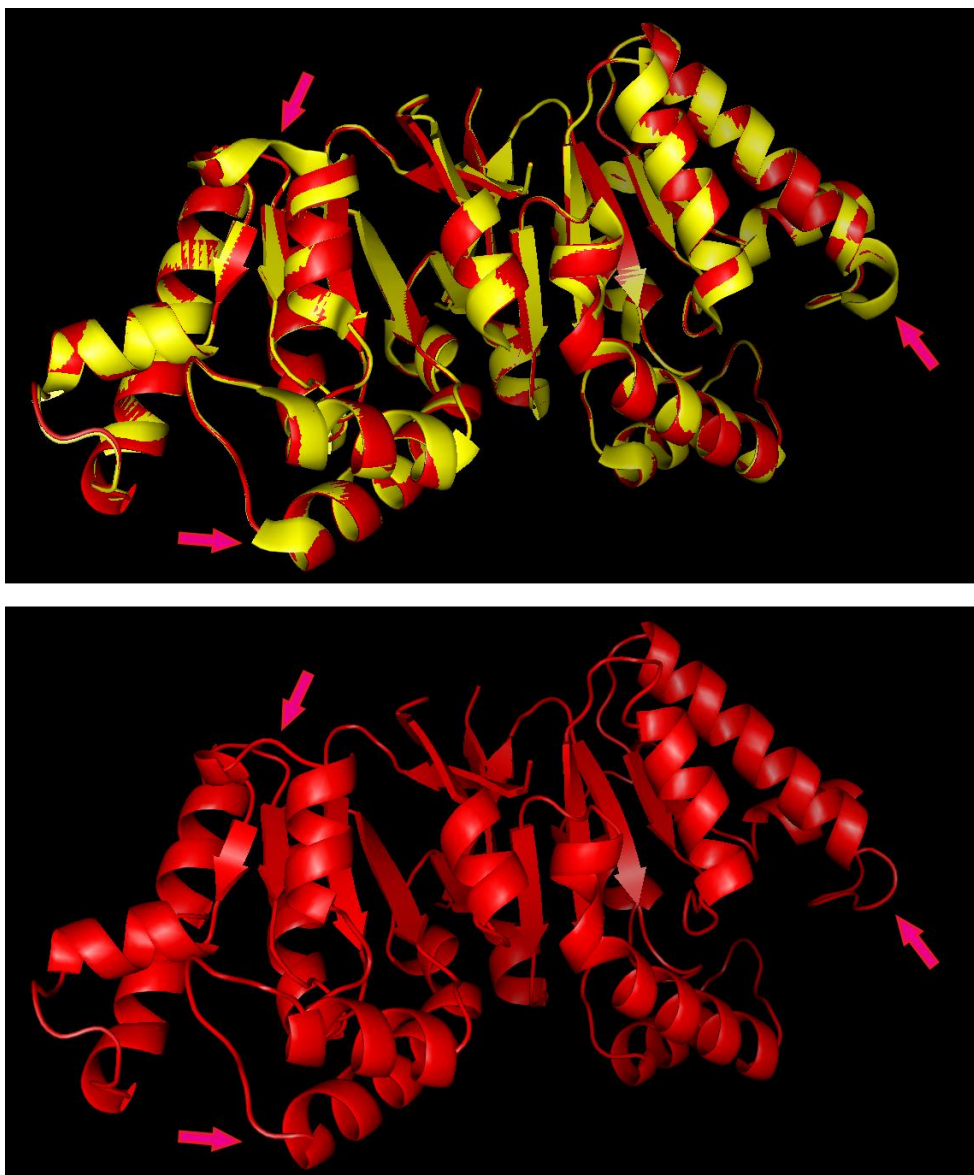


Рис. 14. Совмещение исходной (желтая) и оптимизированной модели (красная) структуры 4GP7 (вверху) и изображение оптимизированной модели (внизу). Стрелки указывают на исчезнувшие в оптимизированной модели участки вторичных структур.

На рисунке 15А изображен Leu155 цепи В, радикал которого обращен в пространство между двумя β -листами и который значительно улучшили свое положение после оптимизации структуры, по мнению сервиса PDB_redo. На рисунке 15В изображен остаток Ans49, расположенный в плохо совмещенной петле в цепи В, который после оптимизации только ухудшил свое положение (он так же плохо вписывается в электронную плотность, как и в исходной структуре; низкие показатели качества для данного участка обусловлены некачественным восстановлением электронной плотности в этой области).

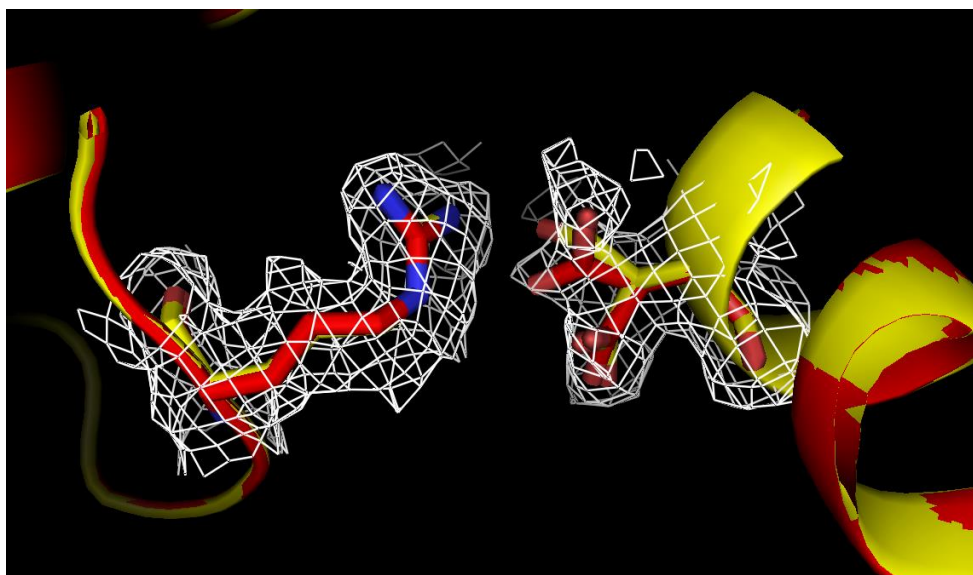


Рис. 15. Совмещение исходной (желтая) и оптимизированной модели (красная) структуры 4GP7. Показаны Asp38 (справа) и Arg123 (слева) в цепи А, которые были сильно сближены в исходной модели и перекрывались, поэтому были маргиналами. В модели из PDB_redo изменена конформация радикала Asp38, расстояние между остатками оптимизировано.

Заключение

Структура 4GP7 неплохо определена, по сравнению со структурами из PDB того же разрешения. Однако она имеет ряд проблем, заключающихся в основном в том, что данного разрешения не хватило для точного определения координат отдельных атомов. Из-за этого возникали трудности в определении конформаций радикалов аминокислот, в то время как остов цепи восстановлен хорошо. Часть допущенных в исходной структуре ошибок расшифровки была исправлена с помощью оптимизации PDB_redo.

Литература

- [1] L. K. Wang, P. Smith, and S. Shuman, "Structure and mechanism of the 2',3' phosphatase component of the bacterial Pnkp-Hen1 RNA repair system," *Nucleic Acids Res.*, vol. 41, no. 11, pp. 5864–5873, 2013.
- [2] G. Kleywegt, M. Harris, J. Zou, T. Taylor, A. Wählby, and T. Jones, "Electron Density Server," *The Uppsala Electron-Density Server*, 2004. [Online]. Available: <http://eds.bmc.uu.se/eds/>.
- [3] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne, "The Protein Data Bank," *Nucleic Acids Res.*, vol. 28, no. 1, pp. 235–242, Jan. 2000.
- [4] [Online]. Available: http://ftp.wwpdb.org/pub/pdb/validation_reports/gp/4gp7/4gp7_full_validation.pdf.
- [5] R. P. Joosten, F. Long, G. N. Murshudov, and A. Perrakis, "The PDB_REDO server for macromolecular structure model optimization," *IUCrJ*, vol. 1, no. Pt 4, pp. 213–220, Jul. 2014.
- [6] [Online]. Available: http://www.cmbi.ru.nl/pdb_redo/gp/4gp7/index.html.
- [7] [Online]. Available: http://www.cmbi.ru.nl/pdb_redo/gp/4gp7/4gp7_dRSCC.svgz.