

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

Факультет биоинженерии и биоинформатики

**Отчет о качестве расшифровки
структуры белка KDGA (PDB:1W37)
методом рентгеноструктурного анализа**

**Отчет студентки 4 курса
Бутусовой А. А.**

Москва, 2014

Аннотация

Этот отчет посвящен анализу качества модели 2-кето-3-дезоксиглутонат альдозаза (1W37.pdb[1]), выделенной из *Sulfolobus solfataricus*. Структура получена с помощью рентгеноструктурного метода определения структур молекул. В отчете собрана некоторая информация о показателях качества и маргинальных остатках, и на основе этих данных сделан соответствующий вывод о качестве самой модели.

Введение

KDGA – 2-кето-3-дезоксиглутонат альдозаза, белок, катализирующий реакцию расщепления D-2-кето-3-дезоксиглюконата (KDG) и D-2-кето-3-дезоксигалактоната (KDGal) до пирувата и D-глицеральдегида у *Sulfolobus solfataricus*. Данный белок представляет гомотетрамером, мономер которого имеет молекулярную массу 33 kDa и 293 аминокислотных остатка на каждый мономер. KDGA является альдозазой первого типа, то есть осуществляет реакцию через образование основания Шиффа между лизином активного центра и участком α -кетокислоты субстрата. Эта альдозаза принадлежит к семейству N-ацетилнейраминат лиазы (NAL).

В статье «*The Structural Basis for Substrate Promiscuity in 2-Keto-3-deoxygluconate Aldolase from the Entner-Doudoroff Pathway in Sulfolobus solfataricus*»[2] представлена структура 2-кето-3-дезоксиглутонат альдозаза, имеющая разрешение 2.0 Å, полученная методом многоволнового аномального рассеяния. Также в статье были описаны результаты по получению кристаллической структуры KDGA в составе основания Шиффа с субстратами (KDG, KDGal), но они не будут рассмотрены в отчете.

Результаты

Структура KDGA была получена Alex Theodossis, Helen Walden, Elaine J. Westwick, Helen Connaris, Henry J. Lamble, David W. Hough, Michael J. Danson и Garry L. Taylor и опубликована статья 2004 году[2]. Для решения фазовой проблемы изначально использовался метод молекулярного замещения (MR) с использованием структур из того же семейства NAL, однако это не увенчалось успехом. Поэтому для разрешения фазовой проблемы

использовался метод многоволнового аномального рассеяния с применением селена в качестве аномально рассеивающего атома в составе селенометионина. Полученные рефлексы были обработаны с помощью программ DENZO и SCALEPACK, в случае исходной альдолазы, SOLVE, DM и CCP4, для белка, содержащего селенометионин.

Из данных EDS[3] получено, что количество уникальных рефлексов 98295, по данным статьи также 98295, количество использованных рефлексов равно 88470, основываясь на статье и на данных pdb-файла[1]. Число рефлексов, сила которых превышает стандартное отклонение более чем в три раза, равно 96132, что составляет 97,8% от общего числа измеренных рефлексов. А также полнота данных равна 99.1%.

Рассматривая данные по разрешению структуры, статья дает данные о высоком разрешении структуры, равном 2.0 Å, те же значения присутствуют в pdb-файле и сервисе EDS[3]. Низкое разрешение, приведённое в pdb-файле, равно 32.0Å, а EDS – 32.14 Å, в статье данных о низком разрешении не приведены.

Полученный кристалл имеет кристаллографическую группу: $P 2_1 2_1 2_1$ (некристаллографическая симметрия), в каждой асимметрической ячейке, согласно статье, содержатся два тетрамера белка (Рис.1). Параметры полученной ячейки $a=83.577$, $b=131.553$, $c=132.556$, $\alpha=\beta=\gamma=90^\circ$, $Z=16$, то есть модифицированная структура содержит 16 молекул селенометионина. Сам же белок можно охарактеризовать, как димер димеров, то есть структура цепей A + D идентичная структуре B+C.

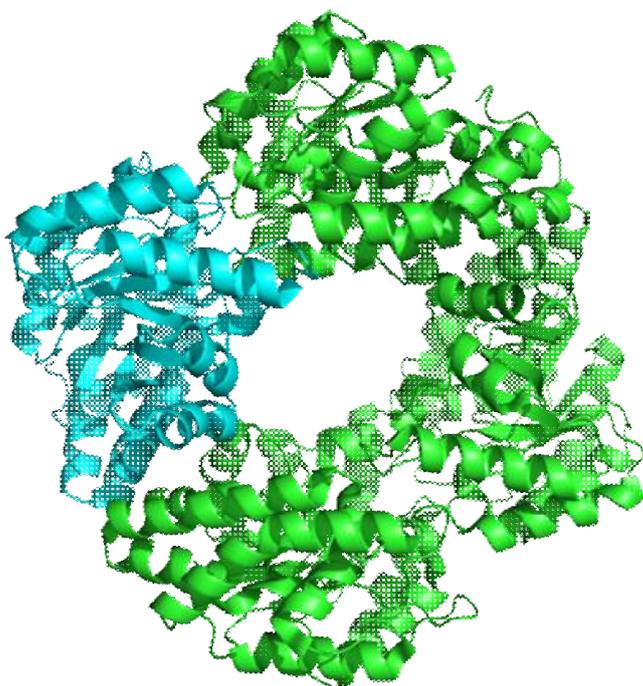


Рис.1: Изображение полной структуры 1w37, голубым изображен мономер представленного гомотетрамера.

К показателям качества модели, приведенным в rdb-файле и статье, относятся: R-фактор, R-free, B-фактор и ранее описанные разрешение, полнота. R-фактор – параметр, оценивающий, насколько полученная кристаллографическая модель соответствует набору экспериментально полученных рефлексов. Для данной структуры R-фактор равен 15.3%, что меньше 25%, значит значение хорошее. R-free оказалось равным 20.2%, это почти хороший результат, так как если говорить строго, то для хорошего R-free должен быть меньше 20%. Для оценки качества также можно рассчитать коэффициент, равный разнице между R-free и R-фактором, если он не превышает 10%, то можно считать, что переоптимизации модели не произошло. В данном случае этот коэффициент равен 4.9%, что подтверждает теорию об отсутствии переоптимизации. Также для оценки модели использовать карту Рамачандрана и пространственный R-фактор, Z-score для RSR.

По данным сервиса MolProbity[4], боковые группы остатков Asn67, His111, His163, Asn195 их всех 4 цепей (A, B, C, D) необходимо «повернуть». Для аминокислотных остатков Asn171 и Gln274 цепи D имеются явные свидетельства в пользу их инверсии. Получено 4.04 недопустимых наложений атомов на 1000 атомов (ClashScore), а также в банке структур содержится ~ 99% структур с разрешением $2.0 \pm 0.25 \text{ \AA}$ имеют ClashScore выше. Обнаружен 47 аминокислотный остаток с «нестандартной» конформацией боковых цепей (4.42% от всех остатков структуры).

Всего 1 остаток, составляющий 0.08% от общего количества аминокислот, лежит вне допустимой области на карте Рамачандрана, это остаток Asn160 цепи C. 1168 из 1172 (97.66%) остатков лежат внутри предпочитаемой области на карте Рамачандрана (Рис.2).

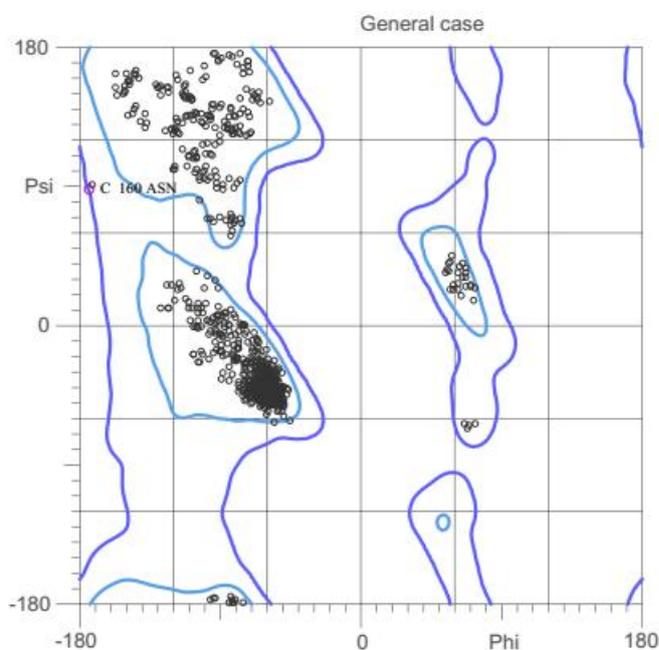


Рис.2: Карта Рамандрана для структуры 1w37, построенная с помощью сервиса MolProbity[4].

Найдено 2 C_{β} -атома с положением, отклоняющимся от типичного более чем на 0.25 \AA , 9 из 9833 связи, образующих остов полипептидной цепи, имеют нетипичную длину, а 14 из 13330 углов между связями, образующими остов, отклоняются от нормы.

RSR – показатель того, насколько построенная модель соответствует «экспериментальной» электронной плотности. Для построенной модели равно $0.095(0.041)$, это значение почти граничит с пороговым значением в 10%. Так как $RSR < 10\%$, то можно считать, что значение хорошее.

Если рассмотреть картину распределения RSR, для всех аминокислотных остатков всех белковых цепей (Рис.3), полученная с сервиса EDS[3], то, нетрудно заметить, что в цепи есть только 2 остатка, чей $RSR > 0.2$; и по одному остатку в цепях В, С, D. При более подробном рассмотрении оказывается, что в каждой цепи Glu294 остаток является маргиналом. Таким образом, во всей структуре всего 5 остатков имеют $RSR > 0.2$, а это 0.42% от всех аминокислотных остатков белка.

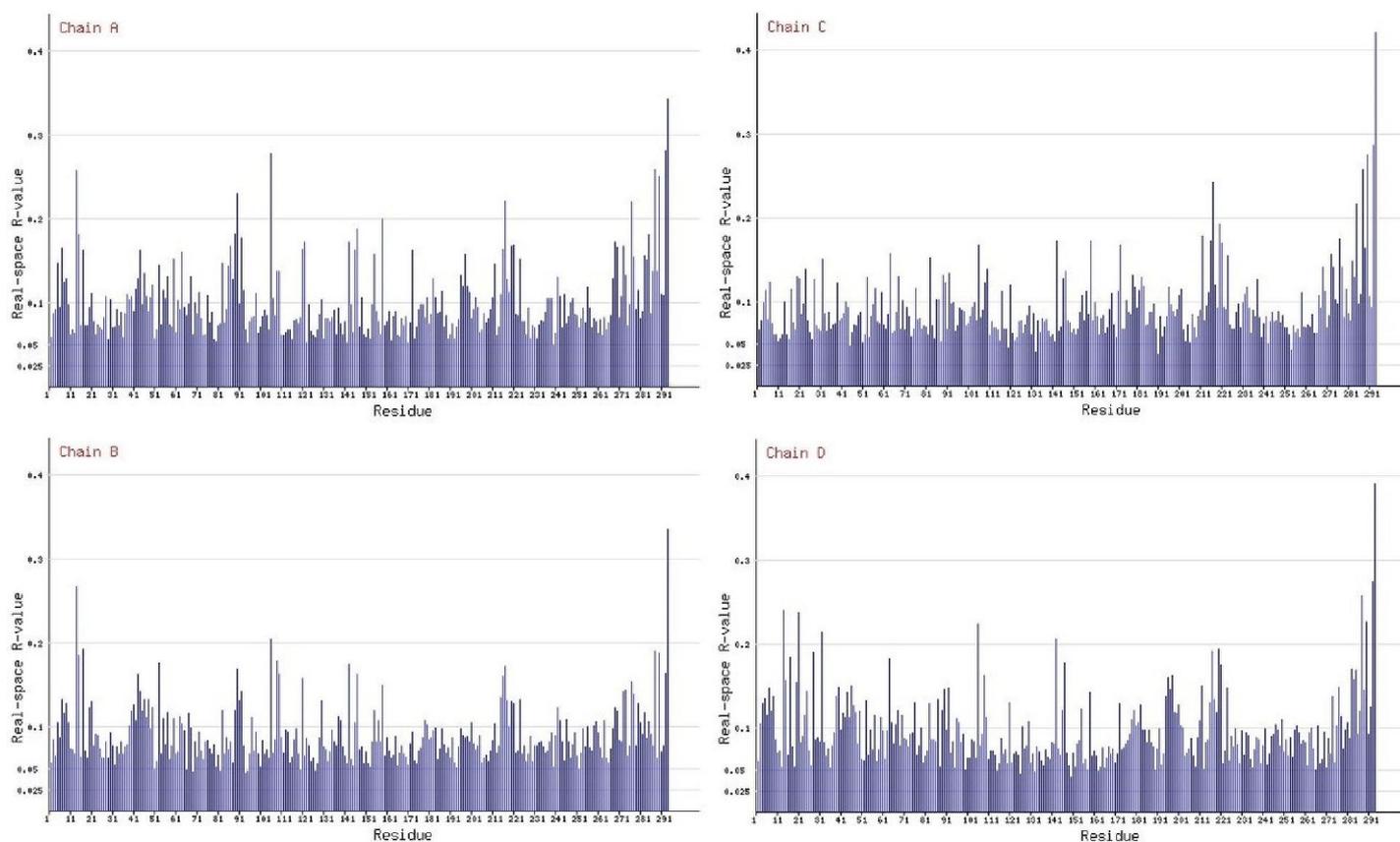


Рис.3: Распределение RSR для всех аминокислотных остатков всего белка, на каждой картинке графики для каждой цепи[3].

Z-score для RSR - параметр, позволяющий сравнивать RSR остатка модели со средним RSR в выборке моделей PDB с таким же разрешением. Проанализировав Z-score этой модели (Рис.4), можно сказать, что почти всех аминокислотных остатков белка (у 0.43% Z-score > 2) имеет RSR хуже, чем в других моделях с разрешением 1.80-2.00Å. Среднее значение Z-score по всем остаткам равно -0.41.

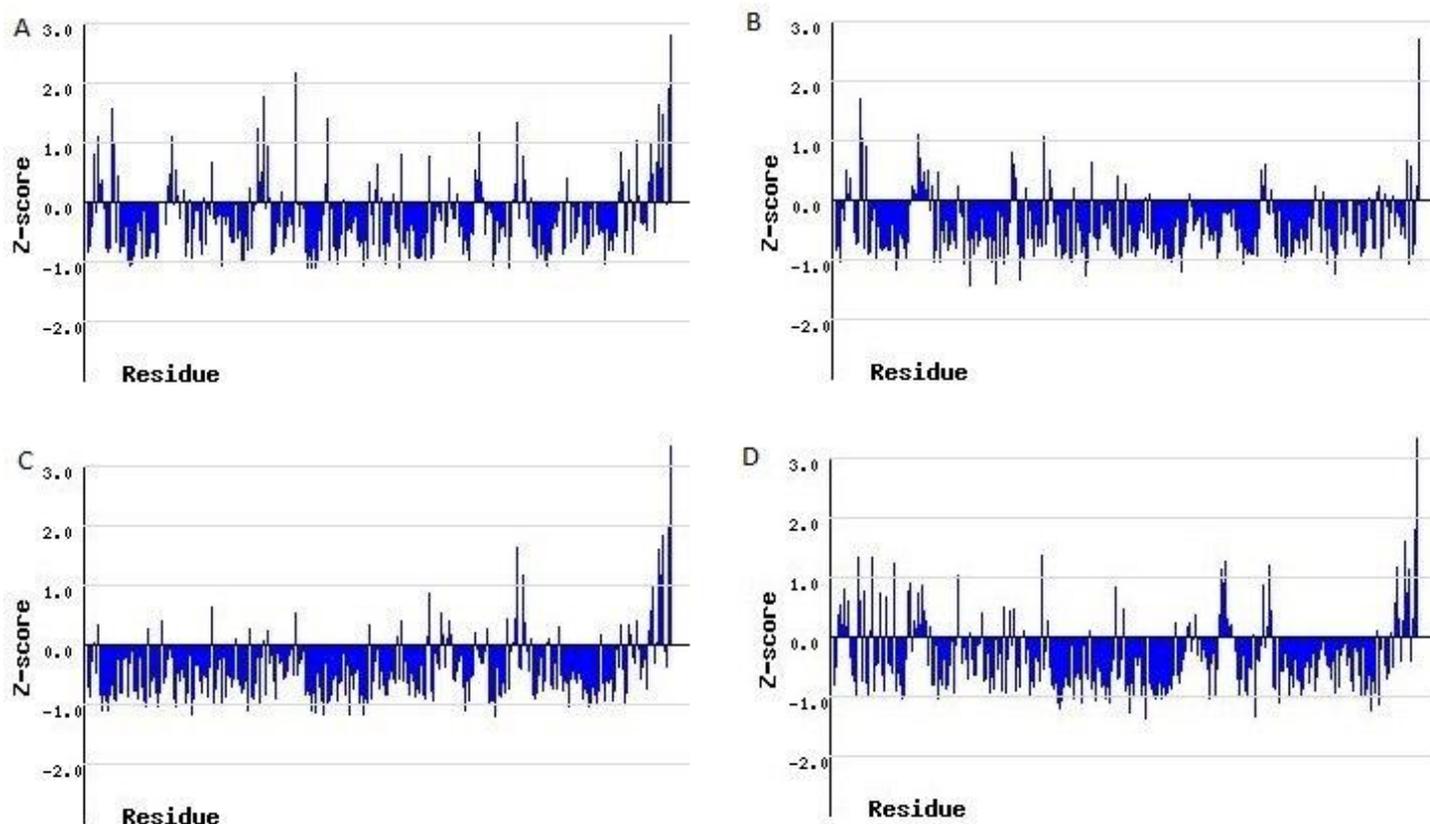


Рис.4: Изображение Z- score для цепи A, B,C, D[3].

Исследование маргинальных остатков.

Табл.1: Примеры маргинальных аминокислотных остатков

Маргинальный остаток	Критерий выбора
Pro7_A	отклонение от нормы углов, образующих остов, RSR=0.165
Met107_A	нетипичная длина связей, образующих остов, RSR=0.105
Lys138_B	отклонение от ротамеров боковой цепи, углов,

	образующих остов, C_{β} -атома, RSR=0.113
Thr157_B	отклонение от углов, образующих остов, RSR=0.107
Glu159_A	нетипичные длины связей, образующих остов, RSR=0.2
Asn160_C	из запрещенной области карты Рамачандрана, RSR=0.077
Arg169_C	нетипичные длины связей, отклонение от значения углов, RSR=0.11
Met182_B	нетипичные длины связей, отклонение от значения углов, RSR=0.087
Glu205_B	отклонение C_{β} -атома, нетипичные длины связей, RSR=0.09
Met215_C	нетипичные длины связей, отклонение от значения углов, RSR=0.135
Glu294_C	Максимальное значение RSR=0.421, Z-score=2.794

В статье помимо выбранной структуры 1w37 рассматриваются также структуры KDGA вместе с субстратами (пируватом, KDG, KDGal). Возможно, именно поэтому в описании структуры, не содержащей субстратов, не описаны никакие аминокислотные остатки. Для дальнейшего рассмотрения были выбраны аминокислотные остатки Tyr130, Thr157, Asn160, Lys138, Met182.

При изучении взаимодействия альдолазы с KDG и KDGal было получено, что остаток Tyr130 находится в окружении субстрата. Именно этот аминокислотный остаток попал в маргинальные по «нестандартности» боковых цепей. При визуализации электронной плотности круг этого остатка модель хорошо вписывается в нее даже на 2 уровне подрезки (Рис.5), также значение RSR=0.104 и Z-score=0.025. Даже если Tyr130 является маргинальным по расположению боковых цепей, вряд ли это ошибка. Возможно маргинальность Tyr130 можно объяснить тем, что он обеспечивает правильное положение Lys155, который входит в состав каталитического центра.

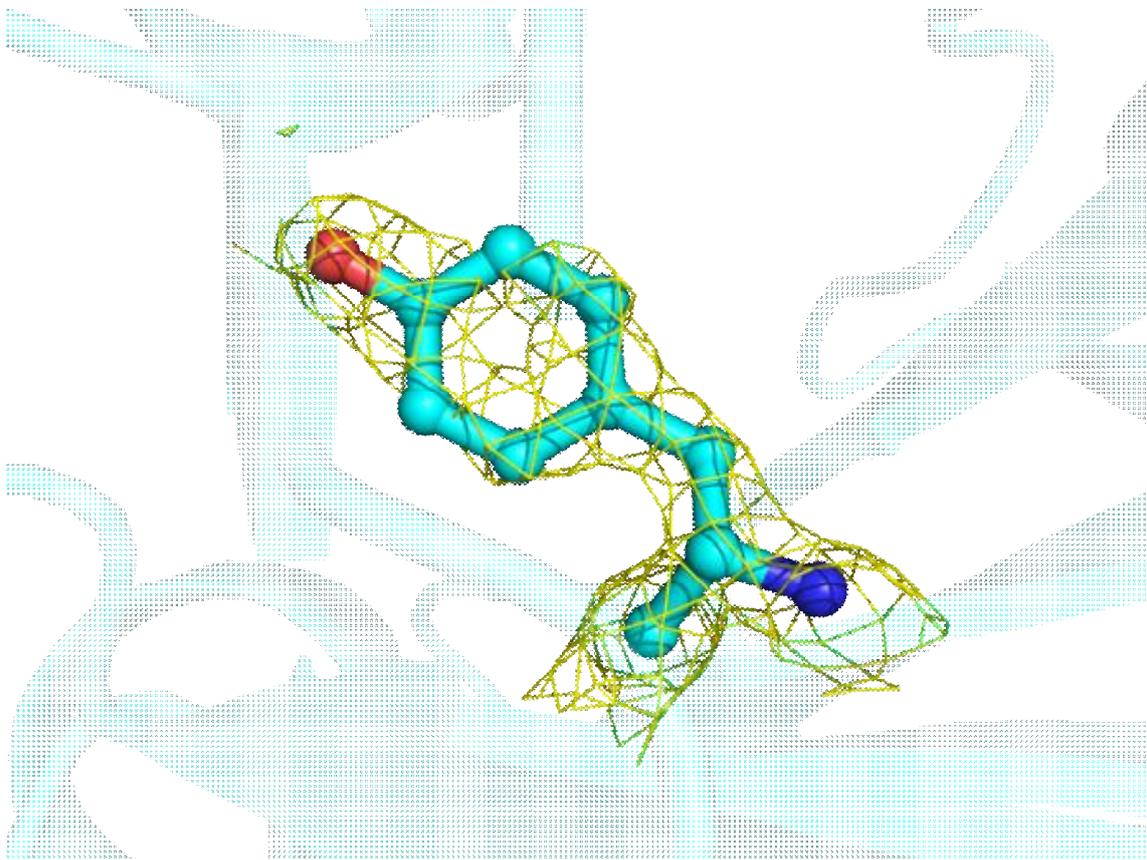


Рис.5: Визуализация электронной плотности вокруг Thr130 цепи А на уровне подрезки 2σ .

Другим остатком, рассмотренным в статье и влияющим на расположение субстрата в активном центре фермента является 157 треонин. По данным анализа с помощью сервиса MolProbity[4] он также попал в маргинальные по значениям углов, образующих остов молекулы. Если также визуализировать электронную плотность вокруг этого остатка, то получим отличное вписывание модели в плотность, при уровне подрезки 2.5 для изображения ЭП (Рис.6).

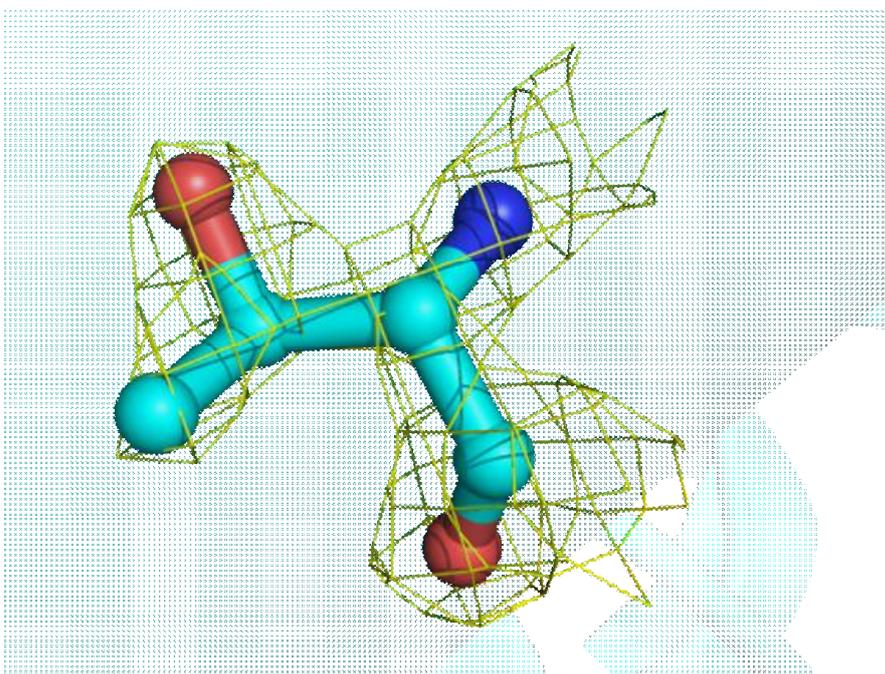


Рис.6: Визуализация электронной плотности вокруг Thr157 цепи В на уровне подрезки 2.5σ .

Опять занесение треонина в маргинальные остатки можно объяснить тем, что между KDG и KDGal образуется водородная связь между γ -кислородом треонина и гидроксильной группой при 4ом углеводе субстрата. (Рис.7)

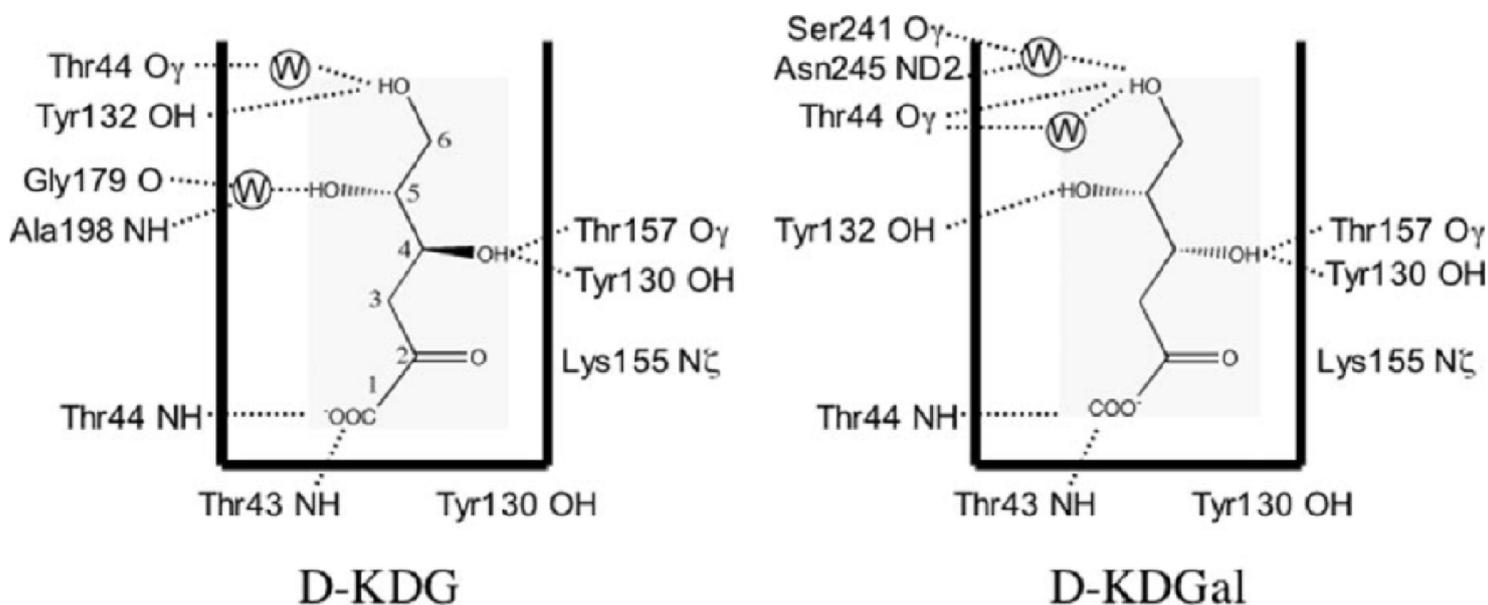


Рис.7: Субстратно-ферментное взаимодействие[2].

Остаток Asn160 цепи С выделен сервисом MolProbity как находящийся в запрещенной области карты Рамачандрана. Его RSR=0.077, а Z-score=-0,62. То есть расшифрован он достаточно хорошо. Значения углов $\phi = -175.2$, $\psi = 89.7$. При визуализации электронной плотности и сравнения со структурами, полученными с другой оптимизацией, получили, что даже с уровнем среза 2.5σ аспарагин очень хорошо списан в плотности и почти совпадает с двумя структурами с PDB-redo[5] (Рис.8). Поэтому то, что этот остаток был определен, как выброс из допустимой области карты Рамачандрана, как сервисом MolProbity[4] и EDS[3], остается для меня загадкой.

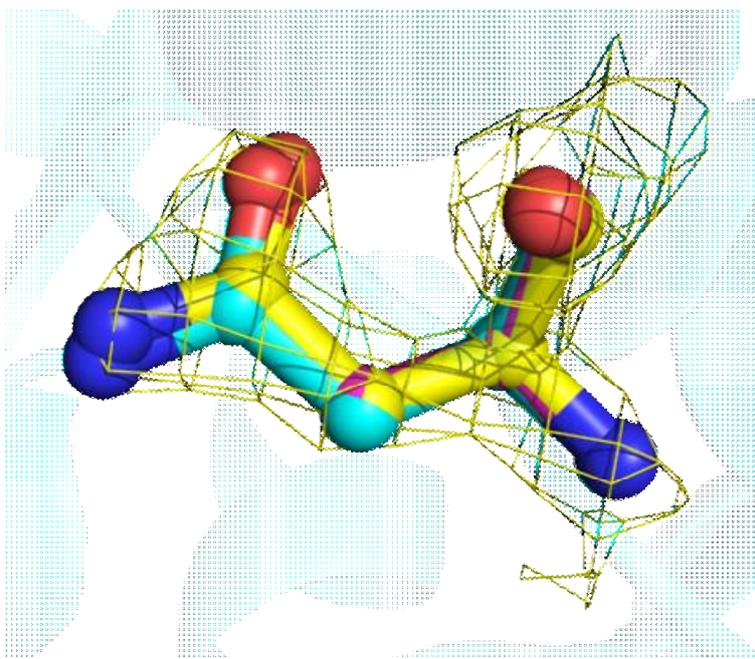


Рис.8: Изображение электронной плотности вокруг остатка Asn160 с уровнем подрезки 2.5σ . Исходный остаток изображен голубым, стандартно оптимизированная – розовым, а полная оптимизация – желтым.

Максимальное $RSR=0.421$ у аминокислотного остатка Glu294, а также у этого остатка необычно высокое значение $Z\text{-score} = 2.794$ по сравнению со значениями других аминокислот белка. При визуализации электронной плотности на уровне подрезки 2.5σ 294 остаток полипептидной цепи не имеет изображения электронной плотности. Если уменьшить уровень подрезки, то результаты вписывания модели в ЭП меняются не значительно:

- 0.5σ (Рис.9А) – визуализированная ЭП имеет пространство, в которое можно было бы вписать еще атомы, следовательно, мы видим не однозначность вписанной модели.
- 1σ (Рис.9В) – наблюдается отсутствие электронной плотности в области кислорода карбоксильной группы $\gamma\text{-C}$.
- 1.5σ (рис.9С) – модель совершенно не описывает ЭП.

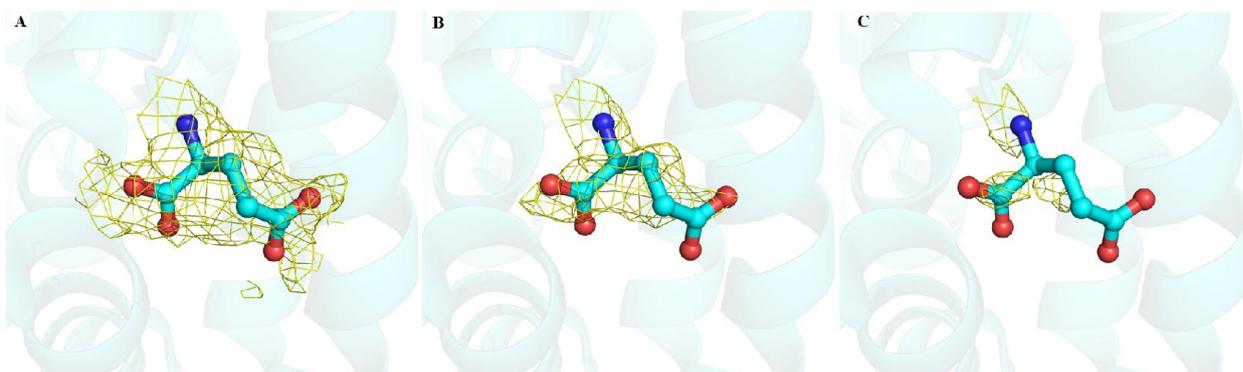


Рис.9: Изображение ЭП вокруг 294 вокруг остатка глутамина цепи А с уровнями подрезки 0.5σ (А), 1σ (В), 1.5σ (С).

Маргинальность 294 остатка глутамата объяснена тем, что это остаток на конце белка, а в процессе выполнения практикумов было выяснено, что остатки на «поверхности» белка хуже вписываются в электронную плотность.

182 остаток метионина цепи А определяется как маргинал по длине связей остова и значениям углов. При изображении ЭП вокруг этого (Рис.10) а.о. получили, что даже на уровне подрезки 0.5σ модель атомов $\delta\text{-серы}$ и $\epsilon\text{-углерода}$ не вписываются в какую-либо электронную плотность. Однако RSR данного остатка равен 0.086 , а $Z\text{-score} = -0.44$.

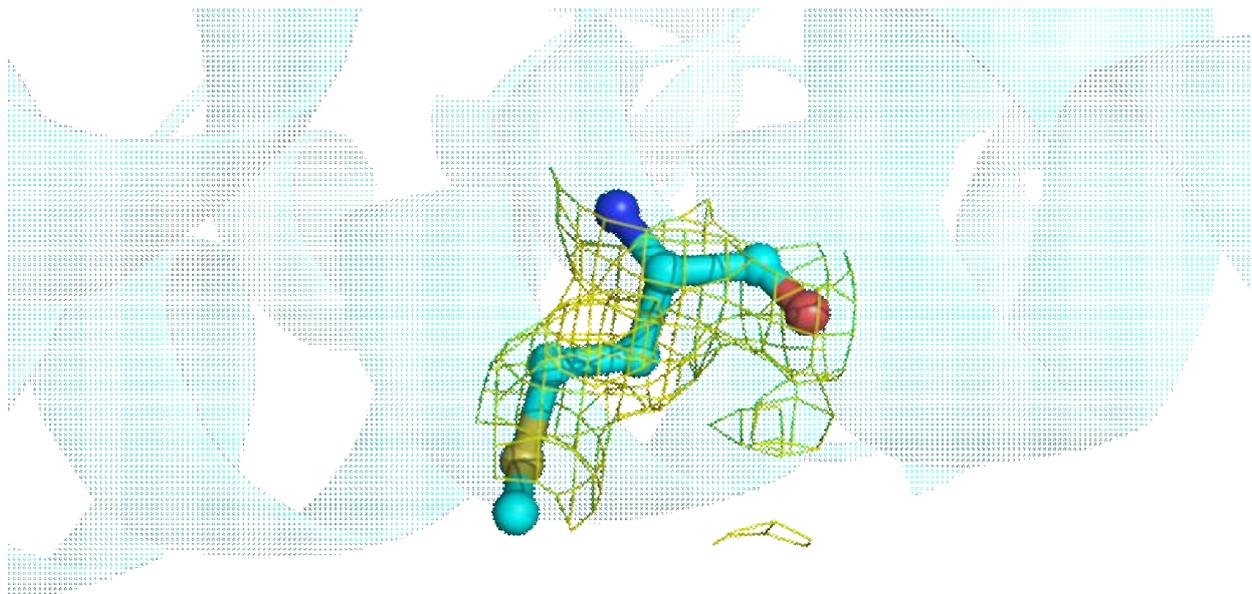


Рис.10: Визуализация ЭП вокруг 182 остатка метионина цепи А, уровень подрезки 0.5 σ .

Заключение

В связи с тем, что по различным оценкам белок содержит различающееся количество маргинальных остатков, по информации с разных сервисов, а также с тем, что авторы статьи подробно не описывали структуры белка без субстратов, нельзя быть уверенным в точности получившейся структуры. Интересно то, что некоторые маргинальные остатки очень хорошо вписываются в электронную плотность. В итоге данную структуру я бы оценила как структуру «среднего качества».

Список литературы

1. Berman, H.M., et al., *The Protein Data Bank*. Nucleic Acids Research, 2000. **28**(1): p. 235-242.
2. Theodossis, A., et al., *The structural basis for substrate promiscuity in 2-keto-3-deoxygluconate aldolase from the Entner-Doudoroff pathway in *Sulfolobus solfataricus**. J Biol Chem, 2004. **279**(42): p. 43886-92.
3. Kleywegt, G.J., et al., *The Uppsala Electron-Density Server*. Acta Crystallogr D Biol Crystallogr, 2004. **60**(Pt 12 Pt 1): p. 2240-9.
4. Davis, I.W., et al., *MOLPROBITY: structure validation and all-atom contact analysis for nucleic acids and their complexes*. Nucleic Acids Res, 2004. **32**(Web Server issue): p. W615-9.
5. ; Available from: http://www.cmbi.ru.nl/pdb_redo/w3/1w37/index.html.

