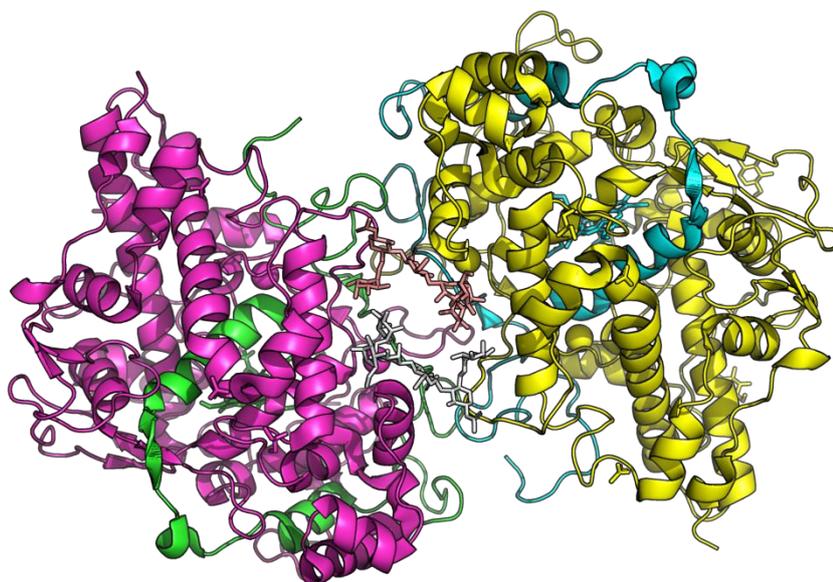


# Отчет по валидации.

## Введение.

Миелопероксидаза – это фермент лизосом нейтрофилов, производящий гипогалогеновые кислоты, которые обладают бактерицидным действием. Гиперактивность миелопероксидазы связывают с развитием хронического воспаления, поэтому поиск ингибиторов миелопероксидазы – очень важная задача в борьбе со многими заболеваниями, связанными с окислительным повреждением тканей. В данной работе рассматривается структура миелопероксидазы (PDB ID: 4C1M) в комплексе с потенциальным ингибитором – ароматическим гидроксаматом, в данной работе сокращенно называемым НХ1.



**Рисунок 1.** Структура миелопероксидазы (PDB ID: 4C1M) с раскраской по цепям. Зеленым и голубым обозначены легкие цепи, розовым и желтым – тяжелые.

Миелопероксидаза представляет собой димер, состоящий из 2 идентичных половин, каждая из которых состоит из 2 полипептидных цепей длиной 108 и 466 аминокислотных остатков (называемые, соответственно, «легкими» и «тяжелыми» цепями). Каждая половина молекулы содержит ковалентно связанный с ней гем. Две половины димера связаны дисульфидной связью через Cys153. Вторичная структура состоит преимущественно из альфа-спиралей с небольшой долей бета-листов.

## Общая информация о модели.

Структура была расшифрована в 2013<sup>1</sup> и демонстрирует миелопероксидазу в комплексе с её обратимым ингибитором НХ1.

**Таблица 1.** Значения некоторых параметров структуры 4C1M (взяты из PDB<sup>2</sup>).

Число измеренных рефлексов	81440	
Метод решения фазовой проблемы	Молекулярное замещение	
Параметры кристаллографической решетки	Длина ( Å )	Угол ( ° )
	a = 111.279	$\alpha = 90$
	b = 63.44	$\beta = 97.36$
	c = 92.381	$\gamma = 90$
Кристаллографическая группа	P 1 21 1	
Наименьшее разрешение	2 Å	
Наибольшее разрешение	55 Å	
Полнота данных	94 %	

Некристаллографическая симметрия в асимметрической ячейке отсутствует.

## Значения индикаторов качества модели в целом.

Основным показателем качества модели являются R-фактор и  $R_{free}$ .

R-фактор характеризует соответствие модели экспериментальным данным и вычисляется по оптимизированной модели и непосредственно измеренным структурным факторам по формуле:

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

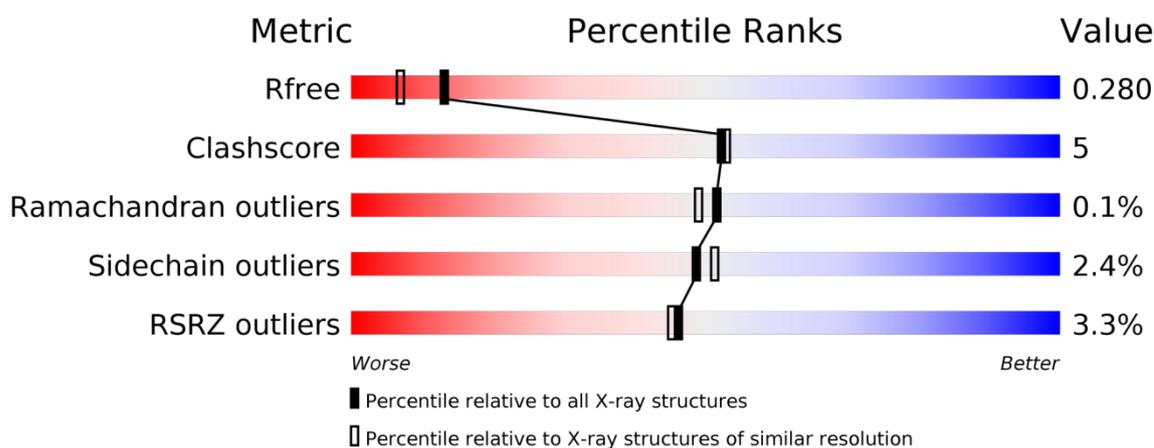
$R_{free}$  вычисляется по той же формуле, что и R-фактор, но набор рефлексов предварительно разделяется на рабочий и контрольный. Это делается для того, чтобы избежать переоптимизации и не «подогнать» совсем уж неправильную модель. Считается, что значение R-фактора должно быть меньше 20% в идеальном случае. Значение R-фактора больше 40% считаются плохим показателем. Если модель правильная, то значение R-фактора и  $R_{free}$  будут близки по значениям. Разница между R-фактором и  $R_{free}$  превышающая 10% может служить показателем переоптимизации.

В PDB значения R-фактора и  $R_{free}$  составляют:

**R-фактор:** 0.222 (22.2%)

**$R_{free}$ :** 0.278 (27.8%)

Значения R-фактора и  $R_{free}$  неидеальны, но не превышают 40%. Разница между R-фактором и  $R_{free}$  составляет 5,6%, что меньше 10%, а значит, переоптимизация вряд ли имеет место быть.



**Рисунок 2.** Показатели качества модели 4C1M по данным PDB.

По данным PDB validation report:

- $R_{free}$ , составляющий 22.2%, в целом хуже, чем у моделей с похожим разрешением.
- Данная структура содержит 1 остаток из запрещенной области карты Рамачандрана, что составляет 0.1% от общего числа аминокислотных остатков.

- Процент остатков с маргинальными боковыми цепями составляет 2,4%
- Процент остатков, маргинальных по значению RSRZ составляет 3,3%

Значения почти всех 5 параметров (за исключением  $R_{free}$ ), по-видимому, являются довольно средними как в целом для структур, полученных методом РСА, так и для структур с похожим разрешением.

Сервер MolProbity<sup>3</sup> позволил нам получить значения некоторых индикаторов качества структуры, среди которых нас в первую очередь интересуют:

1. “Clash score” – число недопустимых наложений атомов на 1000, а также перцентиль данной структуры по отношению к структурам примерно такого же разрешения.
2. “Poor rotamers” – число остатков с маргинальными боковыми цепями.
3. “Ramachandran outliers” – полные маргиналы по карте Рамачандрана, лежат вне допустимой области.
4. “Ramachandran favored” – число и процент остатков в предпочитаемой области (в идеале таких должно быть > 98%).

All-Atom Contacts	Clashscore, all atoms:	5.47		97 <sup>th</sup> percentile* (N=715, 2.00Å ± 0.25Å)
	Clashscore is the number of serious steric overlaps (> 0.4 Å) per 1000 atoms.			
Protein Geometry	Poor rotamers	19	1.91%	Goal: <0.3%
	Favored rotamers	923	92.58%	Goal: >98%
	Ramachandran outliers	0	0.00%	Goal: <0.05%
	Ramachandran favored	1091	97.15%	Goal: >98%
	Rama distribution Z-score	-1.05 ± 0.23		Goal: abs(Z score) < 2
	MolProbity score <sup>^</sup>	1.66		92 <sup>nd</sup> percentile* (N=12522, 2.00Å ± 0.25Å)
	Cβ deviations >0.25Å	2	0.19%	Goal: 0
	Bad bonds:	17 / 9530	0.18%	Goal: 0%
Bad angles:	8 / 12963	0.06%	Goal: <0.1%	
Peptide Omegas	Cis Prolines:	6 / 91	6.59%	Expected: ≤1 per chain, or ≤5%
	Cis nonProlines:	2 / 1042	0.19%	Goal: <0.05%
Additional validations	Chiral volume outliers	0/1425		
	Waters with clashes	27/664	4.07%	See UnDowser table for details

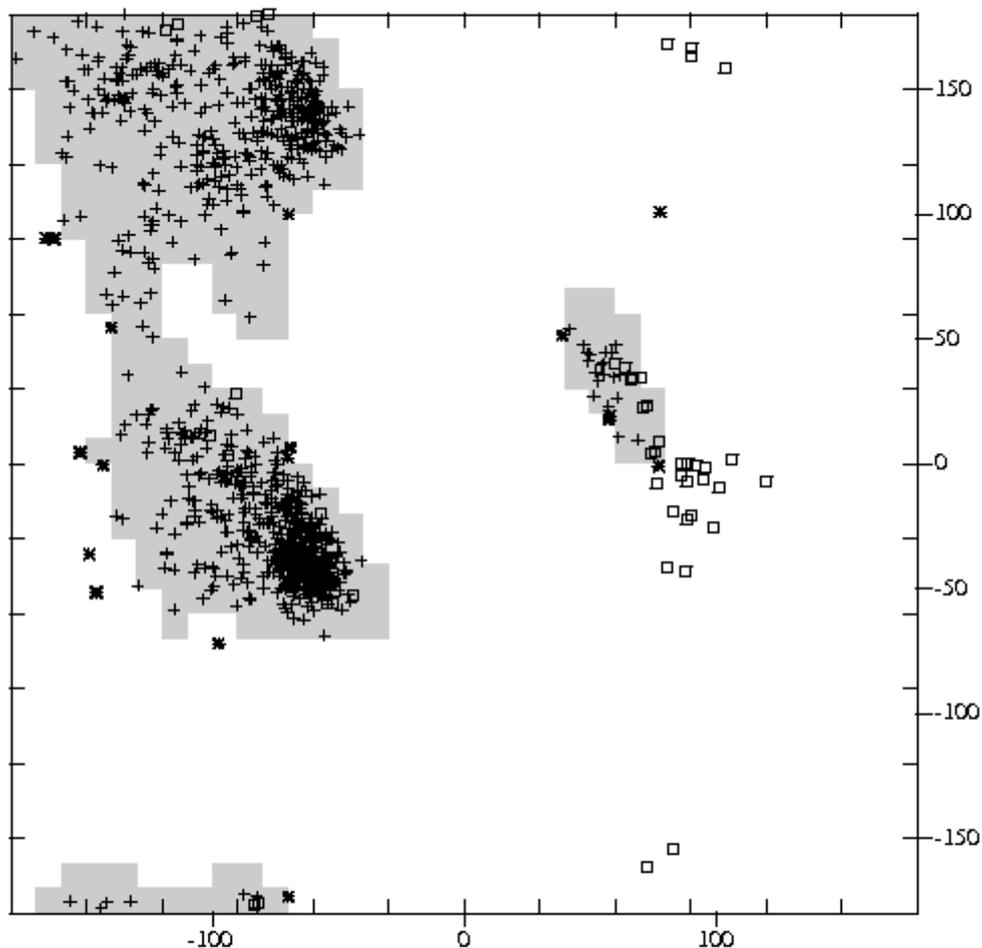
**Рисунок 3.** Показатели качества модели по данным сервера MolProbity.

Как видно из рисунка 3, перцентиль данной структуры по параметру «clashscore» составляет 97% - случай, приближающийся к идеальному (то есть, у 97% структур примерно такого же разрешения ситуация с перекрываниями Ван-дер-Ваальсовых радиусов обстоит хуже).

Число остатков с маргинальными боковыми цепями превышает желаемое значение почти в 6 раз и составляет 1,91%. В целом, при просмотре сводной таблицы по всем остаткам можно заметить, что там действительно наблюдается много отклонений маргиналов по значениям торсионных углов.

Число остатков, лежащих в предпочитаемой области карты Рамачандрана, составляет 97,15%, что немного меньше идеального случая, но близко к нему. Сервер MolProbity не обнаружил остатков из запрещенной области карты Рамачандрана (что противоречит PDB validation report, в котором утверждается, что в структуре имеется один такой остаток), для их определения был использован сервис EDS<sup>4</sup>.

┌



X = PHI mapped to [-180,180>

Y = PSI mapped to [-180,180>

PDB file : /pub/db/eds/sfd/4c1m/pdb4c1m.ent

Glycines (open squares): 68 ; Start/end residues : 12

D-amino acids : 0 ; Residues with missing atoms : 4

Residues in Ramachandran plot checked : 1055 out of 1139

In core regions (plus signs): 1038 ; Outliers (asterisks): 17

Percentage outliers: 1.6

An average <= 2.0 Å model has ~0.5% outliers

See: Kleywegt, G.J. and Jones, T.A. (1996). Structure 4, 1395-1400.

└

**Рисунок 4.** Карта Рамачандрана, построенная сервисом EDS.

Сервис EDS нашёл 17 остатков из запрещенной области карты Рамачандрана, что составляет 1,6% от общего числа аминокислотных остатков и превышает желаемый процент (0,5% или меньше) в 3 раза.

Помимо уже перечисленных показателей качества отдельных остатков, используются также комфортность окружения и RSR-фактор. RSR-фактор сравнивает соответствие электронной плотности модели с экспериментальной электронной плотностью: то есть, фактически, служит показателем того, насколько хорошо остаток «вписан» в электронную плотность. В качестве относительной оценки RSR-фактора для конкретного остатка используют RSR z-score (RSRZ), где RSR сравнивается со средним RSR для того же типа остатков по выборке из PDB с похожим разрешением. Хорошие значения RSRZ должны быть меньше 2. Для определения значений RSRZ для каждого остатка использовался сервис EDS.

Нужно также отметить, что EDS показал, что доля остатков с  $RSRZ > 2$  составляет 2,55%, что отличается от заявленного в PDB validation report.

## Маргинальные остатки

В таблице 2 приведено 14 маргинальных остатков с указанием критерием, по которому они были отобраны как маргинальные. Всего при поиске маргинальных остатков было обнаружено 9 недопустимых ротамеров, 5 остатков с возможной инверсией цепи, 17 остатков из запрещенной области карты Рамачандрана (по версии сервиса EDS) и 29 остатков с  $RSRZ > 2$ .

**Таблица 2.** Маргинальные остатки.

Остаток	Критерий маргинальности
Цепь В 6 LYS	<ul style="list-style-type: none"><li>• Недопустимый ротамер</li><li>• Перекрывание 0.49Å Н и В 6 LYS HE2</li></ul>
Цепь С 113 VAL	<ul style="list-style-type: none"><li>• Высокий z-score (4.66)</li></ul>
Цепь С 142 ASP	<ul style="list-style-type: none"><li>• Запрещенная область карты Рамачандрана (по версии EDS)</li><li>• Слегка повышенный RSRZ (2.17)</li><li>• Перекрывание 0.46Å HB3 и C2038 НОН О</li></ul>
Цепь С 200 ASN	<ul style="list-style-type: none"><li>• Почти наверняка должна быть инверсия</li></ul>
Цепь С 353 MET	<ul style="list-style-type: none"><li>• Запрещенная область карты Рамачандрана (по версии EDS)</li><li>• Недопустимый ротамер</li></ul>
Цепь С 355 PRO	<ul style="list-style-type: none"><li>• Цис-конформация</li><li>• высокий RSRZ (3.56)</li></ul>
Цепь С 421 ASN	<ul style="list-style-type: none"><li>• Перекрывание 0.65Å HD21 и С 336 HIS HD1</li><li>• Подозрение на инверсию боковой цепи</li></ul>
Цепь С 550 ASN	<ul style="list-style-type: none"><li>• Цис-конформация</li></ul>
Цепь D 181 GLU	<ul style="list-style-type: none"><li>• Перекрывание 0.44Å HG3 и B2019 НОН О</li><li>• Недопустимый ротамер</li></ul>
Цепь D 217 HIS	<ul style="list-style-type: none"><li>• Повышенный RSRZ (2.73)</li><li>• Недопустимый ротамер</li></ul>
Цепь D 269 PRO	<ul style="list-style-type: none"><li>• Высокий RSRZ (3.61)</li></ul>
Цепь D 314 ARG	<ul style="list-style-type: none"><li>• Запрещенная область карты Рамачандрана</li><li>• Недопустимый ротамер</li></ul>
Цепь D 318 ASP	<ul style="list-style-type: none"><li>• Запрещенная область карты Рамачандрана</li><li>• Недопустимый ротамер</li></ul>
Цепь D 569 ALA	<ul style="list-style-type: none"><li>• Высокий RSRZ (3.48)</li></ul>

## Анализ 5 маргинальных остатков.

### Остаток 1. ASP 142, цепь С.

Данный остаток является маргинальным по трём признакам: он попадает в запрещенную область карты Рамачандрана (по данным EDS), имеет повышенный RSRZ (не очень высокий, но  $>2$ ) и перекрытие  $0.46\text{\AA}$  HB3 с C2038 НОН О.

Повышенный RSRZ обычно говорит о том, что остаток не слишком хорошо вписывается в свою электронную плотность. В данной ситуации RSRZ составляет 2.17: показатель повышен, но не слишком. Посмотрим на электронную плотность.

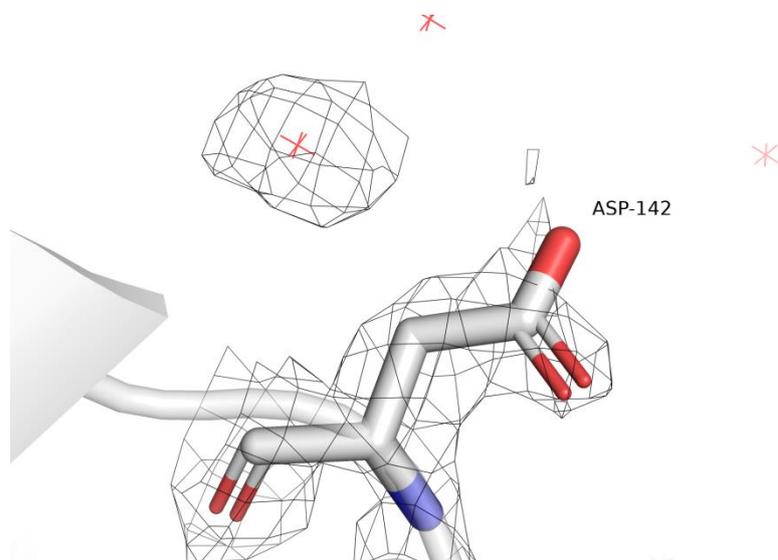
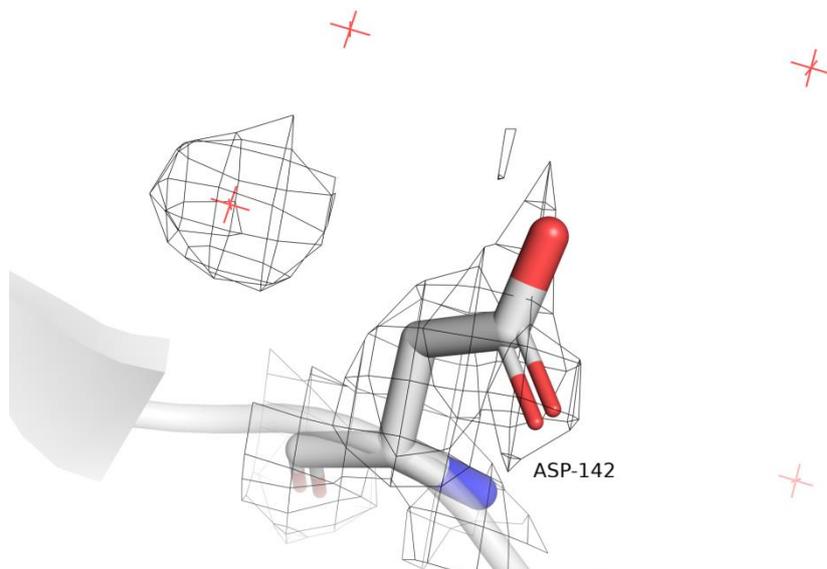


Рисунок 5. Электронная плотность остатка ASP142. Уровень подрезки 1.



**Рисунок 6.** Электронная плотность остатка ASP 142. Уровень подрезки 1.

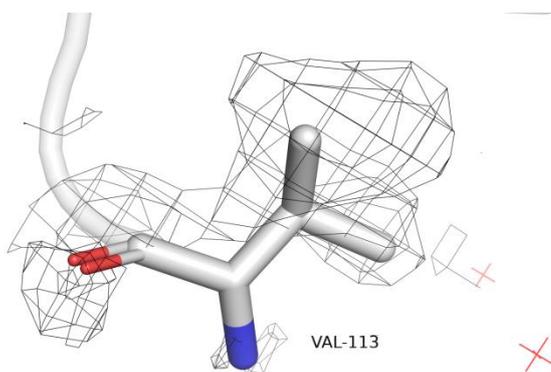
Электронная плотность остатка ASP142 не очень хорошая: вокруг кислорода карбонильной группы будто бы «откушен» кусок электронной плотности, что, скорее всего, связано с его расположением на периферии и, соответственно, повышенной подвижностью (данный остаток имеет B-фактор = 22.5). Однако в остальном остаток нормально вписывается в электронную плотность, так что, по всей видимости, небольшое превышение RSRZ почти не играет роли.

На рисунках 5 и 6 изображена электронная плотность остатка ASP142, а также молекулы воды, с которой, по данным MolProbity происходит перекрывание Ван-дер-Ваальсовых радиусов. Так как перекрывание происходит между молекулами Н и О, можно было бы предположить образование водородной связи, однако атом Н, который потенциально эту связь может образовывать, ковалентно связан с атомом С, а  $-(CH_2)-$  группы почти никогда не образуют водородные связи.

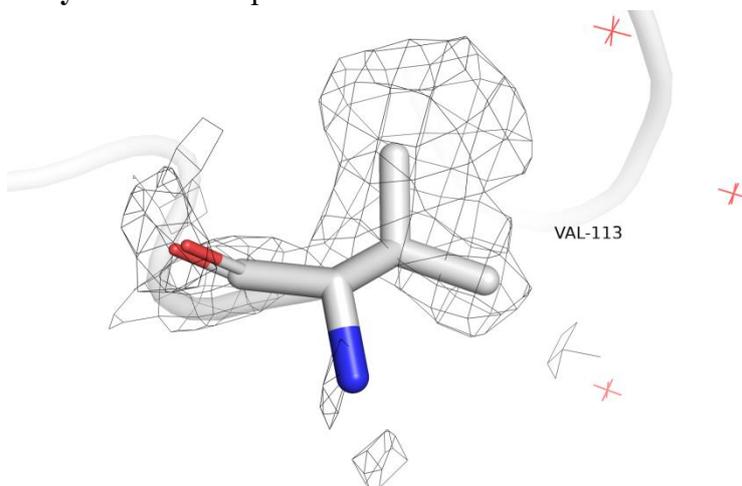
По имеющемуся изображению электронной плотности кажется, что ошибки при расшифровке структуры не допущено, но, возможно, дело в недостающем «куске» электронной плотности.

### **Остаток 2. VAL 113, цепь С.**

Данный маргинальный остаток имеет самый высокий RSRZ среди остатков данного белка (4.66).



**Рисунок 7.** Электронная плотность остатка VAL113. Уровень подрезки 1.

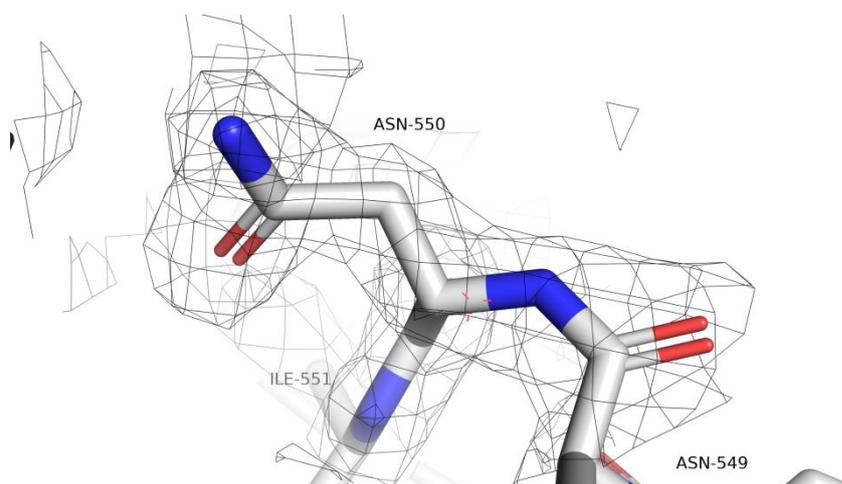


**Рисунок 8.** Электронная плотность остатка VAL113. Уровень подрезки 1.

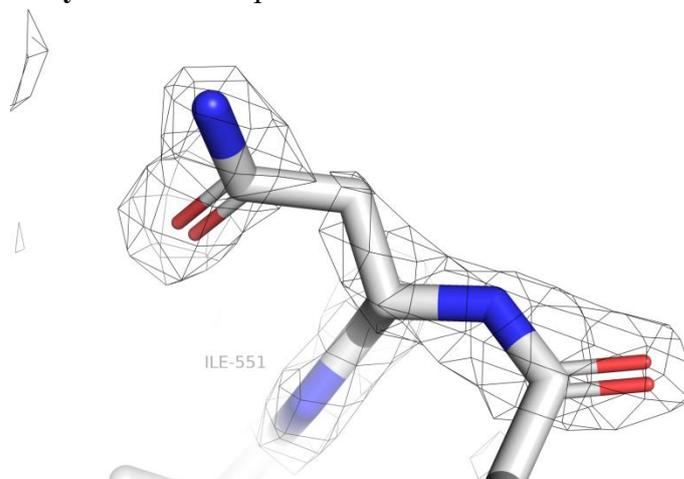
Данный остаток плохо вписывается в свою электронную плотность, его хочется подвинуть вверх и вправо. Скорее всего, это ошибка, допущенная при расшифровке структуры.

### **Остаток 3. ASN 550, цепь C.**

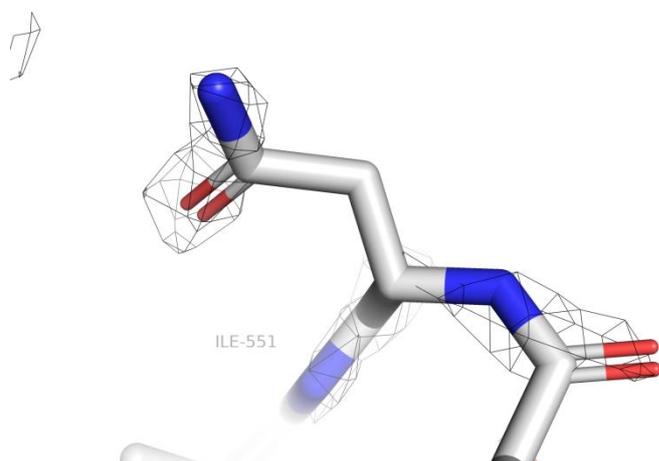
Данный остаток определен как маргинальный из-за его цис-конформации. Несмотря на то, что он не был определен, как остаток с повышенным значением RSRZ, рассмотрим его электронную плотность, чтобы определить, является ли цис-конформация ошибкой расшифровки или особенностью структуры.



**Рисунок 9.** Электронная плотность остатка ASN550. Уровень подрезки 1.



**Рисунок 10.** Электронная плотность остатка ASN550. Уровень подрезки 2.



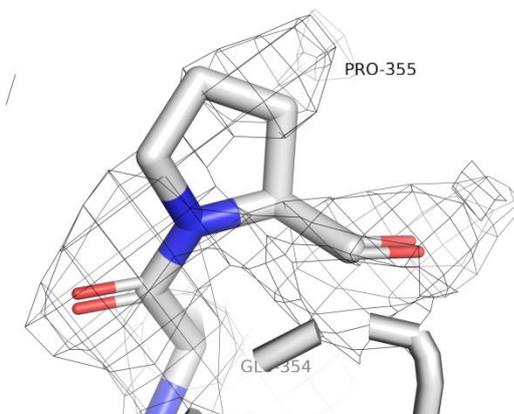
**Рисунок 11.** Электронная плотность остатка ASN550. Уровень подрезки 3.

Визуально остаток хорошо вписывается в электронную плотность.

При увеличении уровня подрезки видим, что, скорее всего, цис-конформация данного остатка является особенностью структуры: остаток в цис-конформации вписывается в электронную плотность лучше, чем это делал бы остаток в транс-конформации.

**Остаток 4. PRO 355, цепь C.**

Данный остаток является маргинальным из-за цис-конформации и высокого значения RSRZ (3.56).



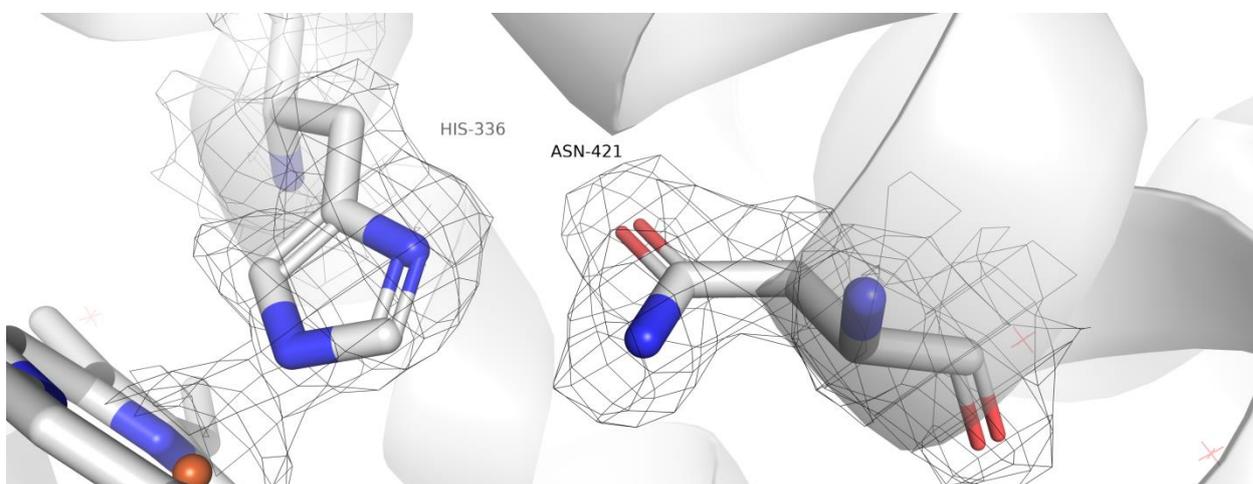
**Рисунок 12.** Электронная плотность остатка PRO355. Уровень подрезки 0.7.

На рисунке 13 видно, что электронная плотность электронная плотность данного остатка плохая и не полностью покрывает иминогруппу пролина даже на уровне подрезки 0.7. B-фактор данного остатка равен 34.79, что является довольно высоким значением (но объясняемым, так как PRO355 располагается на периферии молекулы).

Электронная плотность, даже не слишком хорошая, хорошо объясняет, что цис-конформация пролина является в данном случае особенностью структуры, а не ошибкой расшифровки. Несмотря на то, что цис-конформация пептидной связи довольно редка, пролин имеет тенденцию находиться в цис-конформации чаще других аминокислотных остатков.

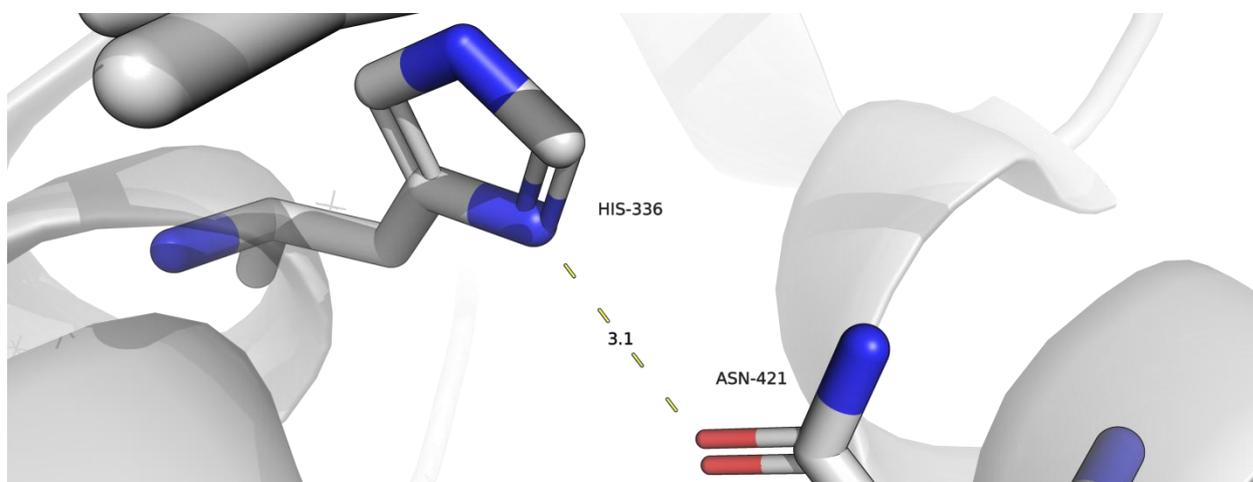
#### **Остаток 5. ASN 421, цепь C**

Данный остаток является маргинальным из-за перекрытия 0.65Å HD21 и C 336 HIS HD1, а также из-за подозрения на инверсию боковой цепи.



**Рисунок 13.** Электронная плотность остатка ASN421. Уровень подрезки 1.

Перекрытие Ван-дер-Ваальсовых радиусов, скорее всего, связано с образованием водородной связи между остатками ASN421 и HIS336.



**Рисунок 14.** Возможная водородная связь между остатками HIS336 и ASN421.

Геометрия для образования водородной связи, если судить по рисунку 15, получается довольно неплохая и вполне может быть. Однако для образования водородной связи инверсии боковой цепи ASN421 быть не должно. Поэтому считаем причиной перекрытия Ван-дер-Ваальсовых радиусов образование водородной связи и скорее отвергаем подозрение на инверсию.

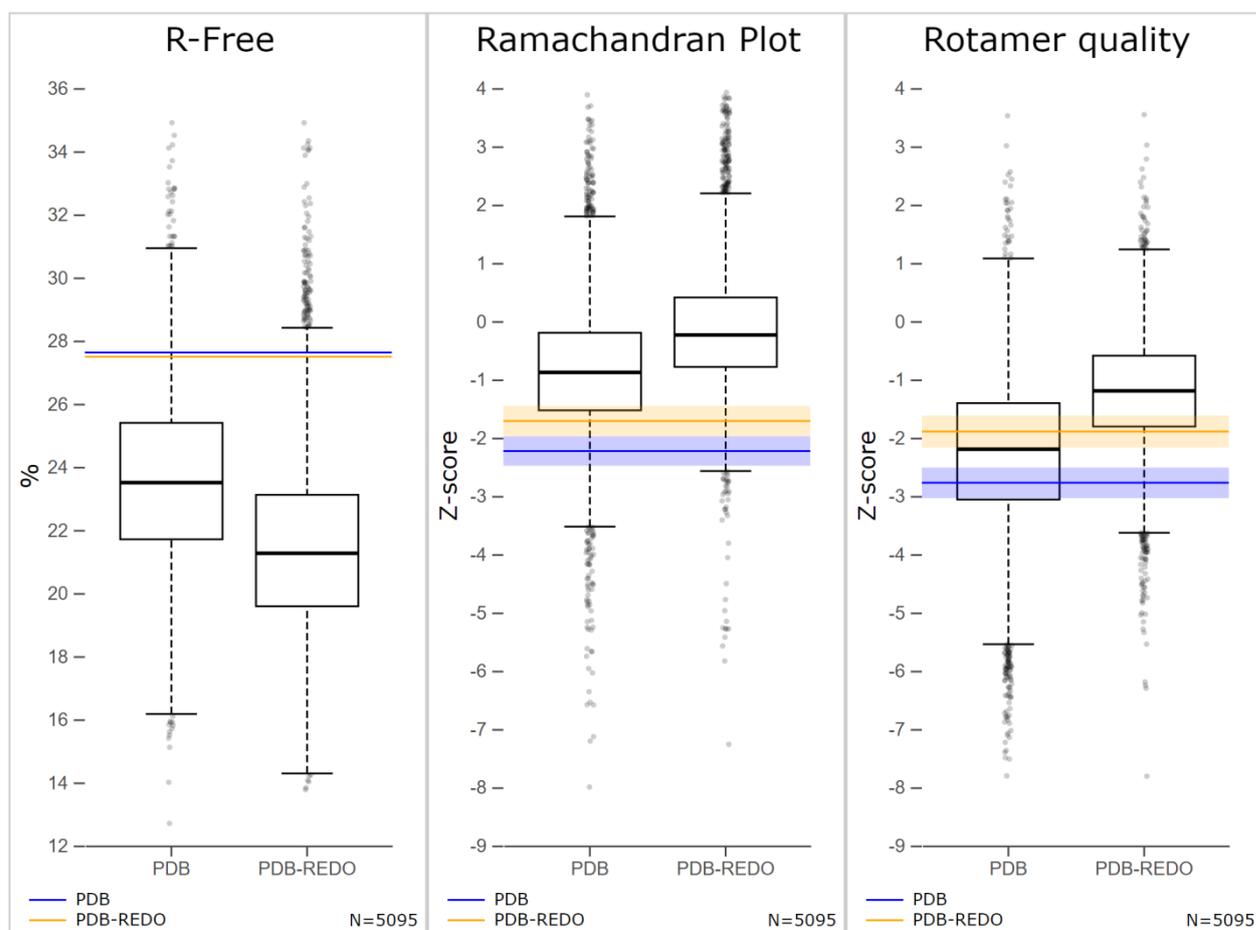
## Сравнение модели из PDB с моделью из PDB\_redo.

База PDB\_redo содержит оптимизированные модели.  
Сравним параметры нашей модели и модели из PDB\_redo<sup>5</sup>.

**Таблица 3.** Сравнение R и R<sub>free</sub> моделей из PDB и PDB\_redo.

	PDB	PDB_redo
R	0,2220	0,2287
R <sub>free</sub>	0,2762	0,2749

Как видно, значения R и R<sub>free</sub> практически не изменились (на десятые доли процента).



**Рисунок 15.** Сравнение параметров модели из PDB и PDB\_redo.

Как видно на рисунке 16, немного улучшилось «качество» торсионных углов и ротамеров (произошло 17 замен ротамеров и удаление 10 молекул воды). Число «preferred regions» на карте Рамачандрана увеличилось (правда, всего на 2), число «allowed regions» уменьшилось (на 4), но возросло число остатков, находящихся на запрещенных областях карты Рамачандрана (с 2 до 4). Надо также отметить, что PDB\_redo говорит, что запись PDB содержит 2 маргинала по расположению на карте Рамачандрана, в то время как PDB validation report утверждает, что такой остаток всего 1.

## Выводы.

Рассмотренная нами структура имеет хорошее разрешение (2 Å при полноте данных 94%), неидеальные, но и не плохие значения R и  $R_{free}$  (равные, соответственно, 0,222 и 0,278), относительно небольшой показатель недопустимых перекрытий атомов (5 на 1000), довольно мало остатков с высоким RSRZ, однако довольно большой процент маргинальных ротамеров (1,91%). В целом, я бы оценила качество расшифровки как удовлетворительное из-за уже упомянутого процента маргинальных ротамеров, некоторых не вписывающиеся в электронную плотность остатков, а также областей (например, в начале и в конце цепи A) с отсутствующей электронной плотностью.

## Список литературы.

- (1) Forbes, L. V, Sjögren, T., Auchère, F., Jenkins, D. W., ¶3, B. T., Laughton, D., Hemsley, P., Pairaudeau, G., Turner, R., Eriksson, H., Unitt, J. F., and Kettle, A. J. (2013) Potent Reversible Inhibition of Myeloperoxidase by Aromatic Hydroxamates \*.
- (2) <https://www.rcsb.org/structure/4C1M>
- (3) <http://molprobity.biochem.duke.edu/>
- (4) <http://eds.bmc.uu.se/eds/>
- (5) <https://pdb-redo.eu/db/4c1m>