

# Валидация

## Задание 1.

3L9H - моторный белок кинезин-5, структура которого была получена в ходе PCA эксперимента. Он играет важную роль в формировании веретена деления, обеспечивая поддержание его биполярности и способствуя его удлинению. Кроме того, кинезин-5 может двигаться к плюсовому или минусовому концу (чаще) микротрубочки, перемещая клеточные молекулы.

В статье, посвященной 3L9H (DOI: 10.1016/j.bmcl.2010.01.110), рассказывается об открытии ингибиторов АТФазной активности кинезина - hexahydro-2H-pyrido[3,2-c]quinolines (HHPQs). Они обеспечивают арест клеточного цикла и приводят к апоптозу. В статье рассмотрена оптимизация одного из ингибиторов (EMD) и его использование в качестве ингибитора роста опухоли.

Проанализируем некоторые параметры качества структуры из отчета о валидации:

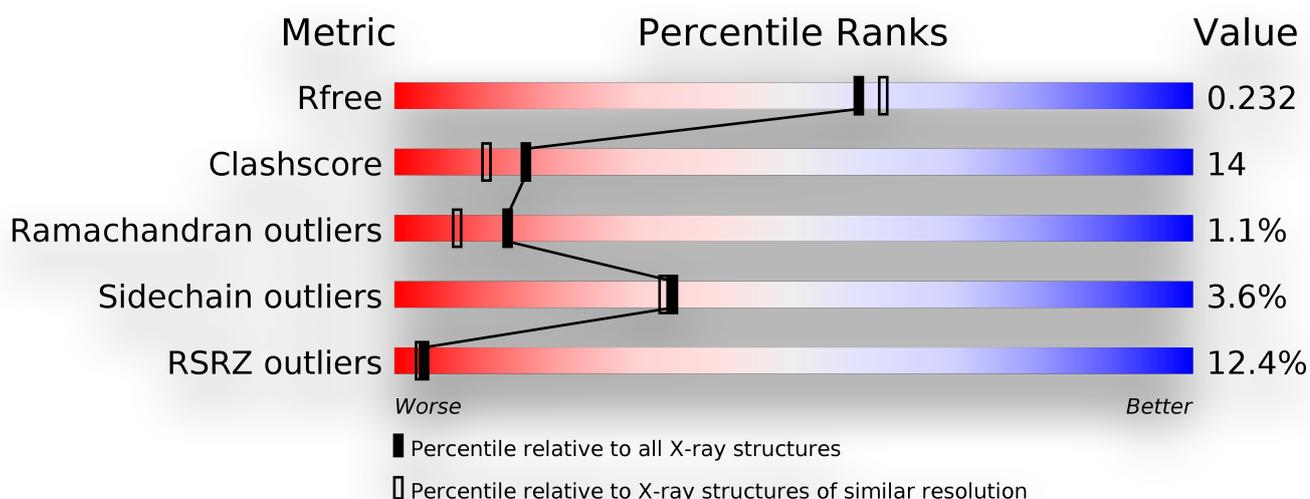


Рис 1. Графический отчет о качестве структуры.

По рисунку 1 сразу можно предположить, что качество структуры ниже среднего, близкое к плохому. Структура была получена в 2010 году, что также согласуется с моим предположением.

Однако рассмотрим метрики качества подробнее:

- ◆ **Разрешение** составляет 2,0 ангстрем при полноте данных 88%. Разрешение хорошее. Это ~ среднее значение разрешения для структур представленных в PDB. Полнота данных при этом не очень высокая. Думаю, можно ожидать неплохое качество восстановления электронной плотности.
- ◆ **Значение R-free и R-фактора** составляют 0.234 и 0.258 соответственно. Значение R-free близко к хорошему. Разность R-free - R не превышает 10%. То есть переоптимизация не наблюдается. Значение R-free

единственная метрик на рисунке 1, которая находится в благоприятном диапазоне значений. Она лучше, чем примерно у 60-65 % структур PCA в PDB.

◆ Проанализируем, как хорошо остатки ложатся на **карту Рамачандрана**.

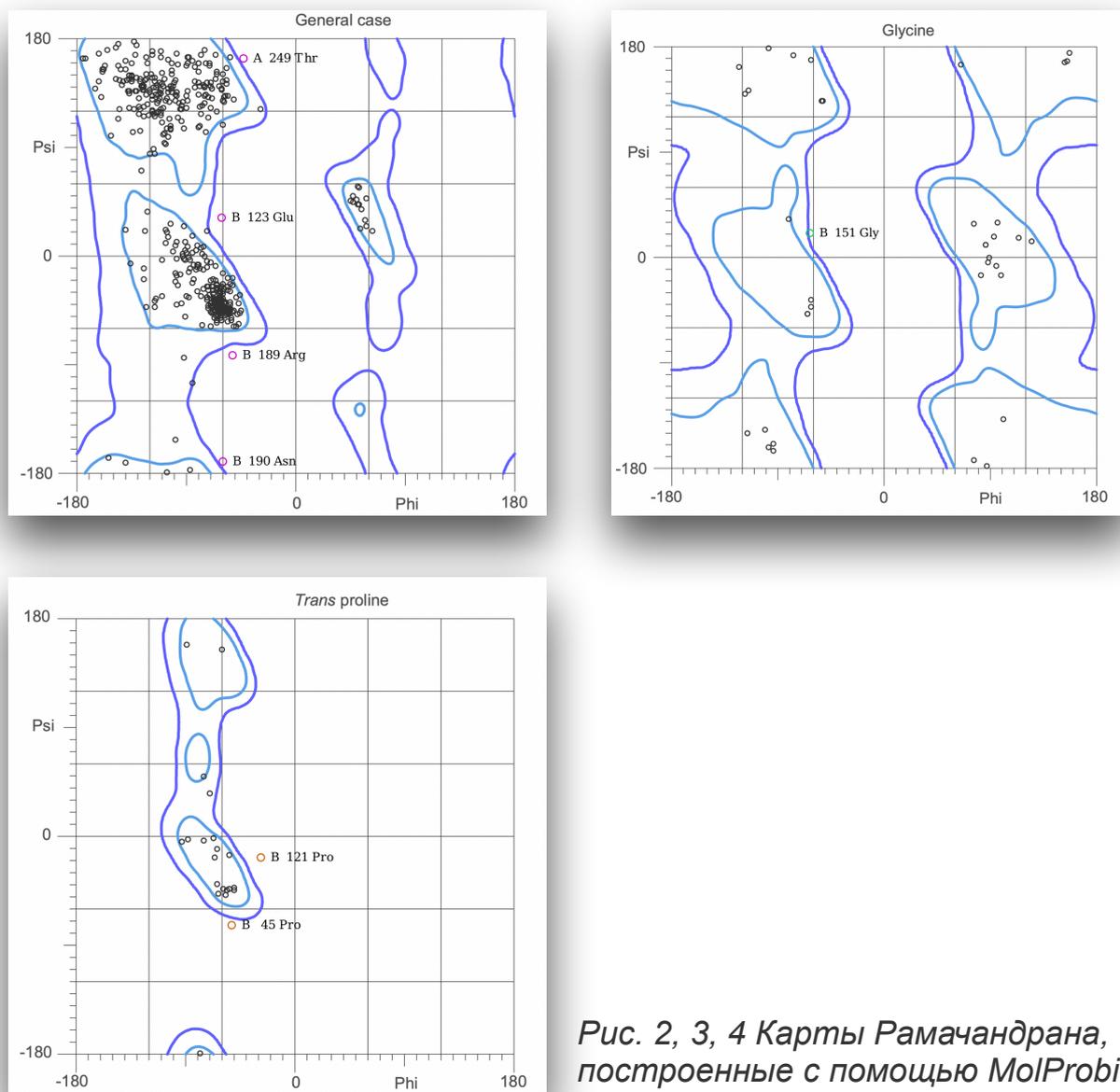


Рис. 2, 3, 4 Карты Рамачандрана, построенные с помощью MolProbity

В данной модели в асимметрической ячейке находятся 2 идентичных по составу мономера, представленных цепями А и В. При этом, все проанализированные (327 из 368) остатки цепи А, кроме **одного** маргинального (0 %), находятся в разрешенной области и 98% в предпочтительной. Однако в цепи В из всех проанализированных остатков (327 из 368) **6** аминокислотных остатков (2%) являются маргинальными по торсионным углам остовов и только 96% а.о. находятся в предпочтительной области. (Рис 2, 3, 4)

По этой метрике можно сделать вывод о неудовлетворительном качестве расшифровки положения атомов остова. Всего у 14 % моделей из PDB процент маргиналов по торсионным углам остова больше.

- ◆ Структура имеет не очень большой процент неротамеров (13 для цепи А и 8 для цепи В) (Рис 5).

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	A	293/322 (91%)	280 (96%)	13 (4%)	28	25
1	B	293/322 (91%)	285 (97%)	8 (3%)	44	46
All	All	586/644 (91%)	565 (96%)	21 (4%)	35	34

Рис. 5 Таблица с данными по ротаметрам из отчета по валидации PDB (раздел 5.3.2)

У 35 % моделей из PDB процент маргиналов по торсионным углам боковых цепей а.о. больше.

- ◆ **Clashscore** - это показатель количества близко расположенных атомов на 1000. При добавлении атомов водорода он составляет 14, что довольно много. Однако большая часть клэшей возникает между расставленными PyMol атомами водорода или между водородами воды, поэтому этот параметр, вероятно, не стоит рассматривать, как определяющий при оценке качества моей модели.
- ◆ **RSRZ** метрика показывает, насколько сильно модельная ЭП отличается от экспериментальной. Процент плохо вписанных в электронную плотность остатков в моей структуре составляет 12.4. Всего у 4% моделей из PDB процент маргиналов по соответствию электронных плотностей больше.
- ◆ Теперь оценим напрямую **качество электронной плотности** модели. В отчете по валидации указано, что атомы, для которых нет подтверждающей их координаты ЭП отсутствуют. Однако, если посмотреть на электронную плотность в PyMol, то при уровне значимости 1 (mesh = 1) некоторые атомы все же не покрыты электронной плотностью. (Рис. 6)

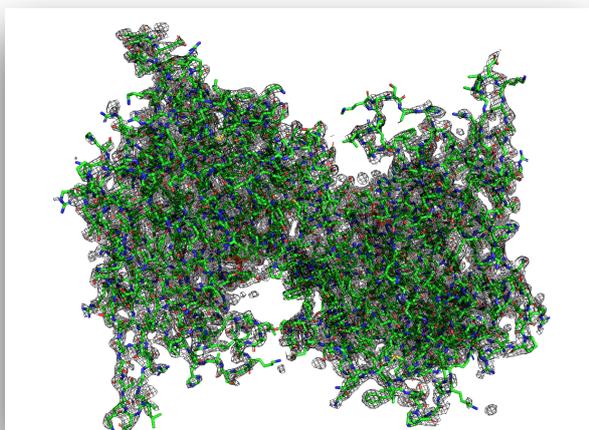


Рис. 6 Структура 3L9H с электронной плотностью (mesh = 1, carve = 1,5)

Не покрыты электронной плотностью преимущественно атомы, расположенные в боковых цепях поверхностных, а как следствие подвижных аминокислотных остатков, которые не являются функционально значимыми.

## Задание 2.

В этом задании я опишу остатки, маргинальные по описанным выше метрикам качества (по данным [MolProbity](#), отчета PDB).

### Маргинал по торсионным углам остова (Рис. 7):

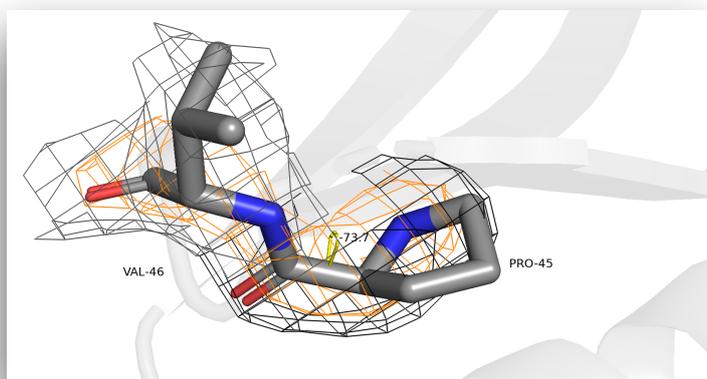


Рис.7 Остаток Пролина- 45, цепь В (Оранжевым покрашена ЭП при mesh = 2, carve = 1.5, черным при mesh = 1, carve = 1.5)

На рисунке 4 по карте Рамачандрана видно, что пролин никак не может попасть в разрешённую область из-за значения угла пси. Скорее всего, торсионный угол изменился бы, если ЭП лучше бы покрывала атом азота валина.

### Теперь рассмотрим маргинал по ротамерам (Рис. 8):

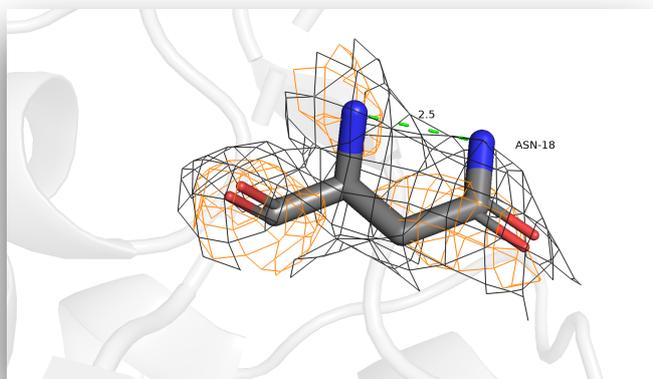


Рис.8 Остаток аспарагина- 18, цепь А (Оранжевым покрашена ЭП при mesh = 2, carve = 1.5, черным при mesh = 1, carve = 1.5)

Думаю, неротамерность данного остатка связана с очень близким расположением атомов азота в его составе. Расстояние между ними составляет всего 2.5 Ангстрема. Яркая выраженная ЭП для азота в

боковой цепи отсутствует. Я бы расположила его чуть правее на рисунке 8. Хотя тогда, вероятно, появилась бы проблема с значением угла CB-CG-ND2. Кроме того, можно было бы изменить конформацию CG атома, тогда бы проблема также исчезла.

### Клэши между атомами (не H) а.о. (Рис. 9):

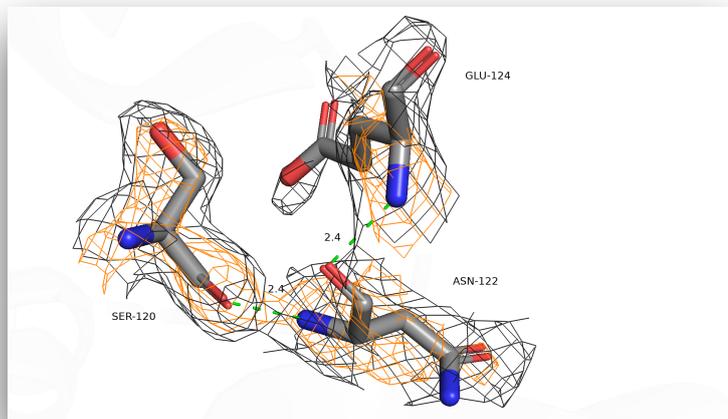


Рис.9 Остаток глутамата- 124, остаток аспарагина-122, остаток серина-120 цепь B, (Оранжевым покрашена ЭП при mesh = 2, carve = 1.5, черным при mesh = 1, carve = 1.5)

Я нашла клэш, состоящий сразу из 3 а.о. Атомы азота и кислорода на рисунке 9 расположены друг от друга на расстоянии всего 2.4 Ангстрема в обоих случаях.

При этом, ЭП кислорода аспарагина и азота глутамата действительно плохо определена. Однако наличие второго клэша между атомами серина и аспарагином довольно неожиданно так как они хорошо вписаны в имеющуюся ЭП. Возможно, в данном случае в ходе эксперимента она была неправильно определена, либо остатки закристаллизовались в таком неестественном положении. Также, думаю, даже внутри хорошо определенной ЭП можно разместить атомы некорректно.

### Маргинал по несоответствию экспериментальной и модельной ЭП (Рис. 10):

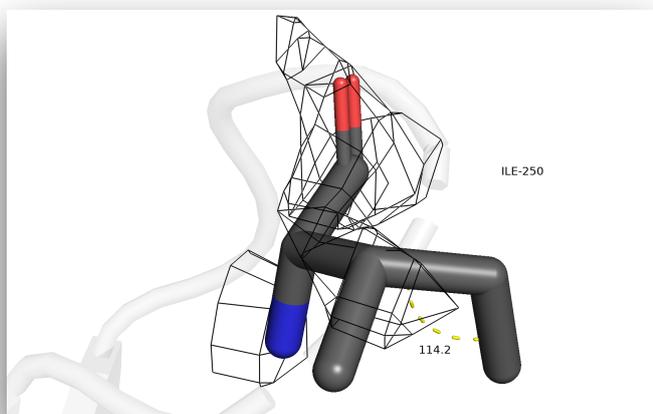


Рис.10 Остаток изолейцина- 250, цепь A, (Черным покрашена ЭП при mesh = 1, carve = 1.5)

Действительно для данного остатка почти отсутствует ЭП. Неудивительно, что он также является маргиналом по ротамерам. Его CD1 атом расположен «слишком низко». В данном случае я бы направила его так, чтобы CG1-CD1 связь была примерно параллельно CA-CB связи и угол CB-CG1-CD1 составлял  $\sim 109$  градусов.

### Задание 3.

Несмотря на большое количество маргиналов по торсионным углам остова и по несоответствию электронных плотностей, я считаю, что данную структуру можно использовать для изучения особенностей исследуемого белка. Тем более можно использовать эту структуру для задачи, поставленной перед авторами статьи (Рис. 11).

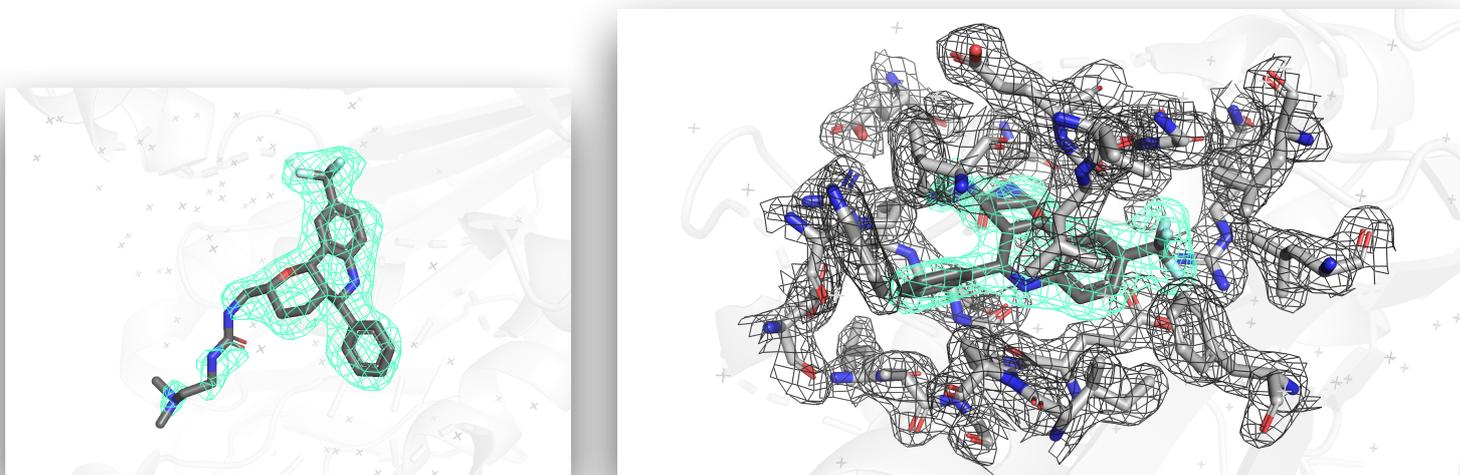


Рис.11 Слева: лиганд EMQ - ингибитор кинезина-5 (светло-зеленым изображена его ЭП при  $mesh = 1$ ,  $carve = 1.5$ ). Справа: лиганд EMQ и остатки, расположенные на расстоянии  $< 5$  Ангстрем от него. (Черным изображена их ЭП при  $mesh = 1$ ,  $carve = 1.5$ ) (для цепи B)

На рисунке 11 видим, что ЭП для лиганда и его сайта связывания очень хорошая.

Маргиналы, присутствующие в структуре располагаются преимущественно на периферии белковой глобулы, поэтому качество их расшифровки не так важно. Кроме того, в следующем задании с помощью сайта [PDB Redo](https://www.rcsb.org/pdb/redo) структура будет оптимизирована, лучше вписана в электронную плотность. Данный WEB сервер работает с 2014 года, а моя структура была загружена на PDB в 2010. Вероятно, тогда еще не было алгоритмов, которые могли бы очень хорошо вписывать остатки в электронную плотность и оптимизировать их структуру. Поэтому я ожидаю улучшения метрик качества после обработки PDB Redo.

### Задание 4.

Validation metrics from PDB-REDO		
	PDB	PDB-REDO
<b>Crystallographic refinement</b>		
R	0,2316	0,1793
R-free 	0,2844	0,2123
Bond length RMS Z-score	0,271	0,767
Bond angle RMS Z-score	0,597	0,883
<b>Model quality (raw scores   percentiles)</b>		
Ramachandran plot appearance	43	86
Rotamer normality	24	47
Coarse packing	N/A	N/A
Fine packing	26	62
Bump severity	48	68
Hydrogen bond satisfaction	38	62
WHAT_CHECK	<a href="#">Report</a>	<a href="#">Report</a>

Качество модели в целом улучшилось. Это можно увидеть на рисунке 12.

Рис.12 Таблица с сайта PDB Redo с данными по изменению метрик качества

Качество было значительно улучшено почти для всех остатков. А сильно ухудшилось лишь для LYS A 191 (A 176) и LYS B 34 (B 19), которые расположены на периферии структуры.

Кроме того исчезли все маргиналы, кроме одного, рассмотренные в задании 2 (Рис 13, 14, 15).

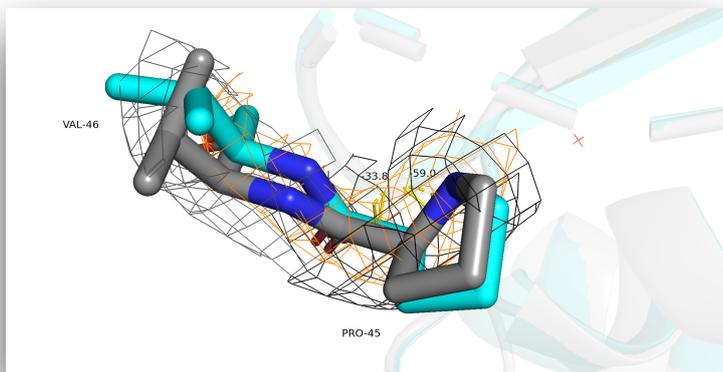
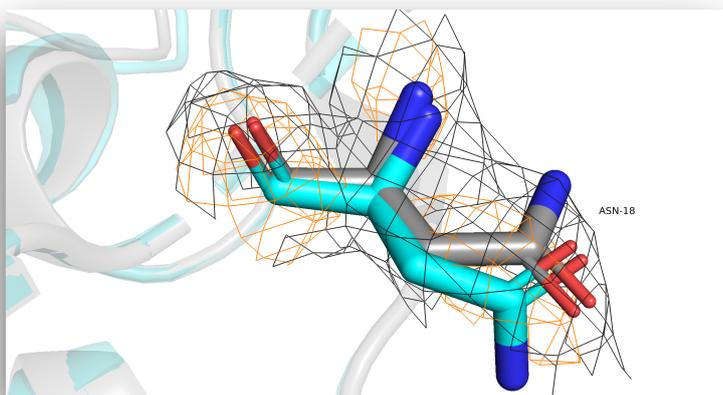
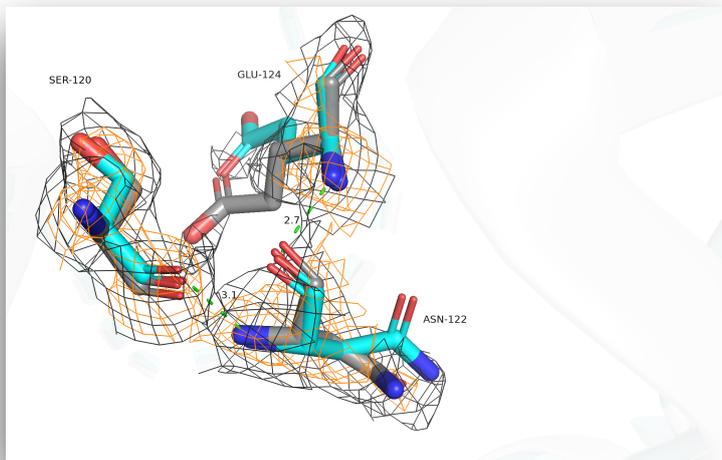


Рис.13 Остаток Пролина- 45, цепь B (Оранжевым покрашена ЭП при mesh = 2, carve = 1.5, черным при mesh = 1, carve = 1.5, бирюзовым -оптимизированная модель)



Углы фи и пси принимают теперь значения, попадающие в предпочтительную область на карте Рамачандрана.

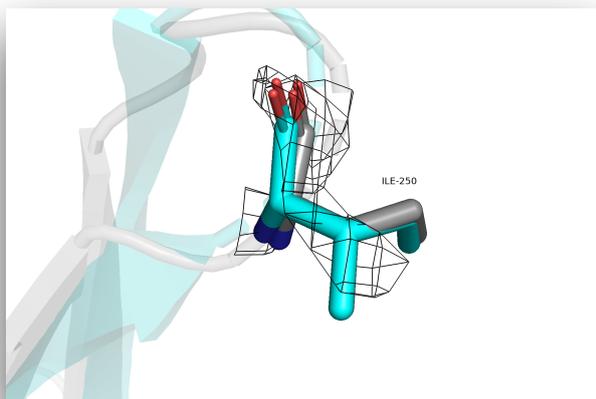
Рис.14 Остаток аспарагина- 18, цепь



А (Оранжевым покрашена ЭП при  $mesh = 2$ ,  $carve = 1.5$ , черным при  $mesh = 1$ ,  $carve = 1.5$ , бирюзовым -оптимизированная модель)

Конформация азота изменилось так, что теперь мы наблюдаем более благоприятную картину на рисунке 14. Такой аспарагин является ротамером.

Рис.15 Остаток глутамата- 124, остаток аспарагина-122, остаток серина-120 цепь B, (Оранжевым покрашена ЭП при  $mesh = 2$ ,  $carve = 1.5$ , черным при  $mesh = 1$ ,  $carve = 1.56$  бирюзовым -оптимизированная модель)



Также пропали маргиналы по клешам. Теперь расстояние между атомами составляет 2.7 и 3.1 Ангстрема.

Рис.16 Остаток изолейцина- 250, цепь A, (Черным покрашена ЭП при  $mesh = 1$ ,  $carve = 1.56$  бирюзовым -оптимизированная модель)

Однако изолейцин 250 все еще плохо вписан в экспериментальную электронную плотность (Рис. 16). Дело в том, что она почти отсутствует. Поэтому в данном случае сложно повысить качество. Странно однако, что сервисом Redo не было оптимизировано положение атомов этого остатка так, чтобы он перестал быть маргиналом по ротамерам.

Для изучения данного белка я бы однозначно взяла эту структуру из PDB Redo.