

Московский государственный университет имени М.В. Ломоносова

Факультет биоинженерии и биоинформатики

Отчет по структурной биоинформатике студента четвертого курса

Холиной Татьяны Дмитриевны

Качество расшифровки структуры белка лептина (PDB-идентификатор 1AX8) методом рентгеноструктурного анализа

Преподаватели:

Е. А. Аксянов

А. В. Алексеевский

С. А. Спирин

В. Ю. Лунин

Москва, 2016

Введение

Лептин - сигнальный белок длиной 149 аминокислотных остатков, контролирующий чувство голода. Он был открыт на мутантных мышах, страдающих ожирением - у них обнаружили неактивную форму лептина. Лептин продуцируется жировой тканью и через кровь переносится к центру голода в головном мозге, где подавляет аппетит. При голодании количество продуцируемого лептина падает, что воспринимается организмом как сигнал к запасанию энергии. Лептин играет важную роль при ожирении, которое возникает как при мутантных формах лептина, так и при нарушениях в его сигнальном пути [2].

Лептин состоит из четырех α -спиралей (A-D), соединенных двумя кроссоверными петлями и одной короткой петлей. Последние 5 остатков спирали D формируют 3_{10} -спираль [1].

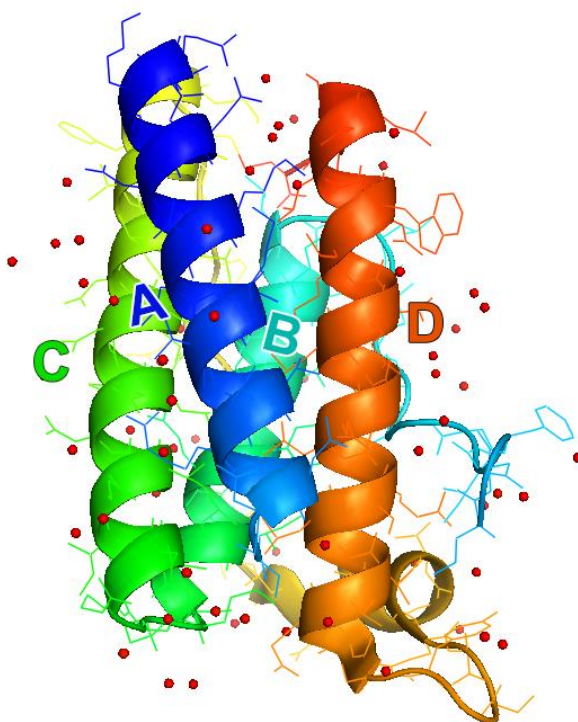


Рис.1. Структура лептина

Расшифровка кристаллической структуры

Кристаллическая структура лептина была расшифрована в 1997 году командой Zhang et al. Расшифрованная структура хранится на сервере PDB под идентификатором 1AX8 [3]. Лептин был выбран «молекулой месяца» RCSB в мае 2012 г [2].

В записи [3] хранится молекула лептина длиной 130 аминокислотных остатков и молекулы воды. Длина белка 146 аминокислотных остатков, но в модели присутствует только 130.

Разрешение структуры 2.40 Å. В молекуле лептина остаток Trp100 заменен на глутамат для уменьшения агрегации белка и облегчения кристаллизации [1].

Кристаллографическая ячейка имеет размеры a=88.14 Å, b=88.14 Å, c=48.05 Å, с углами alpha=90.00, beta=90.00, gamma=120.00. В асимметрической ячейке нет некристаллографических симметрий, т.к. лептин асимметричен.

Решение фазовой проблемы

Фазовая проблема решена множественным изоморфным замещением с помощью двух производных, содержащих тяжелые атомы – один с платиной, второй с осмием [1].

Улучшение модели

В записи PDB число измеренных рефлексов равно 8253, а число рефлексов для R-free равно 877. Полнота данных составляет 99.0%, и диапазон разрешений равен от 2.4 до 8.0 Å. R-value равняется 18.5, а R_{free} равняется 28.3, что дает разницу 9.8%. В статье же число рефлексов равно 7900 (число рефлексов для R-free не приведено), R-value равняется 20.5, а R_{free} – 29.0, что дает разницу 8.5%. В обоих случаях разница между R-value и R_{free} приемлема. В EDC [4] полнота данных равна 97.5%, и диапазон разрешений равен 29.89 - 2.40 Å.

Качество модели

Отчет PDB

Приведенная модель белка неполная: в ней отсутствуют остатки 1-2 и 25-38 (11% от длины белка). 63% остатков не являются маргинальными, тогда как 23% маргинальны по одному признаку и 3% - по двум [5].

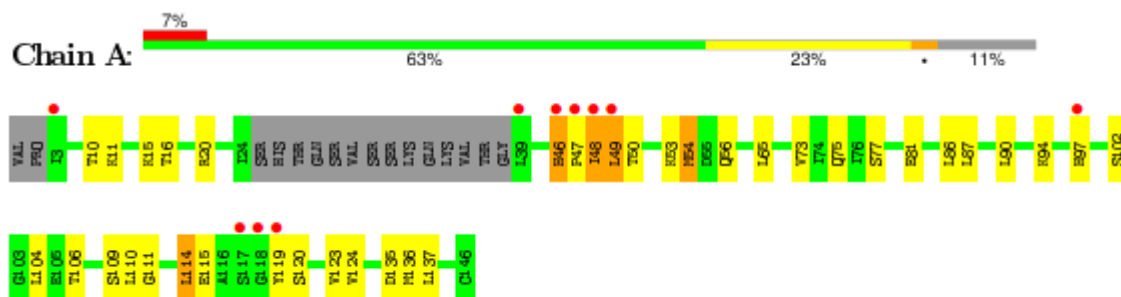


Рис.2. Качество остатков модели 1AX8. Серым отмечены остатки, отсутствующие в модели. Остальные остатки покрашены по количеству геометрических критериев, по которым они выбиваются: зеленые = 0, желтые = 1, оранжевые = 2. Красными точками отмечены остатки, плохо подходящие под электронную плотность (RSRZ>2)[5].

PDB-отчет приводит следующие маргинальные остатки:

Таблица 1. Маргинальные остатки из отчета PDB

Вид маргинальных остатков и атомов	Количество	Остатки
Число слишком близких контактов	38 атомов (19%)	
Остатки с нестандартным углом связи боковой цепи	2 остатка	Met54, Met136
Остатки, выбивающиеся по Рамачандрану	2 остатка	His46, Ile48
Остатки с неротамерной боковой цепью	9 остатков	Leu49, Ser77, Lue87, Leu90, Lys94, His97, Leu114, Asp135, Leu137
Остатки с инвертированной боковой цепью	1 остаток	Gln56
Остатки, выбивающиеся по RSRZ	10 остатков	Ile3, Leu39, His46, Pro47, Ile48, Leu49, His97, Ser117, Gly118, Tyr119

Качество модели по MolProbability

Маргинальные остатки можно подтвердить сервисом MolProbability^[6]. Он также позволяет добавить атомы водорода к модели, после чего можно определить перекрывающиеся атомы (Clashscore) и инвертированные боковые цепи.

Таблица 2. Маргинальные остатки по версии сервиса MolProbability

All-Atom Contacts	Clashscore, all atoms	19.21		67 th percentile* (N=331, 2.40Å ± 0.25Å)
Protein Geometry	Poor rotamers	8	6.78%	Goal: <0.3%
	Favored rotamers	101	85.59%	Goal: >98%
	Ranachandran outliers	2	1.59%	Goal: <0.05%
	Ramachandran favored	120	95.24%	Goal: >98%
	MolProbability^ Score	2.74		49 th percentile* (N=8058, 2.40Å ± 0.25Å)
	C-β deviations > 0.25 Å	1	0.81%	Goal: 0
	Bad Bonds	0/1016	0.00%	Goal: 0%
	Bad Angles	2/1380	0.14%	Goal: <0.1%

Peptide Omegas	Cis Prolines	0/5	0.00%	Expected: ≤ 1 per chain, or $\leq 5\%$
----------------	--------------	-----	-------	---

* 100th percentile is the best among structures of comparable resolution; 0th percentile is the worst. For clashscore the comparative set of structures was selected in 2004, for MolProbity score in 2006.

^ MolProbity score combines the clashscore, rotamer, and Ramachandran evaluations into a single score, normalized to be on the same scale as X-ray resolution.

Количество маргинальных атомов в каждой категории совпадает с приведенным в отчете PDB, хотя в PDB на один остаток с неротамерной цепью больше. MolProbity также нашел плохой угол связи в лейцине 49, но не нашел в метионине 54.

Таблица 3. Некоторые маргинальные остатки по данным MolProbity

Остаток	Clash > 0.4Å	Рамачандран	Плохие ротамеры	Плохой угол связи	Плохая длина Cβ-связи
His46	++	++	+		
Pro47	++	+			
Leu48	++	++			++
Leu49	++		++	++	
Leu87			++		
Leu90			++		
Lys94			++		
His97			++		
Ser109			++		
Leu110	++		+		
Leu114	++		++		
Met136	++			++	
Leu137			++		

Анализ маргинальных остатков

1. Глутамин 56

У этого остатка инвертирована боковая цепь, т.е. ее нужно повернуть, чтобы избежать конфликтов атомов водорода. По другим параметрам остаток не выбивается из группы.

На рис. 3 изображена электронная плотность Gln56. Видно, что боковая цепь очень хорошо вписывается в электронную плотность.

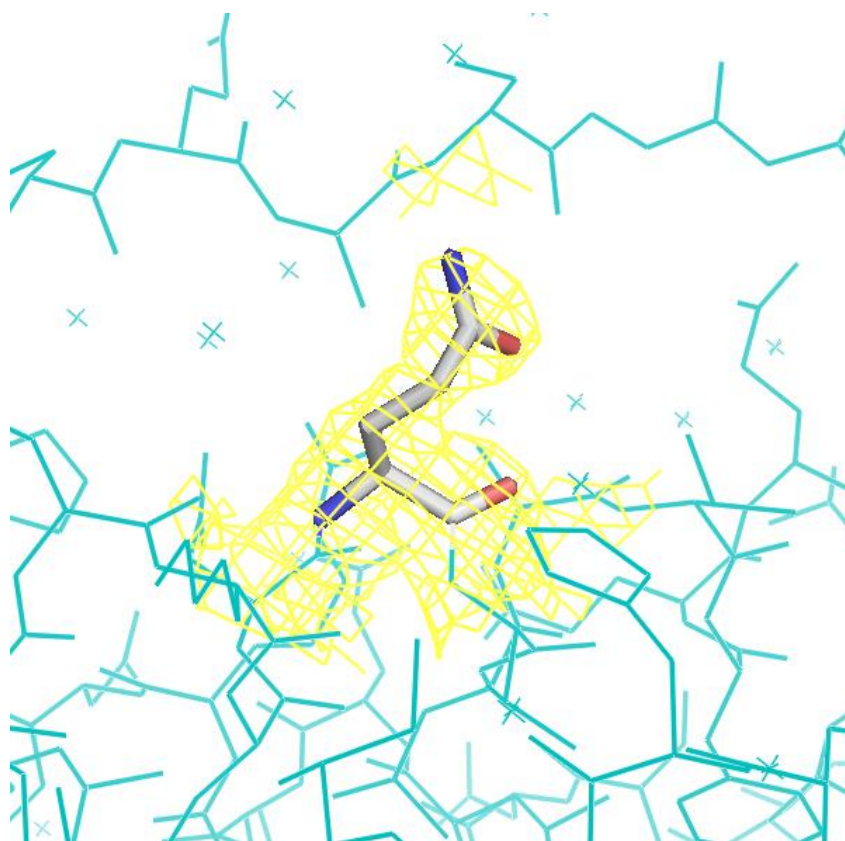


Рис. 3. Электронная плотность вокруг глутамина 56 при уровне подрезки 1.5 σ

2. Изолейцин 48

Этот остаток маргинален сразу по трем параметрам: он перекрывается с 53 LYS HD3, выбивается по Рамачандрану (значения углов -97.4,-148.3), и у него увеличена длина C- β связи (0.31Å).

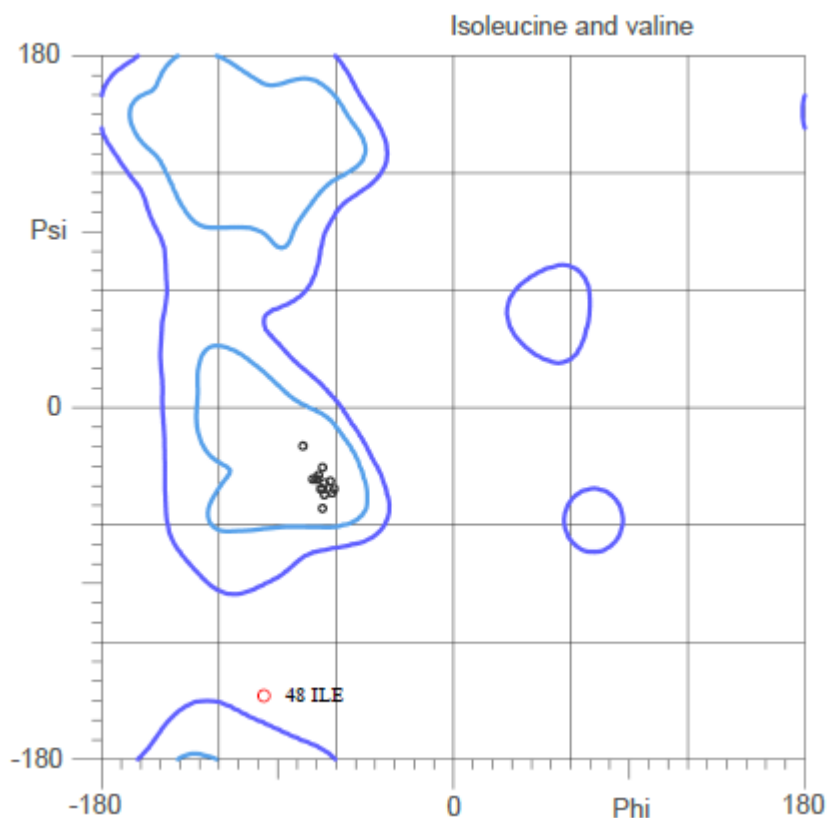


Рис. 4. Карта Рамачандрана для остатков изолейцина и валина в белке 1AX8 [7]

При рассмотрении электронной плотности (рис. 5) данного остатка видно, что электронное облако покрывает лишь часть остова аминокислоты и вообще не покрывает боковую цепь. Ситуация не улучшается даже при снижении уровня подрезки. Таким образом, электронная плотность не дает достаточно данных для правильного определения положения аминокислоты.

3. Лейцин 49

Этот остаток также маргинален по трем параметрам: перекрывается с 48 ILE HG22, у него плохие ротамеры (значения углов χ : 28.2,55.3), и нестандартный угол между атомами (угол N-C α -C равен 4.6 $^\circ$). На рис. 5 Показана электронная плотность Leu49 вместе с Ile48. На уровне подрезки 1.5 σ видно, что положение боковой цепи лейцина скорее всего определено неверно – связь C β -C γ следует повернуть в другую сторону для лучшего вписания в электронную плотность. При этом угол N-C α -C, определенный MolProbability как нестандартный, хорошо вписывается в электронную плотность. Необычный угол может быть связан с расположением остатка в повороте петли.

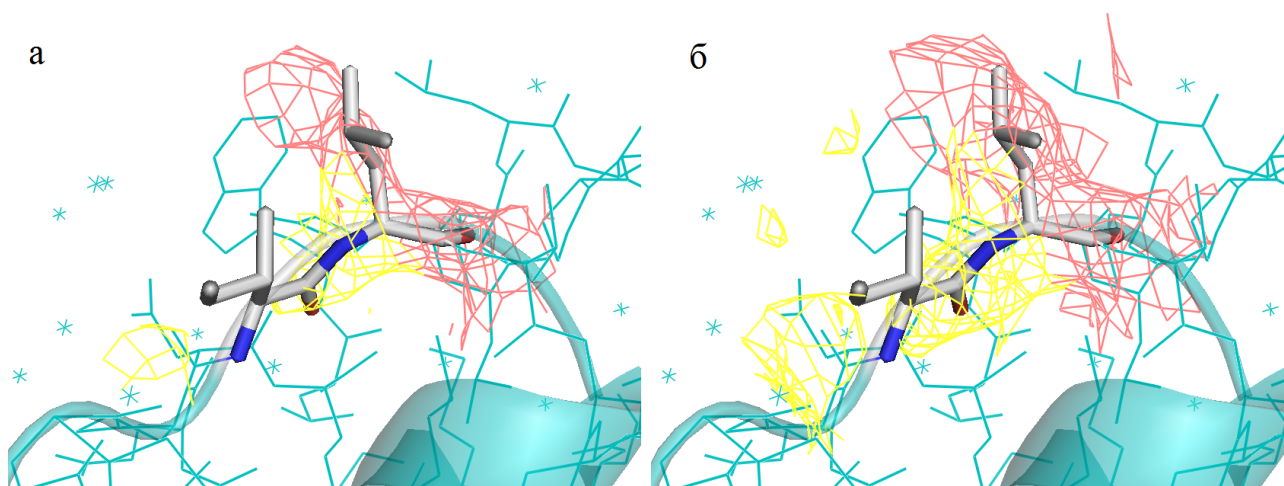


Рис. 5. Электронная плотность изолейцина 48 (желтый) и лейцина 49 (розовый). а) – подрезка на уровне 1.5 σ , б) – на уровне 0.5 σ

4. Метионин 136

Этот остаток имеет плохой Clashscore (He1 перекрывается с A 58 LEU HD21), а также нестандартный угол CG-SD-CE в боковой цепи (угол в модели примерно равен 110, тогда как оптимальный угол равен 100). Электронная плотность вокруг остатка Met136 показана на рис. 6.



Рис. 6. Электронная плотность метионина 136. Отмечен угол и атомы CG-SD-CE. Уровень подрезки 1.5 σ

В общем-то боковая цепь вписано достаточно неплохо. Видно расширение электронного облака вокруг крупного атома серы. Однако можно было бы подвинуть атом Cepsilon с уменьшением угла CG-SD-CE, и вписаться в плотность даже лучше, чем в модели.

5. Гистидин 46

Маргинальный остаток по Рамачандрану (группа остатков перед пролином, углы 34.5,37.2), и плохой ClashScore (O перекрывается с 48 ILE HG13). При этом он не сильно выбивается на карте, находясь на границе допустимой (рис. 7, темно-синий) зоны.

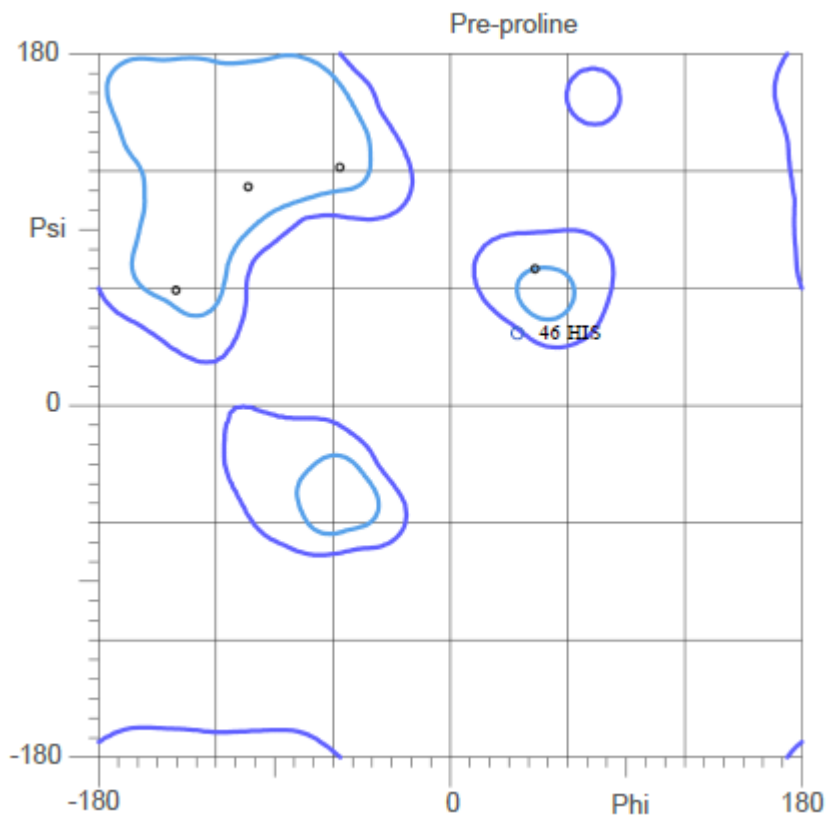


Рис. 7. Карта Рамачандрана для остатков перед пролином в белке 1AX8 [7]

На рис. 8 показана электронная плотность вокруг His46 и соседних Leu45 и Pro47. В общем-то, гистидин и лейцин неплохо вписываются в электронную плотность как по остовам, так и боковыми цепями. Кольцо пролина также расположено в месте с высокой электронной плотностью. Возможно, это просто необычный изгиб петли, хотя я не нашла в окружении остатков, способных его стабилизировать.

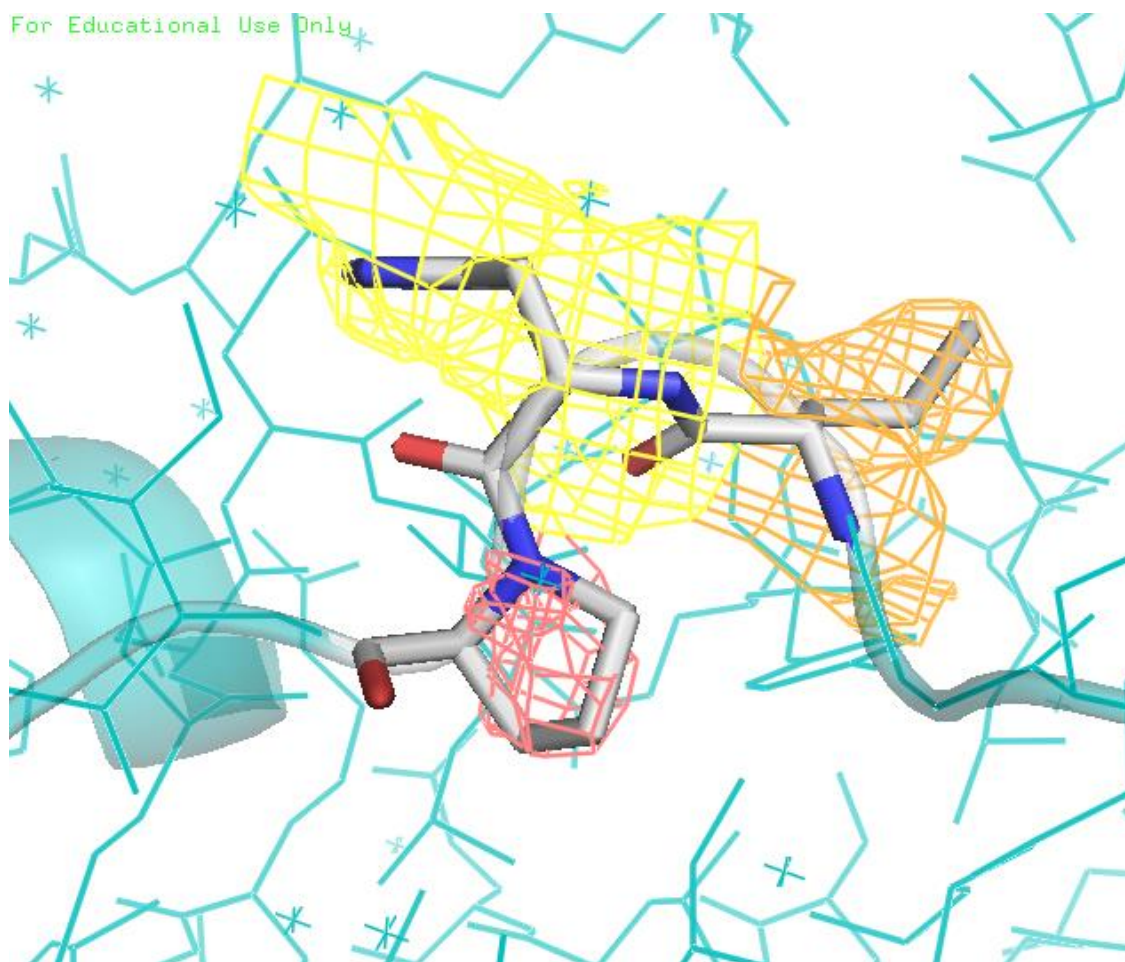


Рис. 8. Электронная плотность вокруг лейцина 45, гистидина 46 и пролина 47.
Уровень подрезки 1.0 σ

Сравнение с моделью из PDB_Redo

PDB_redo ^[8] – сервис, оптимизирующий кристаллическую структуру молекул. Он использует несколько сторонних программ, таких как протокол проверки качества WHAT_CHECK, для улучшения модели. PDB_redo очищает входные данные, например убирает неизвестные атомы из PDB-файла и плохие рефлекссы, извлекает нужные параметры, улучшает и перестраивает структуру в несколько шагов.

Для белка 1AX8 есть запись, созданная в 2015 году ^[9].

Таблица 4. Сравнение параметров оригинальной и улучшенной модели по алгоритму WHAT_CHECK

	Original PDB entry	Re-refined (only)	Re-refined and rebuilt
1st generation packing quality ¹	0.006	0.307	0.356
2nd generation packing quality ¹	0.080	0.394	0.982
Ramachandran plot appearance ¹	-1.541	1.523	1.387
Chi-1/Chi-2 rotamer normality ¹	-2.967	-1.633	-0.574
Backbone conformation ¹	1.019	0.975	0.966

Bond length RMS Z-score ²	0.438	0.471	0.451
Bond angle RMS Z-score ²	0.705	0.720	0.678
Total number of bumps ³	30	7	3
Unsatisfied H-bond donors/acceptors ³	10	7	8

¹ Higher is better, ² Should be lower than 1.000, ³ Fewer is better

PDB_redo сильно улучшила многие параметры, такие как карты Рамачандрана и качество ротамеров, однако немного уменьшила качество конформации остова (Backbone conformation).

Выданный PDB_redo PDB-файл я подала сервису MolProbability:

В частности, исчезло предложение провести инверсию Glu56, но появились инверсии Gln75 и Asn82. При этом Glu56 сильно рекомендовал перевернуть цепь, Gln75 пишет, что «Both orientations clash but flip was preferred», Asn82 – «Some evidence recommending flip».

Таблица 5. Маргинальные остатки для улучшенной структуры по версии сервиса MolProbability

All-Atom Contacts	Clashscore, all atoms:	2.46		100 th percentile* (N=331, 2.40Å ± 0.25Å)
	Clashscore is the number of serious steric overlaps (> 0.4 Å) per 1000 atoms.			
Protein Geometry	Poor rotamers	6	5.08%	Goal: <0.3%
	Favored rotamers	109	92.37%	Goal: >98%
	Ramachandran outliers	1	0.79%	Goal: <0.05%
	Ramachandran favored	123	97.62%	Goal: >98%
	MolProbity score [^]	1.65		99 th percentile* (N=8058, 2.40Å ± 0.25Å)
	Cβ deviations >0.25Å	0	0.00%	Goal: 0
	Bad bonds:	0 / 1016	0.00%	Goal: 0%
	Bad angles:	0 / 1380	0.00%	Goal: <0.1%
Peptide Omegas	Cis Prolines:	0 / 5	0.00%	Expected: ≤1 per chain, or ≤5%

Таблица 6. Маргинальные остатки в оригинальной модели 1AX8 и модели с сервера PDB_redo

Статистика	Оригинальная модель	Модель WHAT_CHECK
Clashscore	19.21	2.46
Плохие ротамеры	8	6
Выбросы по Рамачандрану	2	1
Выбросы C β	1	0
Плохие углы	2	0
MolProbity score	2.74	1.65
Инверсии	1	2

Очевидно, что модель, полученная PDB_redo, значительно лучше оригинальной.

Заключение

Несмотря на достаточно среднее разрешение (2.4 Å), структура разрешена достаточно неплохо. Один из главных недостатков – отсутствия участка 27-38, который, как пишут авторы статьи, был слишком хаотичен, чтобы определить структуру. В ней достаточно мало маргинальных остатков. Анализ отдельных маргинальных остатков показал, что конформация большинства из них либо не сильно выбивается из данных электронной плотности, либо не может быть точно определена по имеющимся данным, то есть исследователи не допускали крупных ошибок. Однако перерасчитанные параметры структуры на сайте PDB_redo показывают, что из оригинальных данных можно было получить структуру, гораздо лучшую, чем выложенная на PDB.

Список литературы, ссылки и использованные сервисы

1. Zhang F1, Basinski MB, Beals JM, Briggs SL, Churgay LM, Clawson DK, DiMarchi RD, Furman TC, Hale JE, Hsiung HM, Schoner BE, Smith DP, Zhang XY, Wery JP, Schevitz RW, "Crystal structure of the obese protein leptin-E100". Nature. 1997 May 8;387(6629):206-9. DOI [10.1038/387206a0](https://doi.org/10.1038/387206a0)
2. <https://pdb101.rcsb.org/motm/149>
3. <http://www.rcsb.org/pdb/explore/explore.do?structureId=1AX8>
4. <http://eds.bmc.uu.se/cgi-bin/eds/uusfs?pdbCode=1ax8>
5. http://ftp.wwpdb.org/pub/pdb/validation_reports/ax/1ax8/1ax8_full_validation.pdf
6. <http://molprobity.biochem.duke.edu/>
7. http://www.rcsb.org/pdb/images/1AX8_ram_m_500.pdf
8. http://www.cmbi.ru.nl/pdb_redo/index.html
9. http://www.cmbi.ru.nl/pdb_redo/ax/1ax8/index.html