

Валидация

Задание 1

Нам было предложено проанализировать качество структуры [2cn3](#), полученной методом PCA (разрешение 1.96 А при полноте данных 100%). В кристаллографической ячейке, отображаемой в PDB-файле, есть две цепи. Структура получена для комплекса мутанта фермента ксилотриглюканазы *Xgh74A-D70A* в рамках исследования каталитических свойств фермента. Нормальная ксилотриглюканаза из целлюлосомы бактерии *Clostridium thermocellum* катализирует гидролиз ксилотриглюкана, являющегося важным компонентом клеточной стенки растений. После получения структуры WT были идентифицированы остатки, которые предполагались важными для катализа, одним из них был Asp70. Для подтверждения его важности получили мутантный фермент, который, как и предполагалось, был неактивным, и определили уже его структуру. [\[1\]](#)

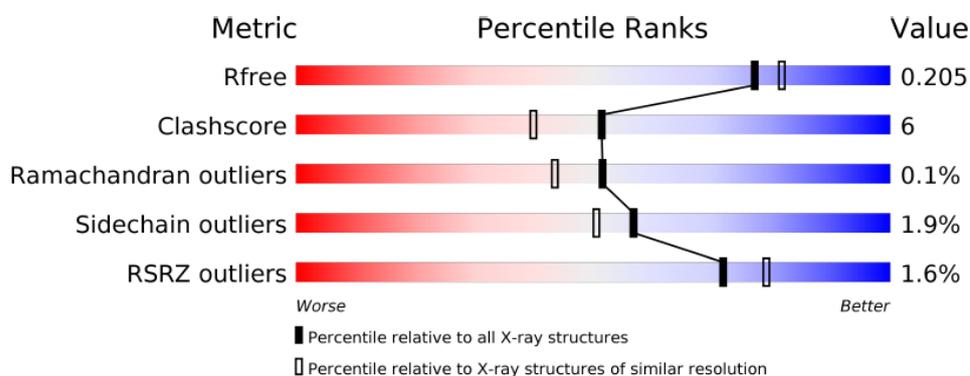


Рис.1. Данные о качестве структуры из отчета о валидации

R_free = 0.205, показывающий значение R_фактора на тестовом наборе рефлексов, отличается от рабочего R_фактора, равного 0.178, но не очень сильно. Clashscore, показывающий меру перекрытия Ван-дер-Ваальсовых радиусов, относительно неплохой (согласно отчету, на 12865 неводородных атомов двух цепей приходится 127 перекрытий, при этом 5 из них были без добавленных автоматически водородных атомов). По количеству аутлайеров по двугранным углам остова (1 из 726 - пролин), торсионным углам радикалов (11 из 593) и пространственным R_факторам (12 из 728) структура тоже весьма неплоха.

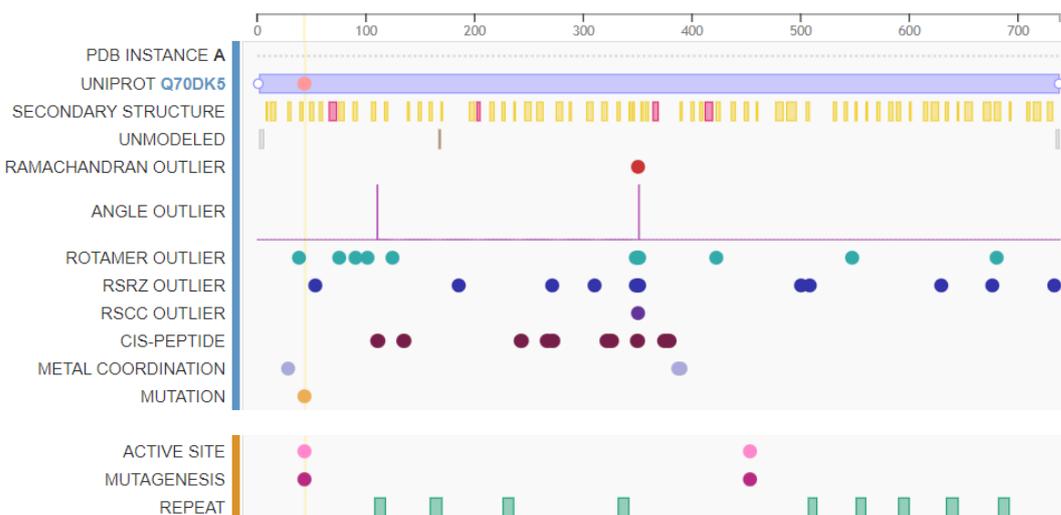


Рис.2. Карта аутлайеров

В структуре есть остатки, расположение которых не подтверждается электронной плотностью. Их можно поискать среди аутлайеров по RSRZ. Вот, например, 3 таких остатка, маргинальных по RSRZ, и ЭП на уровне подрезки 1.5 (carve 1.3). Видно, что как минимум у аспартата и глутамата ЭП не очень подтверждает RRSZ.

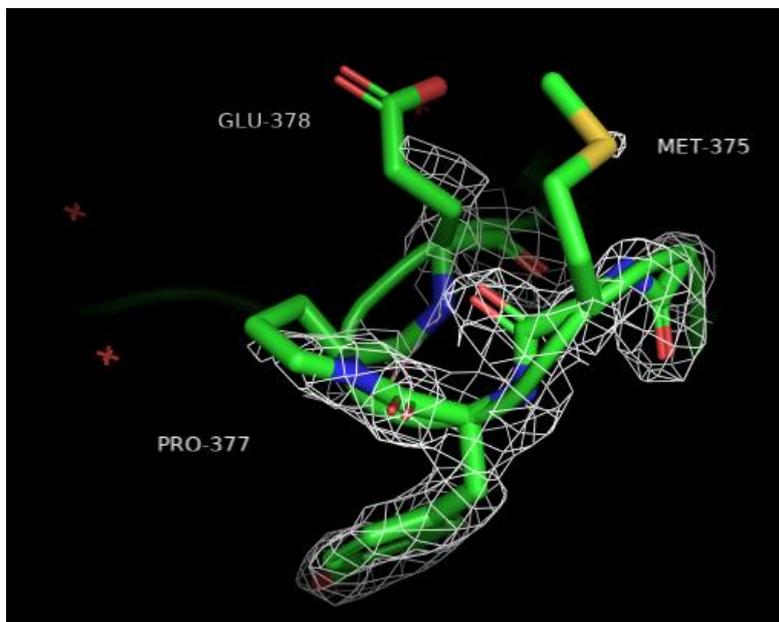


Рис.3. Остатки, положение которых не подтверждается ЭП

Эти остатки лежат на отрезке петли от бета-листа, выходящей на поверхность. Остальные остатки с экстремальным значением RSRZ я не рассматривала – они лежат достаточно хаотично, но все на поверхности. Скорее всего это не имеет биологического смысла. Это не должно напрягать, так как они расположены, тем не менее, достаточно далеко от кармана активного сайта.

Задание 2

Рассмотрим несколько аутлайеров поподробнее. В помощь нам идет отчет MolProbity (не сохранился).

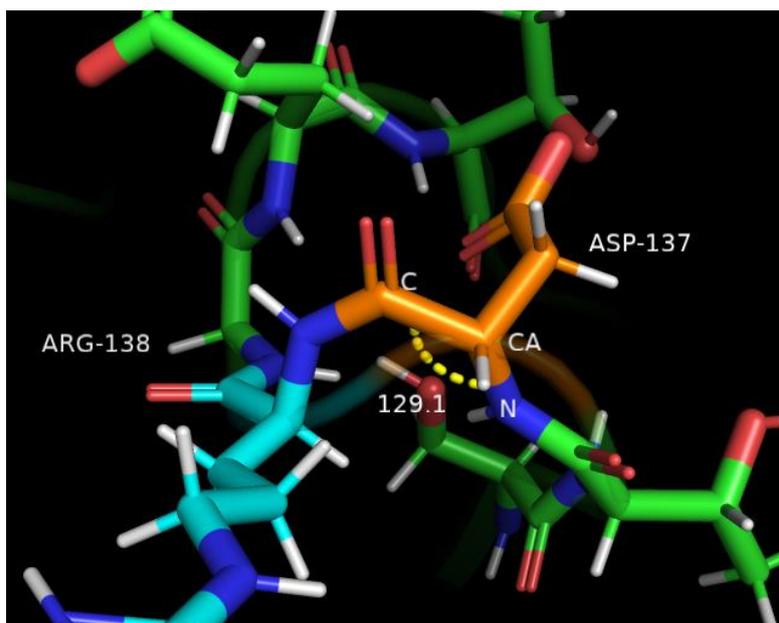


Рис.4. Цис-пептидная связь и отклонение в углах остатков Asp137 и Arg138

В цепи А встречается целых 5 непролиновых цис-пептидных связей (в двух цепях их 9/1364 – это очень много). Одна из них находится между остатками 137 и 138 (двугранный угол омега = 15.39 – не обозначен). Кроме того, в этих остатках встречаются аномальные углы связей. Например, в остатке аспарагина угол C-CA-N при sp^3 -гибризованном атоме равен 129.1° при норме около 110° (отклонение 6.5σ). Надо сказать, не все цис-пептидные связи в структуре сопряжены с аномальными углами.

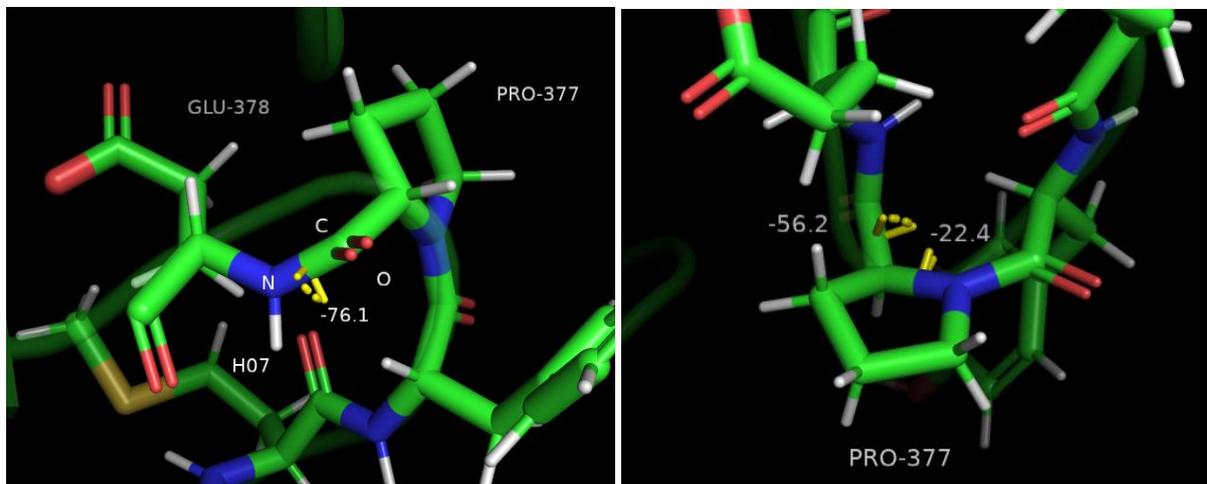


Рис. 5. Смещенная пептидная связь между Pro377 и Glu378 и цис-пролин

Вернемся к остаткам, описанным в конце задания 1. На изображении выше видим коктейль из аутлайеров: Pro377 имеет цис-конфигурацию пептидной связи (угол омега 2.67° ; на картинке – подальше). К сожалению, это не единственные его проблемы: этот остаток единственный во всей структуре является сильным аутлайером на карте Рамачандрана (остальные цис-пролины таковыми не считаются): $\phi = -22.4^\circ$, $\psi = -56.2^\circ$.

Пептидная связь между Pro377 и Glu378 вообще не плоская, чего очевидно быть не может (угол омега -76.1°).

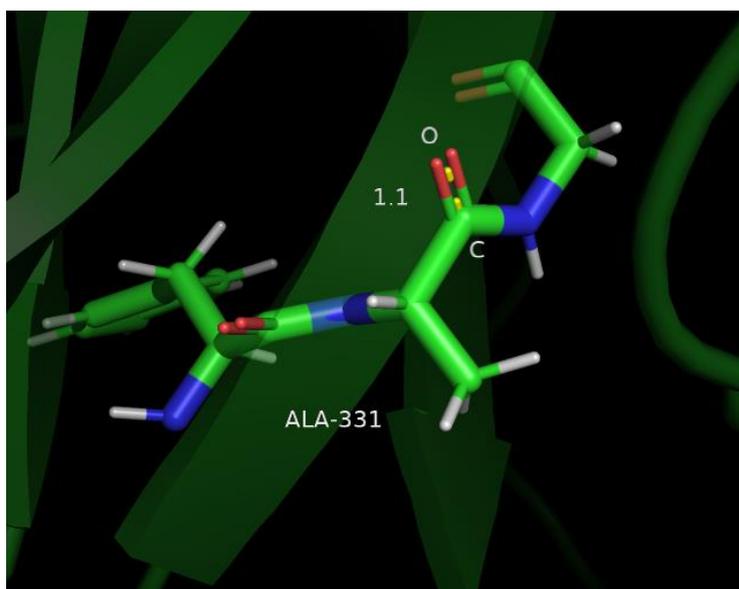


Рис. 6. Аномальная длина связи C-O в остатке Ala331 цепи В

На рисунке выше остаток Ala331 цепи В содержит аномальную C-O связь длины 1.1 А при оптимальной длине 1.2 А (отклонение 4.5σ).

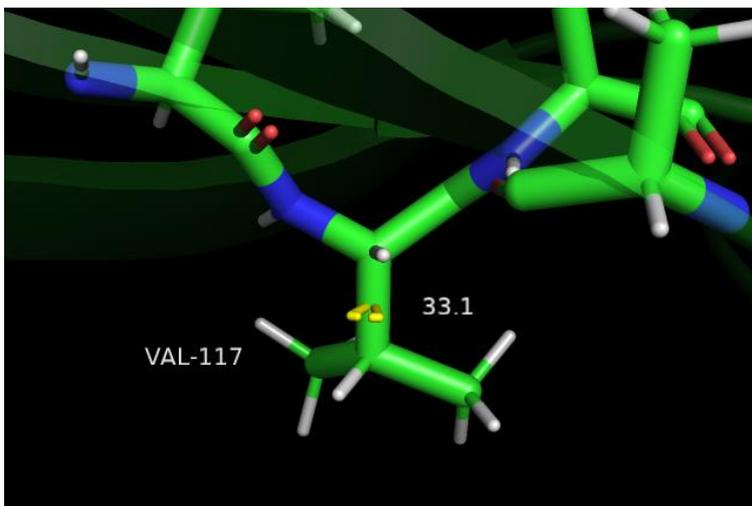


Рис. 7. Аномальный угол в остатке Val117 цепи А

Остаток Val117 цепи А является аутлайером по ротамерам (не подходит ни под один из известных ротамеров валина). У валина в ротамерах оценивается один угол – угол поворота вокруг связи Са-Сб. Угол в 33.1° не выглядит реалистичным просто из-за стереометрических причин.

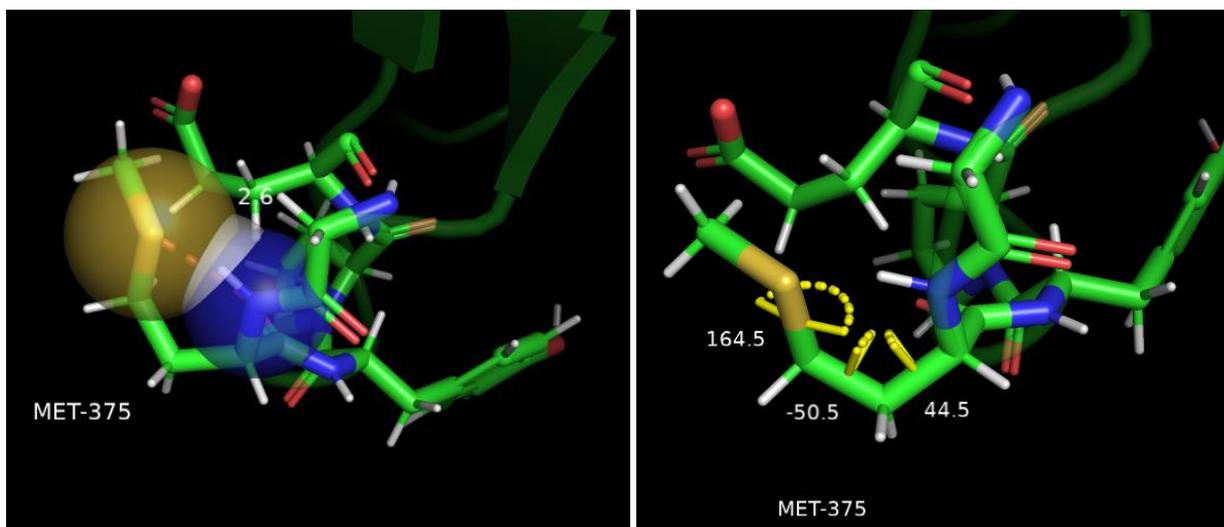


Рис.8. Наложение Ван-дер-Ваальсовых сфер атомов SD и N и нетипичные углы в остатке Met375.

Как я сказала раньше, в структуре есть 5 перекрытий Ван-дер-Ваальсовых радиусов неводородных атомов. Выше – пример такого перекрытия – между атомами SD и N в остатке Met375. Расстояние между центрами атомов при этом 2.62 Å, а само перекрытие составляет 0.63 Å.

Этот же остаток метионина является аутлайером по ротамерам (характеризуется тремя торсионными углами).

Как видно, львиную долю аутлайеров разных типов представляют всего три остатка, расположенные рядом – Met375, Pro377, Glu378. По неожиданному (нет) совпадению, это именно те остатки, положение которых плохо подтверждается ЭП (см. задание 1).

Задание 3

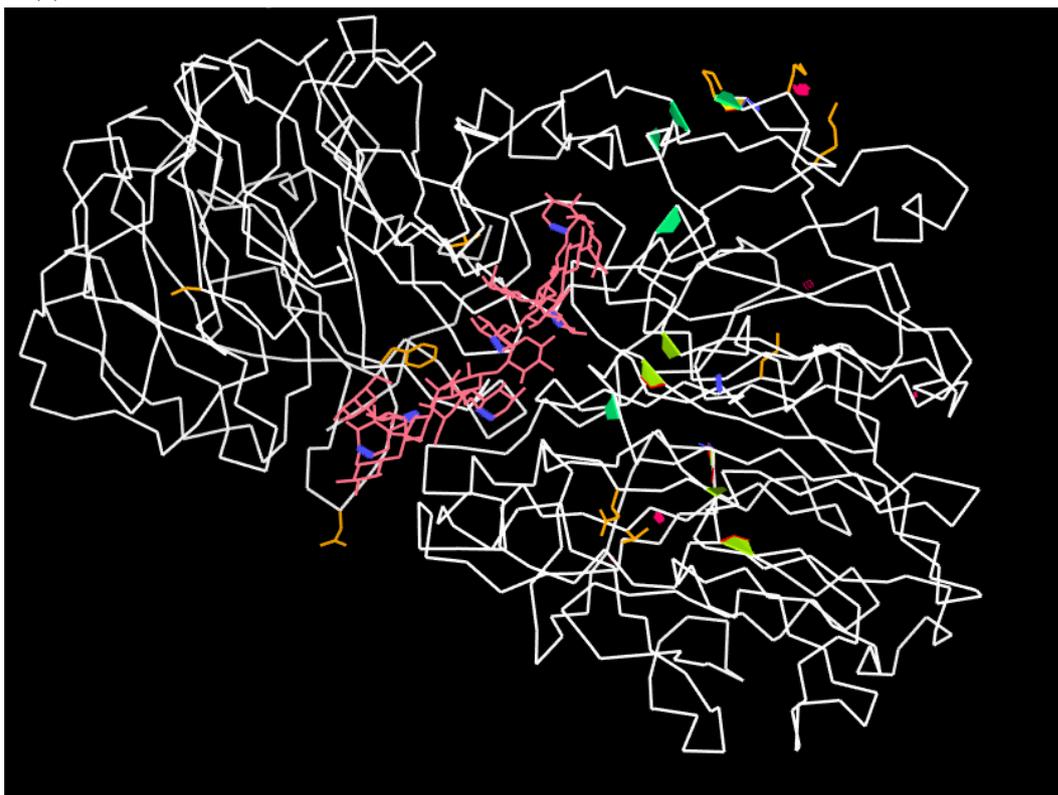


Рис.9. Визуализация аутлайеров в цепи А структуры с помощью сервиса NGL

Согласно краткому отчету по валидации, структура достаточно неплоха – R-free очень хороший, аутлайеров по углам остова и боковой цепи среднее количество, а аутлайеров по RSRZ очень немного. Однако, похоже, именно в этих немногих аутлайерах по RSRZ заключается проблема – рассмотренные нами остатки такого характера содержали по несколько «патологий» за раз (указанные выше Met375, Pro377, Glu378). В целом, аутлайеров нет в зоне контакта с сахаром. Разрешение структуры при высокой полноте данных составляет 1.96 А. Исходя из сказанного выше, мне кажется, что не нужно использовать эту структуру для исследования.

Задание 4

На сайте [PDBRedo](https://www.pdb.org/pdb/redox) можно найти отредактированную структуру, полученную с помощью сервиса WHAT_CHECK. Валидация в нем проходит по другому алгоритму, поэтому валидационные метрики для исходной модели и исправленной различаются.

Validation metrics from PDB-REDO		
	PDB	PDB-REDO
Crystallographic refinement		
R	0,1731	0,1489
R-free	0,1969	0,1732
Bond length RMS Z-score	0,492	0,481
Bond angle RMS Z-score	0,612	0,730
Model quality (raw scores percentiles)		
Ramachandran plot appearance	55	64
Rotamer normality	90	97
Coarse packing	46	43
Fine packing	69	67
Bump severity	18	18
Hydrogen bond satisfaction	38	40
WHAT_CHECK	Report	Report

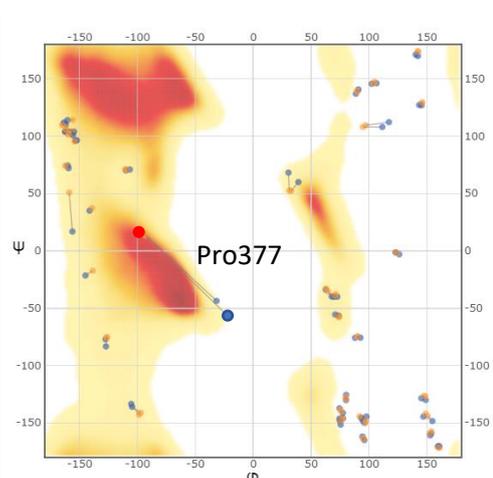


Рис. 10. Валидационные метрики и карта Рамачандрана для исходной и исправленной моделей

Видно, что некоторые метрики при исправлении заметно улучшились: R_free уменьшился, количество остатков, попадающих в предпочтительные области карты Рамачандрана, увеличилось с 55 до 64 (на карте выше голубые точки – до исправления, оранжевые - после), остатков, соответствующих известным ротамерам, также стало больше (97 против 90 до исправления). Какие-то параметры ухудшились (плотность упаковки).

Significant model changes	
Description	Count
Rotamers changed	29
Side chains flipped	1
Waters removed	13
Peptides flipped	0
Chiralities fixed	0
Residues fitting density better	46
Residues fitting density worse	0

Рис.11. Изменения в модели

Как можно судить исходя из данных таблицы выше, значительное количество ротамеров стали лучше подходить к данным библиотек. Большое количество остатков стали лучше подтверждаться электронной плотностью. Это даёт надежду на исправление замеченных ранее ошибок.

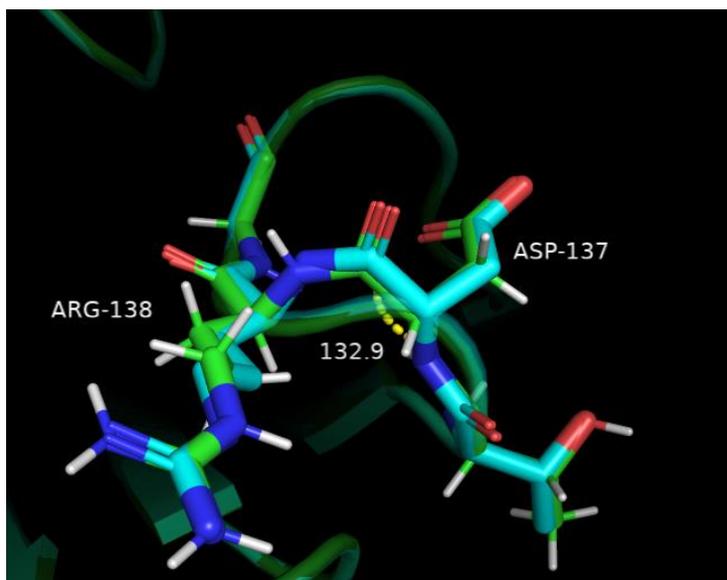


Рис.12. Угол в Asp137 исправленной структуры

Согласно данным PDBRedo, качество остатка Asp137 стало лучше (Δ RSCC 0.176). Но какого-то существенного изменения углов, к которым были вопросы, не произошло. С чем это может быть связано – непонятно, остаток находится далеко от активного центра.

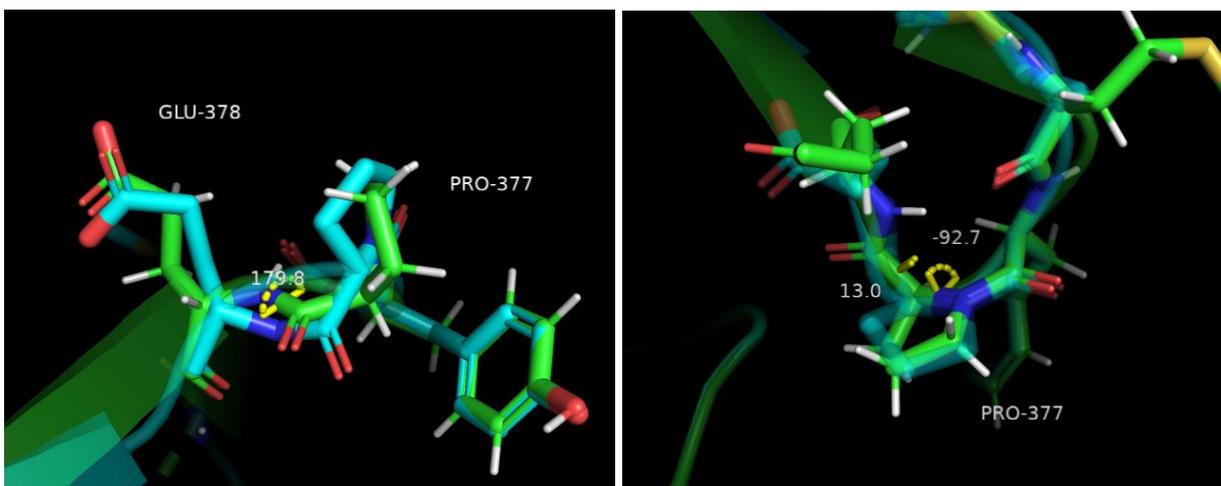


Рис. 13. Исправления в остатках Glu378 и Pro377

После исправления «нормальность» проблемного региона улучшилась: пептидная связь между пролином и аспаратом теперь очевидно в транс-конформации. Углы фи и пси для пролина заметно улучшились (на рис.10 есть жирные точки, обозначающие этот пролин до и после исправления). Таким образом, пролин теперь попадает в область допустимых углов.

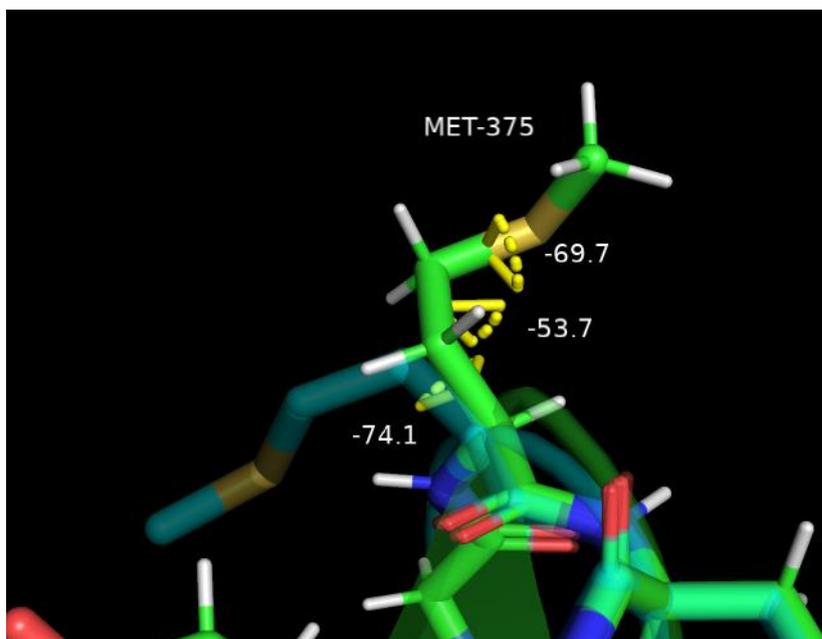


Рис. 14. Исправления в остатке Met365

Углы в ротамере Met365 тоже сильно изменились – по-видимому, в лучшую сторону. Перекрывание Ван-дер-Ваальсовых радиусов N и SD тоже, очевидно, ушло (показывать сферы не стала).

Вывод

При исправлении структуры заметно снизилось количество аутлайеров. Те же, что остались, можно попытаться объяснить с точки зрения функционирования белка (но я этого делать сейчас не хочу). Таким образом, я думаю, исправленную структуру можно использовать в исследованиях и делать выводы об особенностях белка.