

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

ФАКУЛЬТЕТ БИОИНЖЕНЕРИИ И БИОИНФОРМАТИКИ

Отчёт по качеству RSA расшифровки структуры с PDB ID 4QUO

Карань Анна
4 курс, 402 группа

**Москва
2018 г.**

СОДЕРЖАНИЕ

Аннотация	3
Введение	4
Результаты и обсуждение	6
Общая информация о модели	6
Значения индикаторов качества модели в целом	6
R-фактор	7
Перекрытия Ван-дер-Ваальсовых радиусов атомов (Clashscore)	7
Доля маргинальных остатков по карте Рамачандрана (Ramachandran outliers)	8
Доля маргиналов по положению боковой цепи (ротамеров) (Sidechain outliers)	10
Доля маргиналов по RSR (RSRZ outliers)	10
Анализ маргинальных остатков	11
GLU711	12
ARG361	13
THR4	13
ASN63	14
Сравнение модели из PDB с моделью из PDB_redo	14
Заключение	15
Список литературы	16

Аннотация

В данной работе был проведен анализ качества расшифровки кристаллической структуры аминоксептидазы N в комплексе с LL-(R,S)-hPheP[CH₂]Phe(3-CH₂NH₂), аналогом фосфинового дипептида, с PDB ID 4QUO: рассмотрены ключевые индикаторы качества модели, несколько маргинальных остатков и структура из PDB-REDO.

Введение

У данного белка существует множество названий, отражающих его функцию и локализацию:

- аланинаминопетидаза
- аланин специфическая аминопептидаза
- аланиламинопетидаза
- аминолигопептидаза
- аминопетидаза М
- аминопептидаза N
- миелоидный гликопротеин плазматической мембраны CD13
- цистеингликиназа
- цистеинилглицин дипептидаза
- L-аланин аминопептидаза
- мембранная аланин аминопептидаза
- мембранная аминопептидаза I
- микросомальная аминопептидаза
- пептидаза E
- псевдолейцин аминопептидаза

Аминопептидаза N представляет собой металлопротеазу типа II, которая принадлежит к семейству M1 клана MA, состоящую из 868 аминокислот с коротким N-концевым цитоплазматическим доменом, одной трансмембранной частью и большим эктодоменом, содержащим активный центр. APN имеет молекулярный вес 98,7 кДа, существует в двух формах: мембранная аминопептидаза N и растворимая аминопептидаза N. Более того, он демонстрирует присутствие различных изозимов с различными функциями. APN - это повсеместный фермент, присутствующий во множестве организмов в частности в самых разных органах, тканях и клетках человека. Это многофункциональный фермент, связанный с онкогенезом, иммунной системой, болью и т. Д. Кроме того, он также служит рецептором для коронавирусов и других вирусов человека. Помимо проявления различных других функций, APN также участвует в обрезании антигена и процессе презентации антигена. Эти функции облегчают модуляцию биоактивных пептидных ответов (устранение боли, высвобождение вазопрессина) и влияют на иммунные функции и основные

биологические события (пролиферация клеток, секреция, инвазия, ангиогенез), обеспечивая тем самым варианты лечения для многих видов заболеваний [1].

Данный белок в классификации ферментов обладает номером EC 3.4.11.2, т.е. относится к гидролазам (3), гидролизующим пептидную связь (4), конкретно аминопептидазам (11).

Release of an N-terminal amino acid, Xaa⁺Yaa⁻ from a peptide, amide or arylamide. Xaa is preferably Ala, but may be most amino acids including Pro (slow action). When a terminal hydrophobic residue is followed by a prolyl residue, the two may be released as an intact Xaa-Pro dipeptide

Конкретно данный фермент осуществляет реакцию высвобождения N-концевой аминокислоты, Xaa⁺Yaa⁻ из пептида, амида или ариламида. Xaa предпочтительно является Ala, но может быть и любой другой аминокислотой, включая Pro, просто реакция будет осуществляться с заметно более низкой эффективностью. Когда за конечным гидрофобным остатком следует пролильный остаток, они могут высвобождаться как интактный дипептид Xaa-Pro.

Структура с PDB ID 4quo была получена на основе белка *Neisseria meningitidis* MC58. Работа по определению структуры проводилась с целью определения оптимальных взаимодействий ингибитор-фермент [2], так как *N. meningitidis* является опасным патогеном, вызывающим менингококковый нозофарингит и пневмонию, генерализованные формы - менингококцемию, менингококковый менингит и менингоэнцефалит, а также смешанные формы (артрит и др.) [3].

Результаты и обсуждение

Общая информация о модели

Состав комплекса: 1 полипептидная цепь А в комплексе с LL-(R,S)-hPheP[CH₂]Phe(3-CH₂NH₂), аналогом фосфинового дипептида.

- Атомы белка: 6847
- Атомы нуклеиновых кислот: 0
- Гетерогенные атомы: 134
- Атомы растворителя: 1041

Год: 10.09. 2014 - дата загрузки структуры в БД.

Фамилии авторов: Vassiliou S, Węglarz-Tomczak E, Berlicki L, Pawełczak M, Nocek B, Mulligan R, Joachimiak A, Mucha A.

Метод решения фазовой проблемы: Одноволновое аномальное рассеяние

Число измеренных рефлексов: 117723

Разрешение: 1.65 Å

Полнота набора структурных факторов: 90.6 %

Диапазон разрешений структурных факторов: 27.98 Å - 1.65 Å

Параметры кристаллографической ячейки: a=223.71 Å, b=223.71 Å, c=57.77 Å,
 $\alpha = 90.00$, $\beta = 90.00$, $\gamma = 120.00$

Наличие некристаллографических симметрий в асимметрической ячейке: нет

Тип кристаллографической симметрии: H 3

Значения индикаторов качества модели в целом

Основными показателями качества модели являются: Rfree (свободный R-фактор), критерий перекрывания Ван-дер-Ваальсовых радиусов атомов (Clashscore), доля маргинальных остатков по карте Рамачандрана (Ramachandran outliers), доля маргиналов по положению боковой цепи (Sidechain outliers) и доля маргиналов по RSR (RSRZ outliers). На Рис. 1 показаны значения этих параметров для рассматриваемой структуры в сравнении с другими структурами, полученными методом PCA. Далее каждый критерий будет рассматриваться подробнее.

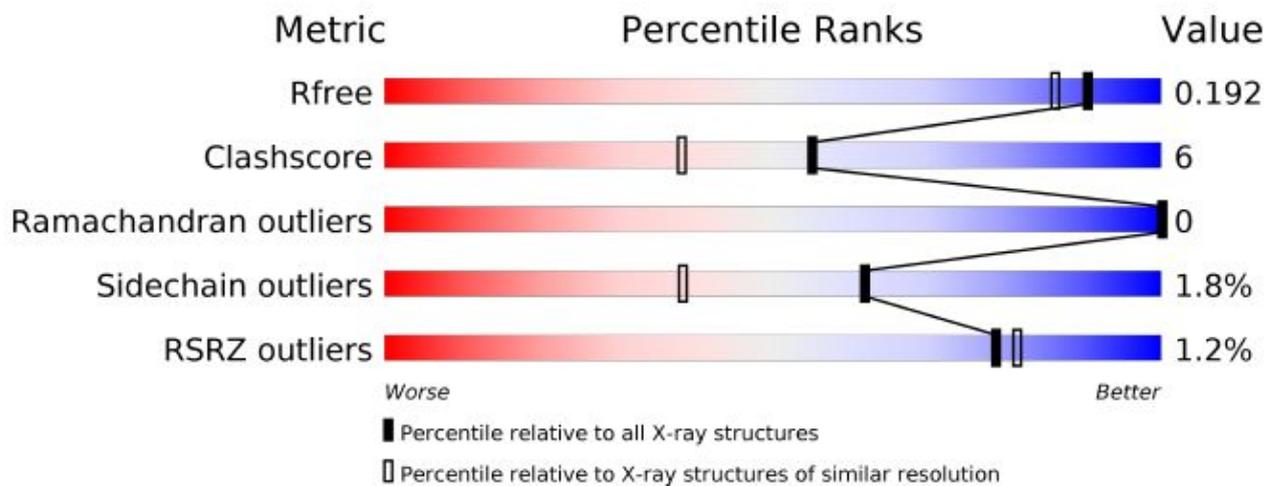


Рис. 1. Основные показатели качества модели (черным – относительно всех PCA моделей; белым – относительно PCA моделей близкого разрешения).

R-фактор

R-фактор характеризует соответствие модели ее экспериментальным значениям (модулям структурных факторов). F^{obs} - модули структурных факторов (они же амплитуды гармоник Фурье), полученные в PCA-эксперименте, F^{calc} - модули структурных факторов, посчитанные по построенной модели.

$$R = \frac{\sum_{hkl} |F_{hkl}^{calc} - F_{hkl}^{obs}|}{\sum_{hkl} F_{hkl}^{obs}} * 100\%$$

Rfree – фактор, считаемый аналогично R-фактору, только на некотором количестве (в данной структуре 5.1%) случайно выбранных рефлексов, которые не использовались при оптимизации модели по R-фактору. Хорошие значения: Rfree < 25%, плохие значения: Rfree > 40%. Значения (Rfree – R) > 10% настораживают в отношении переоптимизации. R-фактор для данной структуры равен 0.148, Rfree - 0.182, т.е. и значение Rfree хорошее, и разница Rfree и R не говорят о переоптимизации.

Перекрытия Ван-дер-Ваальсовых радиусов атомов (Clashscore)

Этот параметр равен количеству перекрытий Ван-дер-Ваальсовых радиусов больше, чем на 0.4 Å, на 1000 атомов. Для данной структуры он равен 6. Это значение лучше, чем у 54% структур с похожим разрешением и 43% всех структур, полученных методом PCA (Рис. 1). Т.е. качество структуры по данному критерию не очень хорошее, среднее.

Доля маргинальных остатков по карте Рамачандрана (Ramachandran outliers)

Карта Рамачандрана показывает распределение аминокислот в соответствии со значениями двух торсионных углов связей атомов остова: ψ и ϕ (Рис. 2).

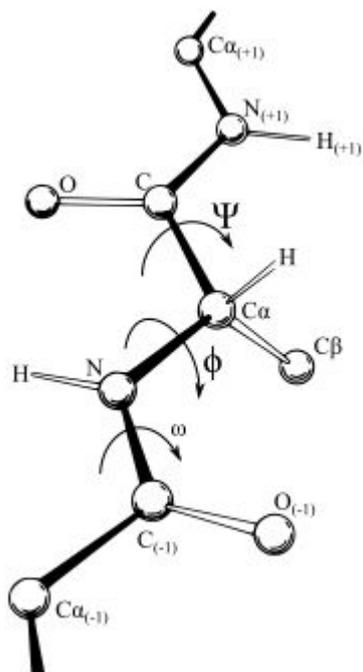


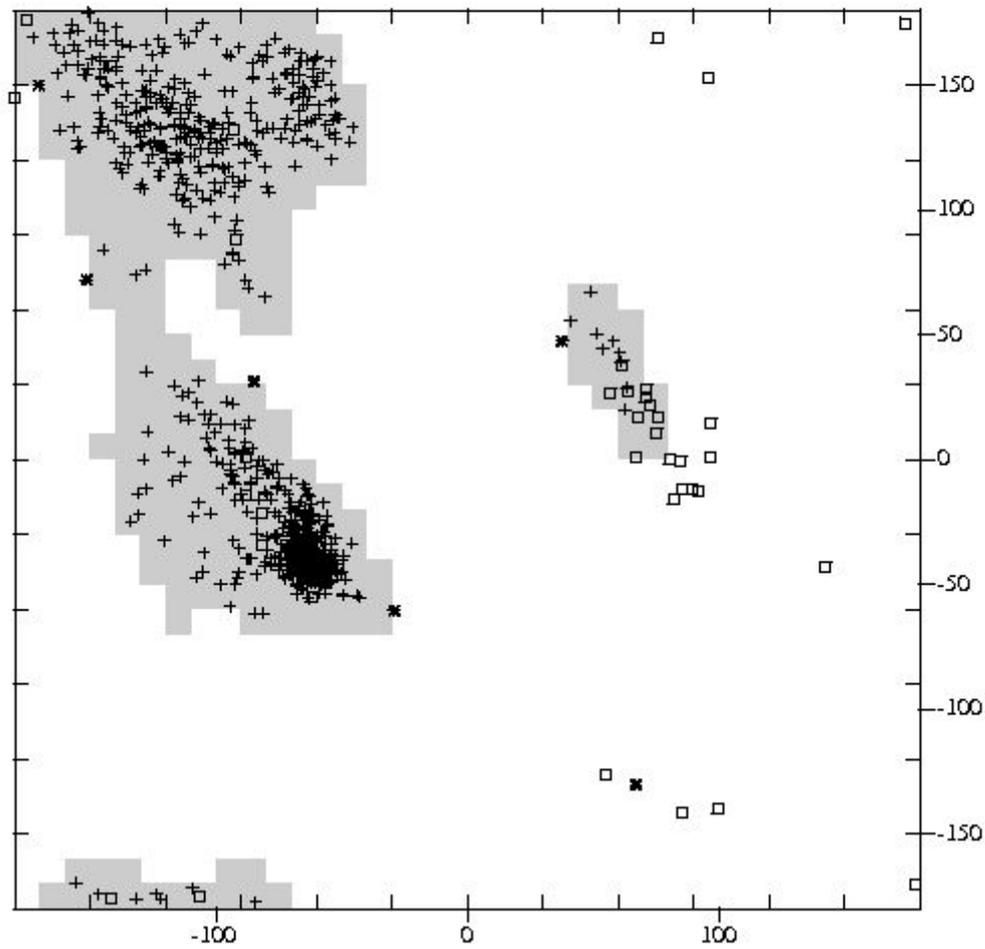
Рис. 2. Торсионные углы пептидной цепи

Получившуюся плоскость можно разбить на три области – запрещенные, разрешенные и предпочитаемые конформации в соответствии со встречаемостью в ранее расшифрованных структурах. Естественно чаще встречаются конформации с наименьшей энергией, именно они и будут для каждого остатка предпочитаемыми. Остатки, обладающие в структуре конформациями с высокой энергией, называются маргинальными (маргиналами).

Карта Рамачандрана для данной структуры была получена с помощью программы MOLEMAN2 [4] (Рис. 3). В результате 6 остатков оказались маргинальными (ASN63, GLU260, LEU264, ALA331, ASN452, CYS830), что составляет 0,7%. Хорошим значением считается 0 - 5%.

Если же использовать программу MolProbity [5], то 98,1% остатка находятся в предпочитаемых конформациях, и 100% в разрешенных. Т.е. обнаружено 0.00% маргинальных остатков. Хорошее значение для данной программы - <0.05%.

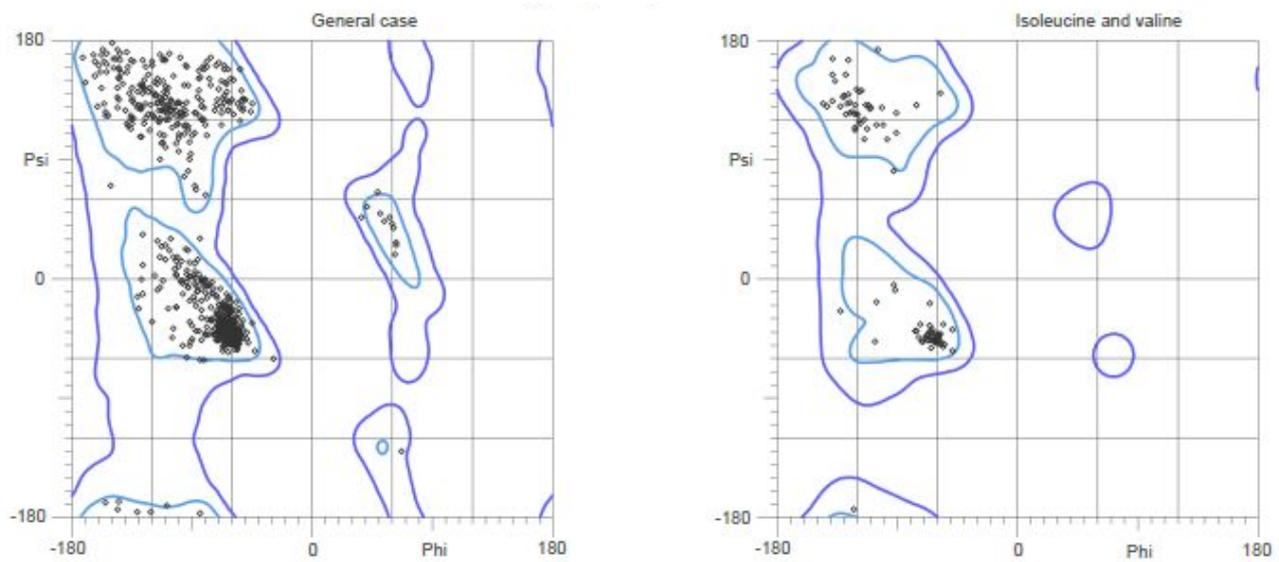
Т.е. по этому критерию (вне зависимости от используемой программы) модель хорошая.



X = PHI mapped to [-180,180>

Y = PSI mapped to [-180,180>

Рис. 3. Карта Рамачандрана, полученная с помощью программы MOLEMAN2.



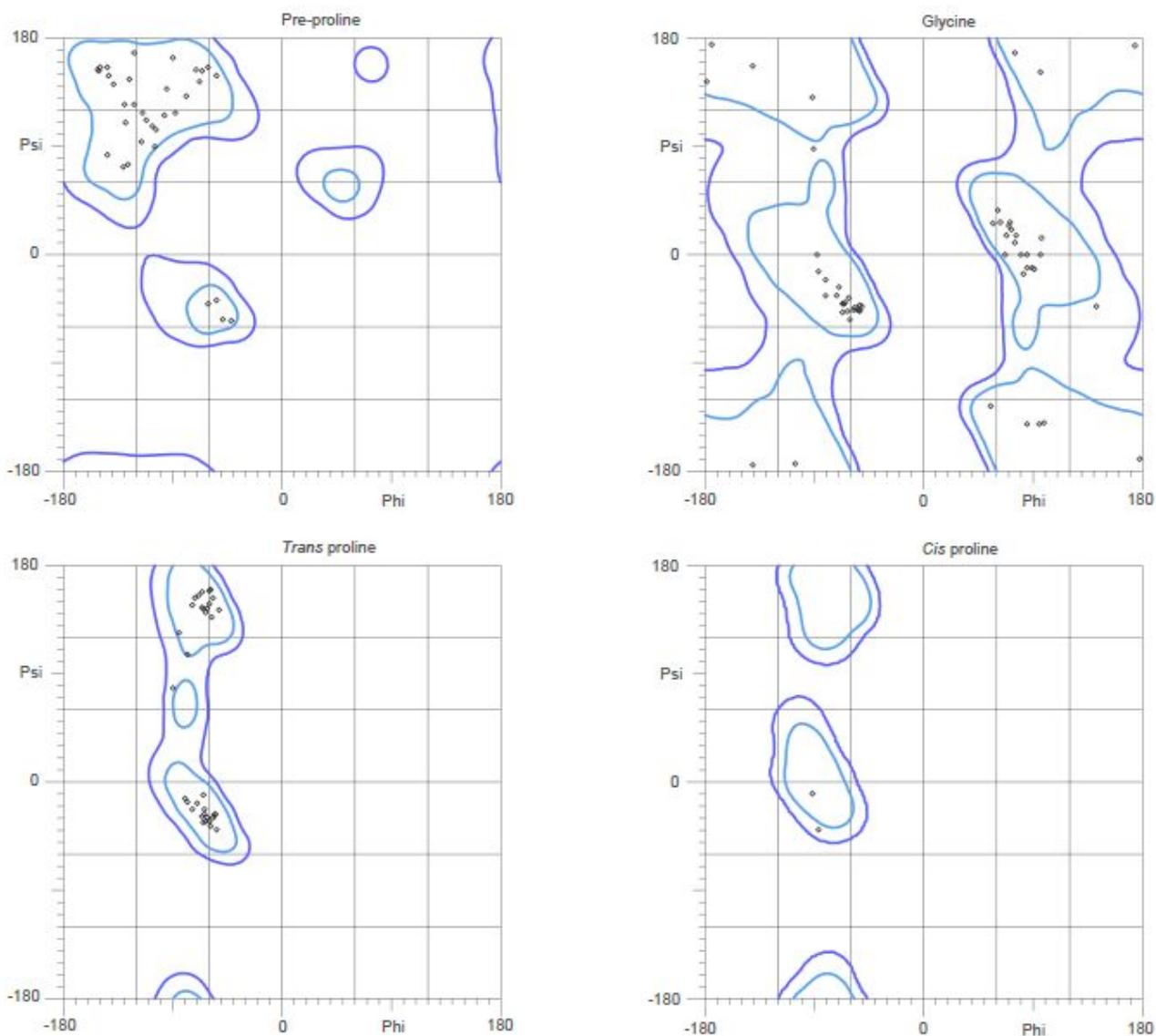


Рис. 4. Карта Рамачандрана, полученная с помощью программы MolProbity.

Доля маргиналов по положению боковой цепи (ротамеров) (Sidechain outliers)

Ротамерами называются боковые цепи в предпочитаемых для данного остатка положениях, которые описываются наборами допустимых значений торсионных углов. Т.е. карта Рамачандрана позволяет найти маргиналов по торсионным углам в остове, ротамеры позволяют найти маргиналов по торсионным углам в боковых цепях аминокислот.

MolProbity [5] выделяет 16 маргинальных остатков (THR4, GLU72, GLU84, VAL219, ASP323, GLN348, ASP353 и др.), т.е. 2,15 %. Хорошими же считаются значения <0.3%. 64% PCA структур имеют значения хуже. Т.е. по данному критерию модель считать хорошей нельзя.

Доля маргиналов по RSR (RSRZ outliers)

Для расчета R-фактора можно использовать как отклонения теоретических структурных факторов от экспериментальных, так и отклонения электронной плотности модели от полученной синтезом Фурье из экспериментальных данных. Расчет R-фактора с использованием электронной плотности позволяет рассчитать его не только для всей модели, но и для ограниченного пространства, в частности вокруг одного остатка, что называется пространственным R-фактором (RSR). Для оценки этого фактора обычно используют Z-score (RSRZ) – на сколько среднеквадратических отклонений RSR остатка в данной модели отличается от среднего значения RSR по другим моделям похожего разрешения. Остаток считается маргинальным при RSRZ большем двух.

Для данной структуры 10 (т.е. 1,16%) остатков являются маргинальными по RSR (Рис. 5), это лучше, чем у 79% структур с близким разрешением (Рис. 1). По данному критерию модель можно считать хорошей.

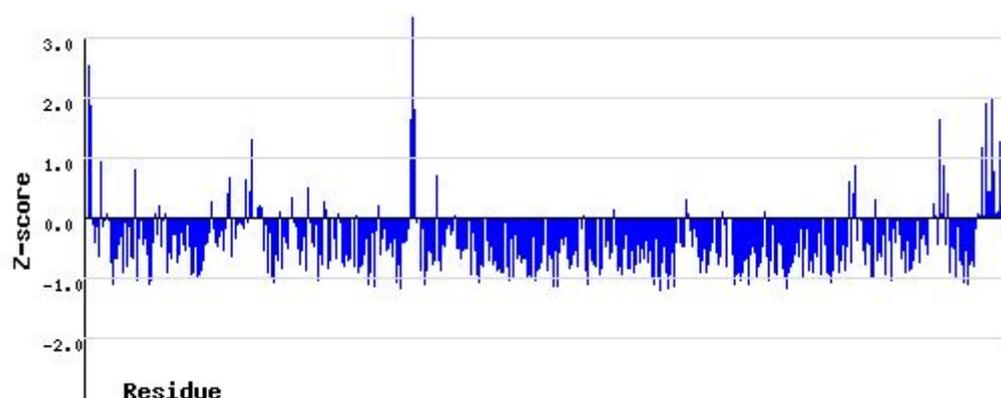


Рис. 5. RSR для обсуждаемой полипептидной цепи

Анализ маргинальных остатков

Таблица 1. Примеры маргинальных остатков (т.е. не все найденные) с указанием критерия качества, по которому они оказались отобранны

Аминокислотные остатки	Критерий качества
GLU711, GLU811, GLU294, SER54	по RSR, длины связей
ARG361, ASP215, LYS227, ASP165	по RSR, углы между связями
THR4, GLU72, GLU84, VAL219	по положению боковой цепи
ASN63, GLU260, LEU264, ALA331	по карте Рамачандрана, если смотреть результаты программы MOLEMAN2

1:A:136[A]:LYS:HG3 с
1:A:177[A]:SER:OG

по перекрываниям Ван-дер-Ваальсовых
радиусов атомов

GLU711

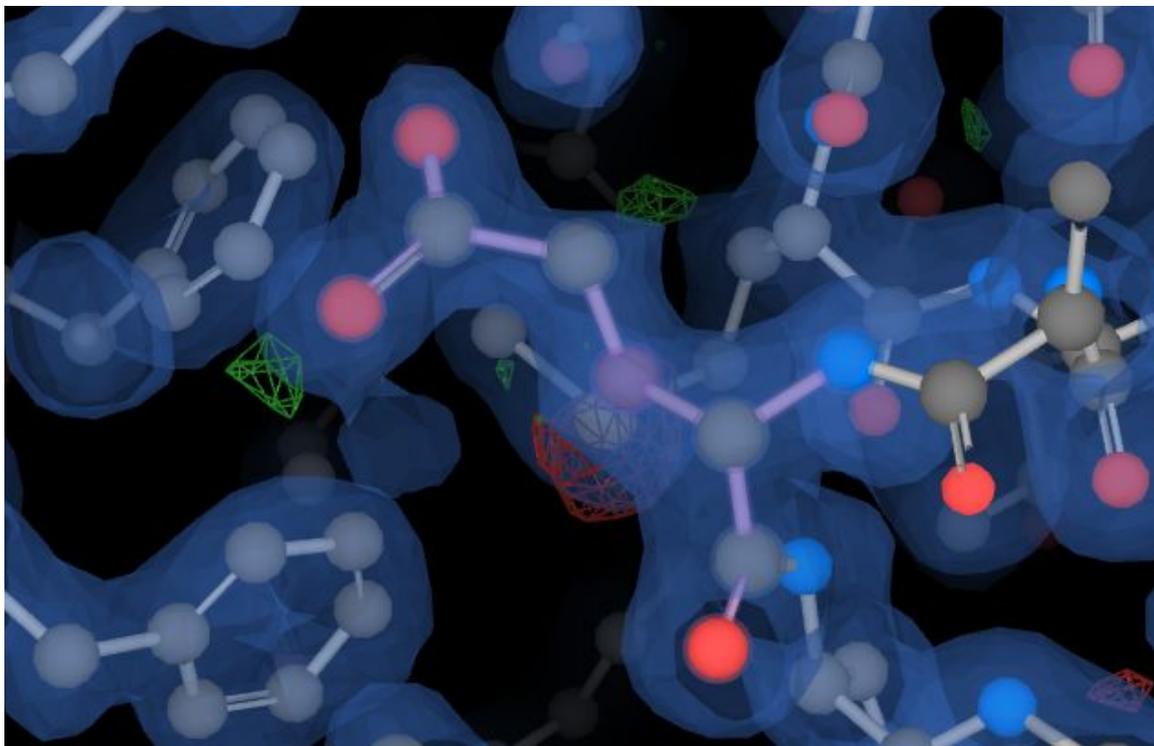


Рис. 6. Атомы остатка GLU711 с изображением поверхности экспериментальной электронной плотности. S выделен сиреневым, остальные C серые, O красные, а N синие.

Таблица 2. Расстояния между атомами GLU711

Атом 1	Атом 2	Расстояние
OE1	CD	1,36 Å
OE2	CD	1,26 Å
CD	CG	1,57 Å
CG	CB	1,53 Å
CB	CA	1,54 Å
CA	N	1,5 Å
CA	C	1,49 Å

Отклоняющимся от идеального значения (1,25 Å) является длина связи OE1-CD. Из химии известно, что длины связей OE2-CD и OE1-CD должны быть одинаковыми (резонансные

структуры), также GLU711 не находится в активном центре фермента, из чего логично предположить, что данный остаток расшифрован ошибочно.

ARG361

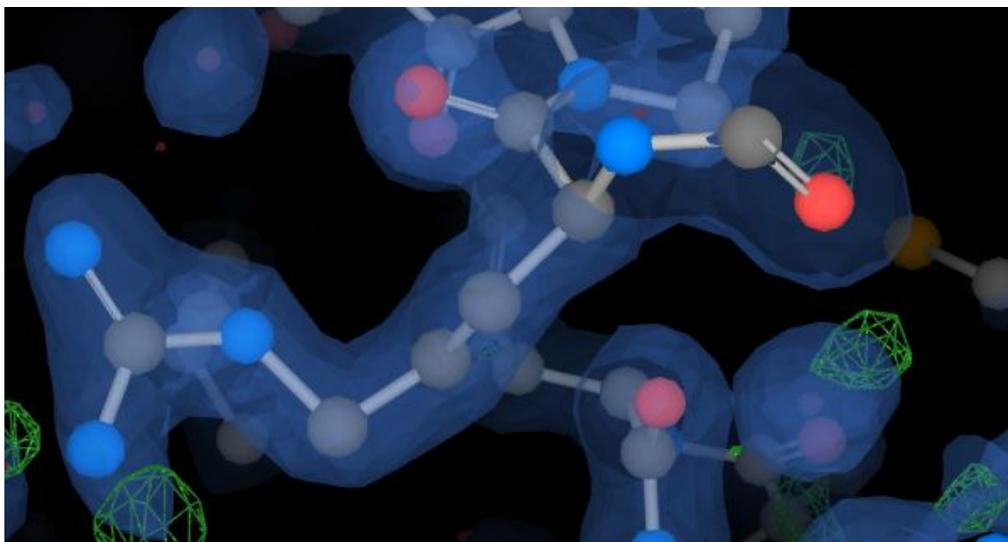


Рис. 7. Атомы остатка ARG361 с изображением предполагаемой моделью поверхности атомов. С серые, О красные, а N синие.

Угол между связями NE-CZ-NH₂ в рассматриваемом остатке равен 117.35, в отличие от идеального - 120.30. Этот остаток находится не в активном центре, а на поверхности фермента взаимодействует с раствором, который и может влиять на показатели, из-за чего логичнее предположить ошибку расшифровки, а не функциональную значимость. К тому же данный остаток находится в петле между α -спиралями. Остатки в петлях расположены достаточно свободно, из-за чего и возникают проблемы при их расшифровке.

THR4

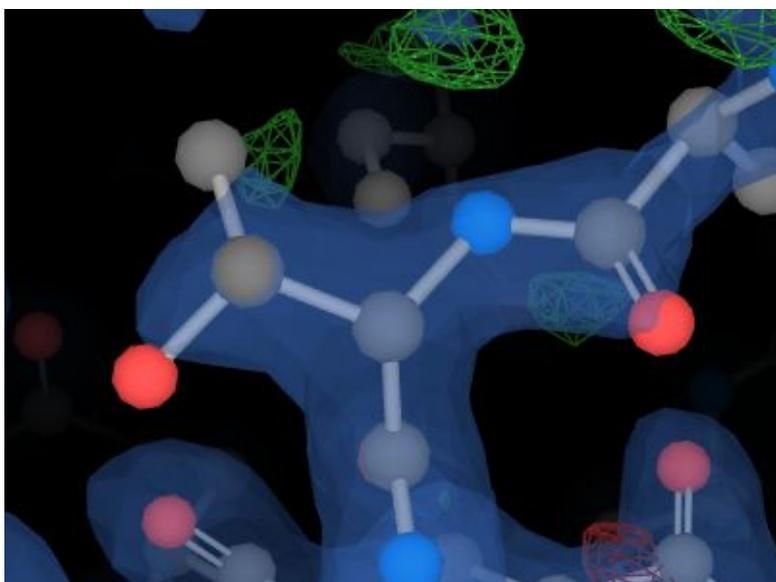


Рис. 8. Атомы остатка THR4 с изображением предполагаемой моделью поверхности атомов. С серые, О красные, а N синие.

Этот остаток находится в самом начале цепи, которая не обладает регулярной структурой, также он не близок к активному центру. Из Рис. 8 очевидно, что атомы остатка “не могут уместиться” в экспериментальную электронную плотность. Отсюда логично предположить ошибку расшифровки, а не функциональную значимость.

ASN63

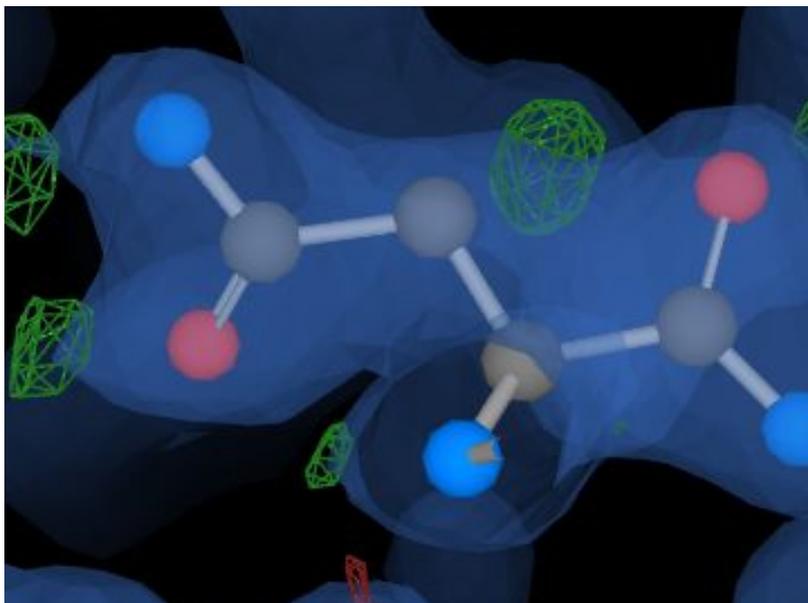


Рис. 9. Атомы остатка ASN63 с изображением предполагаемой моделью поверхности атомов. С серые, О красные, а N синие.

Торсионные углы рассматриваемого остатка по программе MOLEMAN2 : $\varphi = 37.9$, $\psi = 47.5$, т.е. данный остаток находится в области запрещенных конформаций. Однако по MolProbity его конформация среди разрешенных с немного другими значениями торсионных углов, и на Рис. 9 центры атомов адекватно вписываются в экспериментальную электронную плотность. Отсюда можно предположить ошибочность работы MOLEMAN2 и верную расшифровку программой MolProbity.

LYS136 и SER177

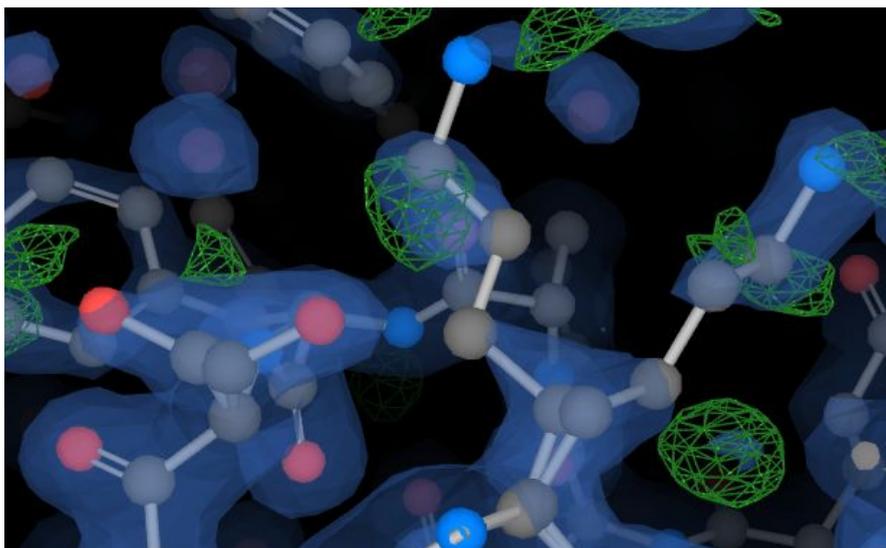


Рис. 10. Атомы остатков LYS136 и SER177 с изображением предполагаемой моделью поверхности атомов. С серые, О красные, а N синие.

Здесь очевидная ошибка расшифровки, на Рис. 10 видно, что и LYS136, и SER177 дублируются, и ни одно из повторений не покрывается хорошо экспериментальной электронной плотностью. Причиной такого может быть несколько молекул растворителя (воды) рядом с полипептидной цепью, которые из-за совсем близкого расположения были расшифрованы, как части аминокислотных остатков, коими на самом деле не являются.

Сравнение модели из PDB с моделью из PDB_redo

Таблица 3. Индикаторы качества моделей PDB и PDB-REDO по результатам работы программы MOLEMAN2 (красным выделены те индикаторы, которые лучше меньше, зеленым - больше, синим - должны быть меньше 1000)

	PDB	PDB-REDO
R-фактор	0.148	0.1568
Rfree	0.182	0.1890
Ramachandran plot appearance	0.257	0.521
Chi-1/Chi-2 rotamer normality	-0.783	-0.411
Backbone conformation	-0.262	-0.272
Bond length RMS Z-score	0.926	0.746
Bond angle RMS Z-score	0.989	0.883

Total number of bumps	36	19
Unsatisfied H-bond donors/acceptors	53	43

По всем параметрам, кроме R-фактора PDB-REDO лучше изначальной модели (PDB).

Можно предположить, что изначальная модель переобучена по R-фактору.

Заключение

Структура с PDB ID 4QUO расшифрована лучше среднего, однако с маргинальными остатками по любым из рассмотренных параметров стоит быть аккуратным, т.е. не говорить об их функциональной значимости, ведь все рассмотренные в работе маргинальные остатки оказались лишь плохо расшифрованы, особенно это относится к остаткам, находящимся на поверхности (близко к большому количеству молекул растворителя).

Список литературы

- 1) Luan, Y., & Xu, W. (2007). The Structure and Main Functions of Aminopeptidase N. *Current Medicinal Chemistry*, 14(6), 639–647. <https://doi.org/10.2174/092986707780059571>
- 2) Vassiliou, S., Węglarz-Tomczak, E., Berlicki, Ł., Pawełczak, M., Nocek, B., Mulligan, R., ... Mucha, A. (2014). Structure-Guided, Single-Point Modifications in the Phosphinic Dipeptide Structure Yield Highly Potent and Selective Inhibitors of Neutral Aminopeptidases. *Journal of Medicinal Chemistry*, 57(19), 8140–8151. <https://doi.org/10.1021/jm501071f>
- 3) Knapp JS, Rice RJ. *Neisseria* and *Branhamella*. In: Murray PR, Baron EJ, Pfaller MA, Tenover FC, Tenover RH. (ed.). *Manual of Clinical Microbiology*. 6th ed. American Society for Microbiology, Washington D. C, 1995.
- 4) xray.bmc.uu.se/usf/moleman2_man.html
- 5) <http://molprobit.biochem.duke.edu/>