

## Аннотация

В данном отчете представлен анализ качества расшифровки структуры оксидазы L-аминокислот (в дальнейшем LAAO) из яда змеи *Calloselasma rhodostoma* (Гладкий щитомордник) методом рентгеноструктурного анализа. Код модели этого белка в PDB 1F8R. Рассматриваются определенные показатели, такие как параметры кристаллической решетки, карта Рамачандрана, R-фактор, R<sub>free</sub> и другие, полученные с помощью различных баз данных и сервисов. Особое внимание уделено нескольким маргинальным аминокислотным остаткам.

## Введение

LAAO представляет собой димерный флавопротеин, содержащий FAD в качестве кофактора и обнаруживаемый у различных видов ядовитых змей.

LAAO катализирует стереоспецифическое окислительное дезаминирование субстрата L-аминокислоты до α-кетокислоты с образованием аммиака и перекиси водорода через промежуточное соединение иминокислоты.

Эти ферменты проявляют заметное предпочтение гидрофобных аминокислот, включая фенилаланин, триптофан, тирозин и лейцин.

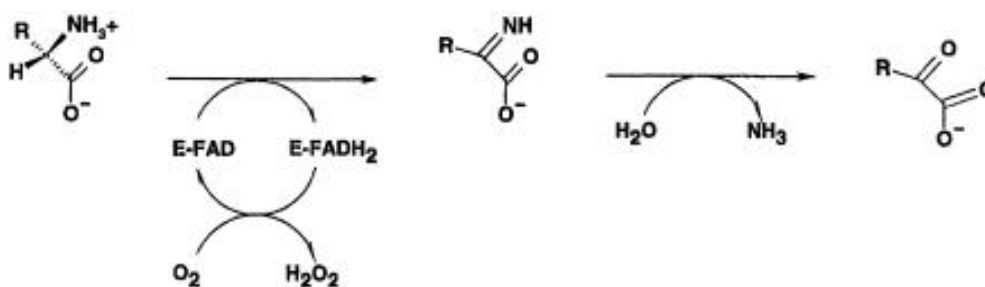


Рисунок 1. Катализируемая реакция

В яде фермент присутствует в высоких концентрациях и считается ответственным за токсическую активность. Хотя способ токсичности яда с LAAO не известен, было показано, что аналогичные ферменты из змей *Crotalus adamanteus* и *Crotalus atrox* могут специфически ассоциироваться с эндотелиальными клетками млекопитающих. После этого LAAO может вызывать апоптоз в этих клетках за счет выработки сильно локализованных концентраций перекиси водорода.

Ферменты LAAO, выделенные от ядов различных видов змей, являются наиболее изученными представителями этого семейства ферментов. [1]

# Результаты и обсуждения

## 1. Общая информация о модели

Состав комплекса: два тетрамера, каждый состоит из 489 аминокислот. Биологическая единица не совпадает с ассиметрической и является димером двух одинаковых цепей. Ассиметрическая ячейка включает 4 димера. Элементарная ячейка содержит 8 молекул.

Кристаллографическая группа: P 1 2<sub>1</sub> 1 (примитивная группа с поворотной осью 1, осью винтовой симметрии 2<sub>1</sub> и поворотной осью 1)

Дата размещения в PDB-банке: 27 августа 2000 года

Авторы: Pawelek PD, Cheah J., Coulombe R., Macheroux P., Ghisla S., Vrieland A.

Метод решения фазовой проблемы: изоморфное замещение, использовались тяжелые атомы ртути (K<sub>2</sub>HgCl<sub>2</sub>) и золота (KAuCl<sub>4</sub>) при различных значениях pH (4.5 и 6.0).

Подробности об экспериментах в таблице 1

Набор данных	pH	Разрешение, Å	Полнота, %	Все рефлексы	Независимые рефлексы	I/σ	R merge
Нативный белок 1	4.5	2.6	87.4	110458	49460	14.7	0.045
Нативный белок 2	4.5	2.0	88.1	391000	139242	18.9	0.046
K <sub>2</sub> HgCl <sub>2</sub>	4.5	2.8	82.9	76227	47166	15.4	0.054
K <sub>2</sub> HgCl <sub>2</sub>	6.0	2.8	88.8	104515	49741	12.0	0.066
KAuCl <sub>4</sub>	4.5	2.8	88.0	96712	50503	16.8	0.042
K <sub>2</sub> HgCl <sub>2</sub> /KAuCl <sub>4</sub>	6.0	2.8	87.3	105067	50092	24.1	0.048

Таблица 1. Статистика собранных данных. Проведенные эксперименты и измеренные в них рефлексы.

Пояснение: R<sub>merge</sub> – сумма всех интенсивностей.

Параметры кристаллографической ячейки:

Векторы: a = 79.496 Å, b = 154.962 Å, c = 102.947 Å

Углы между векторами: α = 90°, β = 109.49°, γ = 90°

Разрешение структуры: 2.0 Å, диапазон 2.0 – 50.0 Å

Некристаллографическая симметрия: за счет образования комплексов с молекулами FAD и АВ (o-аминобензоат); двукратная ось вращения приводит к расположению изоаллоксазиновых колец двух молекул FAD вдоль одной плоскости, а остальные части протетической группы проходят в противоположных направлениях.

Вся информация получена из источников [1] и [2].

## 2. Значения индикаторов качества модели

### R-фактор:

Стандартный фактор достоверности характеризует, насколько полученная модель соответствует экспериментальным данным, в ходе оптимизации минимизируется.

$$R_{x\text{-ray}} = \sum_{h,k,l} (F_{\text{рассчитанный}} - F_{\text{наблюдаемый}})^2 \Rightarrow \min$$

$$R = \frac{\sum_{h,k,l} |F_{\text{рассчитанный}} - F_{\text{наблюдаемый}}|}{\sum_{h,k,l} F_{\text{наблюдаемый}}}$$

Пояснения к формулам:

$h, k, l$  – индексы Миллера (всегда целые числа, известны из оцифровки эксперимента)

$F$  – модуль структурного фактора ( $F_{\text{наблюдаемый}}$  известен из эксперимента,  $F_{\text{рассчитанный}}$  появляется при оптимизации)

По сути R-фактор является интегральной мерой различия  $F_{\text{рассчитанного}}$  и  $F_{\text{наблюдаемого}}$ . Если получается меньше 0.25, то модель считается хорошей.

$R_{x\text{-ray}}=0.185$  (для комплекса с цитратом),  $R_{x\text{-ray}}=0.205$  (для комплекса с АВ), то есть обе модели можно считать хорошими и достоверными.

При подгонке модели существует риск переоптимизации. Чтобы учесть его, используется фактор  $R_{\text{free}}$ , который вычисляется как общий R-фактор, но по случайной выборке из относительно небольшого числа случайных структурных факторов. В моем случае использовали 10% от общего числа.

$R_{\text{free}}=0.210$  (для комплекса с цитратом),  $R_{\text{free}}=0.225$  (для комплекса с АВ). Чтобы модель считалась хорошей,  $R_{\text{free}}$  должен быть близким к  $R_{\text{общ}}$  (а точнее отличаться менее чем на 10%). В моем случае эта разница составляет 1.14 % ( $R_{\text{общ}}= 0.185$ ,  $R_{\text{free}}= 0.21$ ), то есть модель не может считаться переоптимизированной. [1], [2]

## Карта Рамачандрана

С помощью сервиса MolProbity [3] была получена карта Рамачандрана, показывающая положение пар торсионных углов  $\phi/\psi$  полипептидной цепи.

1f8r.pdb, model 1

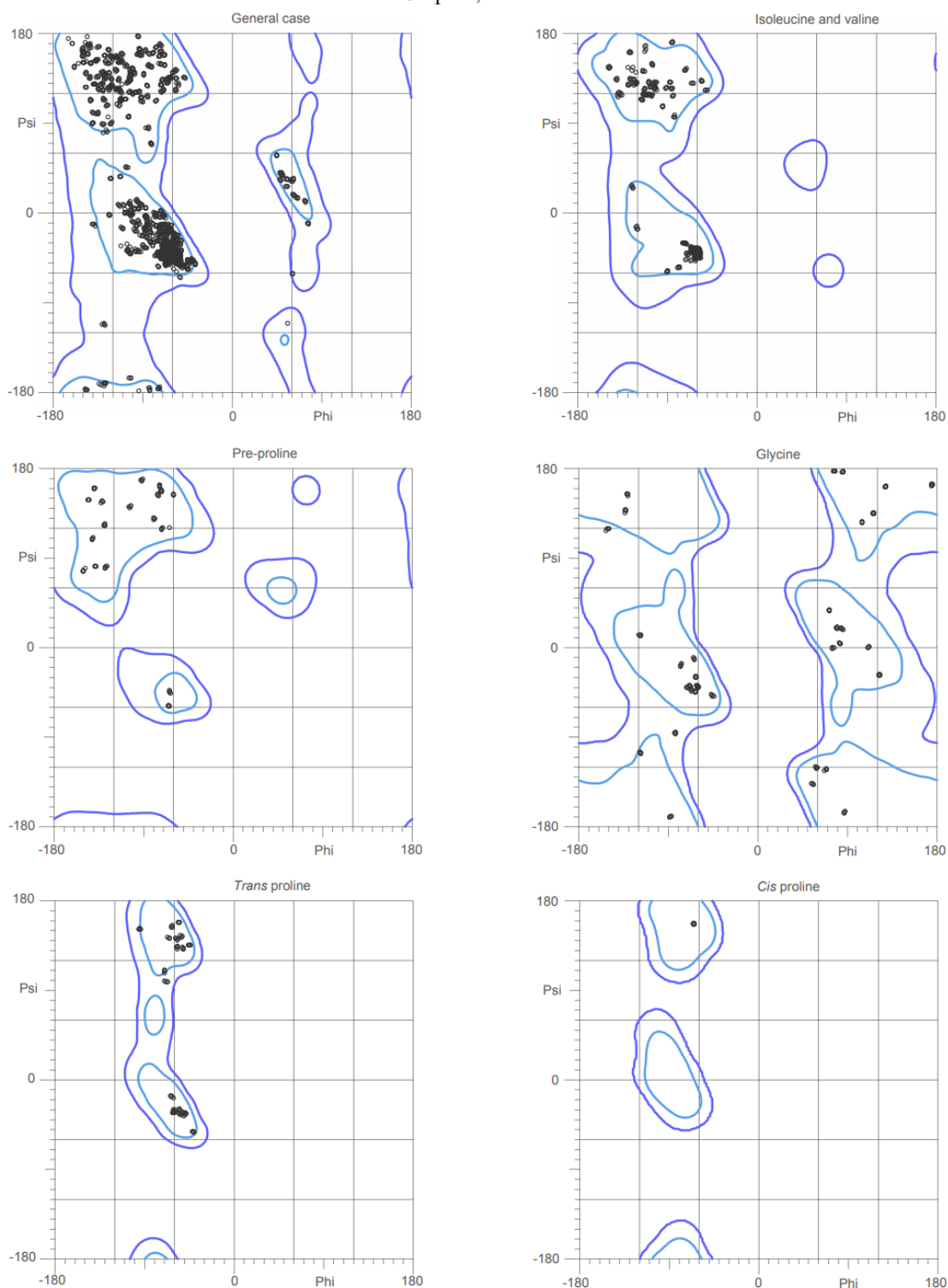


Рисунок 2. Карты Рамачандрана для разных типов остатков модели 1F8R: всех остатков, кроме перечисленных далее; изолейцина и валина; остатков перед пролином; глицина; транс-пролина; цис-пролина. Голубым выделены предпочтительные области значений углов  $\phi$  и  $\psi$ , синим – разрешённые

Согласно сервису MolProbity, 97.3% аминокислотных остатков (1872 из 1924) лежат в предпочтительных областях, ни одного выброса обнаружено не было. Довольно хороший результат.

#### Ротамеры:

Ротамеры – это боковые цепи, находящиеся в типичной для данного типа остатка конформациях.

Сервис MolProbity нашел 41 аминокислотный остаток с информацией боковой цепи, который не соответствует ни одному из ротамеров. Процент таких неправильных остатков 2.54% сильно меньше допустимого порога (0.3%). [3]

#### Clashscore:

Clashscore показывает, сколько недопустимых стерических перекрытий ( $> 0.4 \text{ \AA}$ ) наблюдается в структуре на 1000 атомов.

Для моей структуры этот коэффициент составил 8.31. Это соответствует 89 перцентили (если проще, то процентам) и считается хорошим значением. [3]

#### С $\beta$ -отклонение:

С $\beta$ -отклонение является оценкой геометрии основной цепи белка. Зная геометрию основных атомов остова, можно вычислить идеальное положение атома  $\beta$ -углерода. Большое отклонение от идеальной позиции указывает на ошибку геометрии.

С $\beta$ -отклонений больше  $0.25 \text{ \AA}$  не обнаружено, это хороший показатель. [3]

#### Цис-пролины:

Пептидная связь, соединяющая соседние аминокислоты, имеет характер частичной двойной связи. Поэтому она предполагает либо плоскую транс-, либо плоскую цис-конформацию. Для пролина транс-конформация является более предпочтительной. Обычно только 5% пролинов в полипептидной цепочке находятся в цис-конформации.

В моей модели (1f8r) присутствуют 80 пролинов, из них 4 находятся в цис-конформации, то есть те самые 5%. Вообще модель включает в себя 4 одинаковые цепочки и MolProbity специально подчеркивает, что на одну молекулу белка приходится 1 цис-пролин. [3]

Clashscore (для всех атомов)	8.31 (N=715, 2.00Å ± 0.25Å)
Плохие ротамеры	41 (2.54%)
Выбросы на карте Рамачандрана	0
А. о. в предпочитаемой области	1872 (97.3%)
Сβ-отклонения (> 0.25 Å)	0
Неоптимальные длины связей остова	45/16166 (0.28%)
Неоптимальные углы остова	30/21962 (0.14%)

Таблица 2. Статистика по качеству структуры 1f8r из сервиса MolProbity. Цвета соответствуют его оценке: **зеленый** – хорошо, **желтый** – средне, **красный** – не очень.

### 3. Некоторые маргинальные остатки

Ниже представлена таблица из 15 аминокислотных остатков, информация взята с сервера MolProbity [3].

№	Остаток	Критерии
1	PRO 306 (A)	Пролин в цис-конформации
2	ASP 106 (B)	Плохой ротамер
3	THR 363 (D)	Плохой ротамер
4	GLU 66 (A)	Плохой ротамер
5	GLY 429 (D)	Плохой ротамер
6	GLN 269 (D)	Плохой ротамер
7	ASN 305 (A)	Плохой ротамер
8	VAL 285 (A)	Плохой ротамер
10	GLU 9 (A)	Плохой ротамер
11	FAD 547 (B)	Лиганд, странная геометрия остова (длина связей, углы)
12	CIT 526 (D)	Лиганд, странная геометрия остова (длина связей, углы), очень высокий В-фактор (84.05)
13	ALA 41 (D)	Находится не в предпочитаемой зоне на карте Рамачандрана
14	CYS 412 (D)	Находится не в предпочитаемой зоне на карте Рамачандрана
15	SER 279 (A)	Находится не в предпочитаемой зоне на карте Рамачандрана

Таблица 3. Список маргинальных остатков с указанием номера в цепи и критерия, по которому они были отобраны.

### 4. Детальный анализ 5 маргинальных остатков

Пролин 306 (цепь А), находится в цис-конформации

Находится в разрешенной области на карте Рамачандрана ( $\phi = -65.1$ ,  $\psi = 158.6$ ), допустимый ротамер (значения  $\chi$ -углов 30.4, 317.4, 37.7), В-фактор = 26.02 (в пределах нормы), Сβ-отклонение составляет 0.05Å. В общем вся его необычность заключается в том, что он находится в цис-конформации. Визуализировав его окружение, я обнаружил, что он соединен с другим пролином.

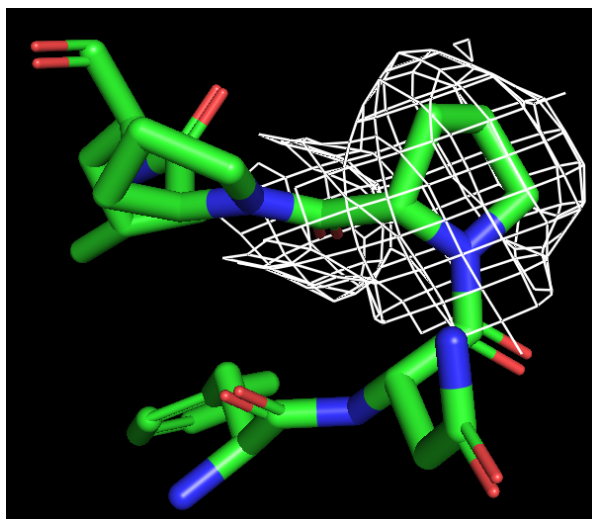


Рисунок 3. PRO306(цепь A) и его электронная плотность на уровне подрезки  $\sigma=1$ . Хорошо вписывается в электронную плотность.

Вывод: это не ошибка и даже не маргинальность, просто особенность связи.

Аспарагин 106 (цепь B), плохой ротамер (классифицируется как выброс)

Расположен прямо перед  $\alpha$ -спиралью на поверхности белковой глобулы, имеет довольно высокий B-фактор =41.14, по карте Рамачандрана находится в разрешенной области ( $\phi = 61.5$ ,  $\psi = 32.5$ ),  $C\beta$ -отклонение очень мало (0.02 Å) Но маргинальный ротамер, значения  $\chi$ -углов 226.2 и 281.

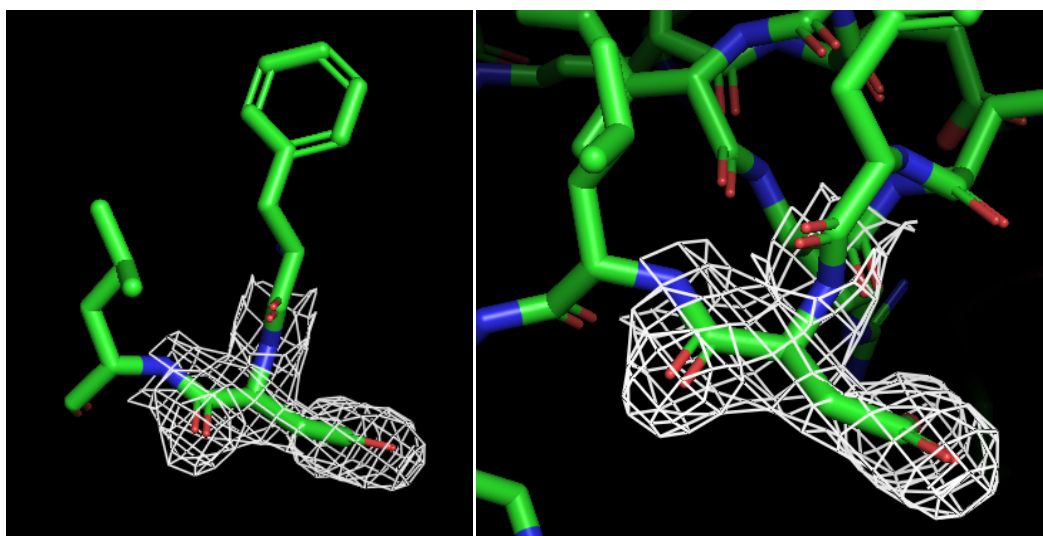


Рисунок 4. Окружение ASP106 и его электронная плотность на уровне подрезки  $\sigma=1$ . Видно, что он хорошо вписывается в электронную плотность и везде имеет правильную конформацию. Скорее всего, на него влияет атом кислорода, расположенный под аспарагином в глубине молекулы белка.

Вывод: необычность конформации можно объяснить электростатическим влиянием.

Аланин 21 (цепь D), необычное, но допустимое положение на карте Рамачандрана

Как я уже писал, MolProbity не нашел остатки из запрещенной области карты Рамачандрана, но некоторые остатки он выделил как странные, но вполне допустимые. Я решил рассмотреть один из них.

B-фактор = 16.91, положение на карте Рамачандрана  $\phi = 105.7$ ,  $\psi = 46.6$ . Ротамер не определен, как и для его соседей по цепи, GLY40 и GLY42. То есть он окружен глицинами, а связи между ними самые подвижные и свободные. Наверное, в этом всё и дело.

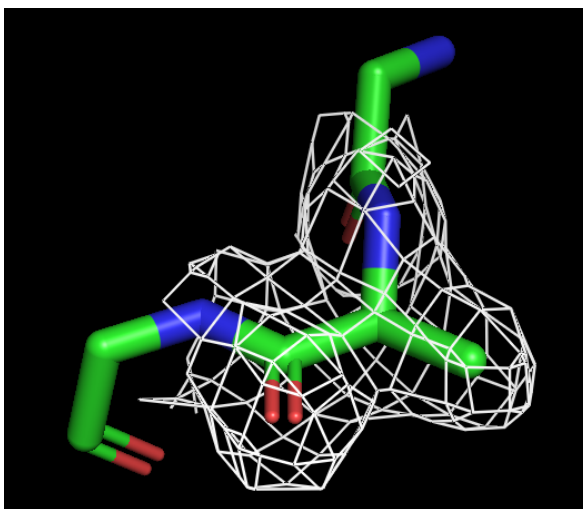


Рисунок 5. Окружение ALA41 и его электронная плотность на уровне подрезки  $\sigma=1$

Видно, что остаток отлично вписывается в электронную плотность. Судя по всему, довольно подвижен, находится в незакрепленной петле.

Вывод: всё нормально, просто этот остаток окружен глицинами, а у них набор разрешенных конформаций больше.

Треонин 363 (цепь B), плохой ротамер и приемлемое положение на карте Рамачандрана

Плохой ротамер,  $\chi$ -угол = 19.2; на карте Рамачандрана располагается в допустимой области карты Рамачандрана ( $\phi = -37.4$ ,  $\psi = -50.4$ ), однако MolProbity не считает это положение правильным. Довольно высокий B-фактор = 41.39, зато C $\beta$ -отклонение = 0.06Å (мало).



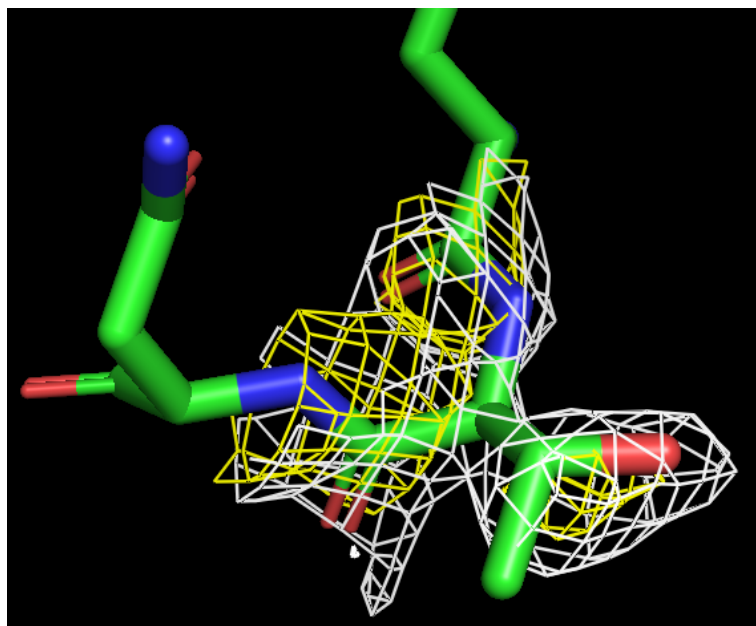


Рисунок 6. THR363 и его электронная плотность при  $\sigma = 0.5$  (желтый цвет) и  $\sigma = 1.0$  (белый цвет). По видимому, здесь экспериментаторы получили не смогли точно вписать в электронную плотность молекулу аминокислотного остатка.

Вывод: маргинальность связана с ошибкой расшифровки, неправильно вписали аминокислоту.

Флавин-аденин динуклеотид (связан с цепью С), аномальные углы и длины связей

Это лиганд, находящийся в активном центре. По понятным причинам для него не были рассчитаны обычные показатели маргинальности, кроме В-фактора = 18.86.

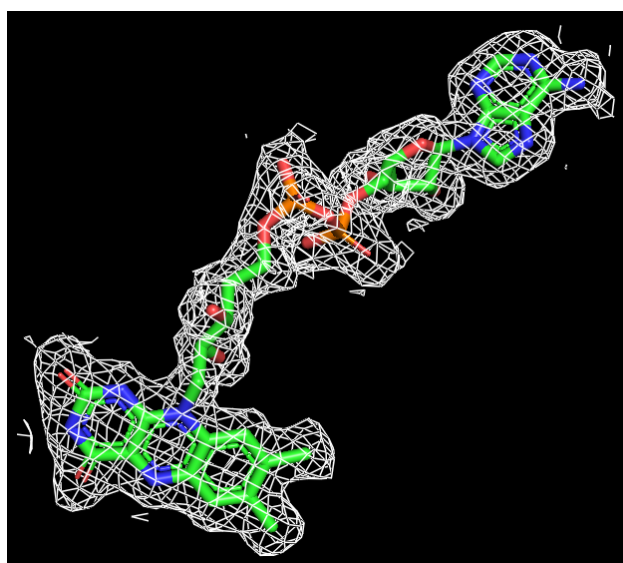


Рисунок 7. FAD в активном центре хорошо вписался в электронную плотность ( $\sigma = 0.5$ ).

Вывод: эта молекула – небелковый кофактор, следовательно, его маргинальность в виде сильного отклонения связей и углов на 12.4  $\sigma$  и 5.8  $\sigma$  соответственно объясняется их непохожестью на обычные, встречающиеся в белках.

## 5. Сравнение модели из PDB с моделью из PDD REDO

На сервере PDD REDO [4] была построена структура 1f8r\_final по имеющимся экспериментальным данным о 1f8r.

	<b>PDB</b>	<b>RMS Z-score</b>
R	0,1798	0,1382
R_free	0,2284	0.1684
RMS Z-score по связям	0.28	0.582
RMS Z-score по углам	0.602	0.75
Хорошие положения на карте Рамачандрана	40	74
Нормальные ротамеры	56	83
Удовлетворительные водородные связи	50	54

Таблица 4. Некоторые характеристики PDB и PDB\_final. Улучшившиеся помечены зеленым цветом

Видно, что модель PDB\_final стала менее переоптимизированной, появилось больше хороших ротамеров и остатков в предпочитаемой области на карте Рамачандрана. Немного возросло число хороших водородных связей. А вот Z-score по длинам связей и углам почти не изменился.

<b>Измененные ротамеры</b>	<b>52</b>
<b>Перевернутые боковые цепи</b>	<b>41</b>
<b>Перемещенные молекулы воды</b>	<b>88</b>
<b>Перевернутые пептидные связи</b>	<b>0</b>
<b>Исправления хиральности</b>	<b>0</b>

Таблица 5. Изменения в модели PDB\_final

На основе имеющихся данных можно сделать вывод, что PDB REDO улучшил структуру 1f8r, устранив часть аномалий.

## Вывод

Разрешение модели 1F8R составляет 2 Å, что считается хорошим результатом. На мой взгляд качество модели можно оценить как высокое, на что указывают её характеристики. Присутствуют маргинальные аминокислотные остатки, но глобальных ошибок мне не удалось найти.

Цель работы достигнута – идентифицирован активный центр фермента, получен комплекс фермент-лиганды (FAD, NAG, FUC, CIT).

## Использованная литература

- 1) *The structure of L-amino acid oxidase reveals the substrate trajectory into an enantiomerically conserved active site.* **Peter D. Pawelek, Jaime Cheah, Rene Coulombe, Peter Macheroux, Sandro Ghisla and Alice Vrielink.** 2000. б.м. : EMBO J, 2000 г.
- 2) База данных PDB, ссылка на структуру <https://www.ebi.ac.uk/pdbe/entry/pdb/1f8r>
- 3) Сервер MolProbity, <http://molprobity.biochem.duke.edu/>
- 4) Сервер PDB REDO, <https://pdb-redo.eu/>