

Практикум № 4. Python для структур **и не** **ТОЛЬКО (!)**

Иосиф Финкельберг

1 Введение

В первых двух заданиях этого практикума мы будем работать со структурой FMN и NADPH-зависимой нитроредуктазы NfnV альфапротеобактерии из рода *Sphingopyxis*, полученной в работе этого года [1]. Нитроредуктаза NfnV участвует в разрушении синтетических соединений, используемых, например, в качестве гербицидов, восстанавливая их нитрогруппы (см. рис 1). Таким образом, потенциально она может использоваться для биоремедиации. Структуру нитроредуктазы NfnV интересно изучать в том числе в контексте ее субстратной специфичности. Так, замена остатка Tyr88 на Phe значительно меняет субстратную специфичность [1].

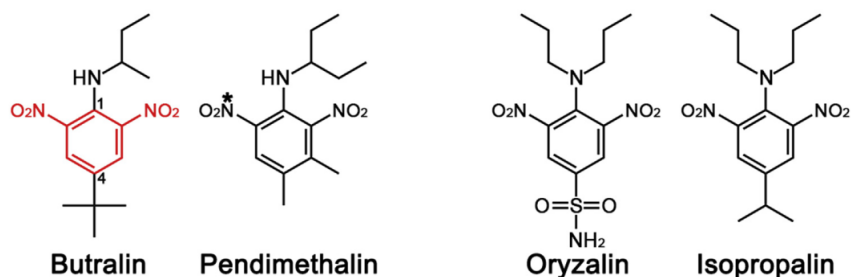


Рис. 1: Субстратами нитроредуктазы NfnV являются распространенные гербициды (рисунок взят из работы [1]).

2 Задание 1. Prody и В-факторы часть 1

Здесь мы воспользовались python-модулем **Prody** для работы с нашей структурой. Мы нашли остаток, средний В-фактор атомов которого максимален. Этим остатком оказался ALA 119 цепи В. Его средний В-фактор якобы равен 103.646. Это кажется довольно подозрительным: по умолчанию стоило бы ожидать, что В-фактор будет высоким у какого-нибудь крупного, подвижного остатка (например, остатка лизина, или ароматического

остатка фенилаланина / тирозина). В связи с этим мы посмотрели на этот остаток в `rusol`, раскрасив атомы структуры согласно их B-факторам (см. рис 2).

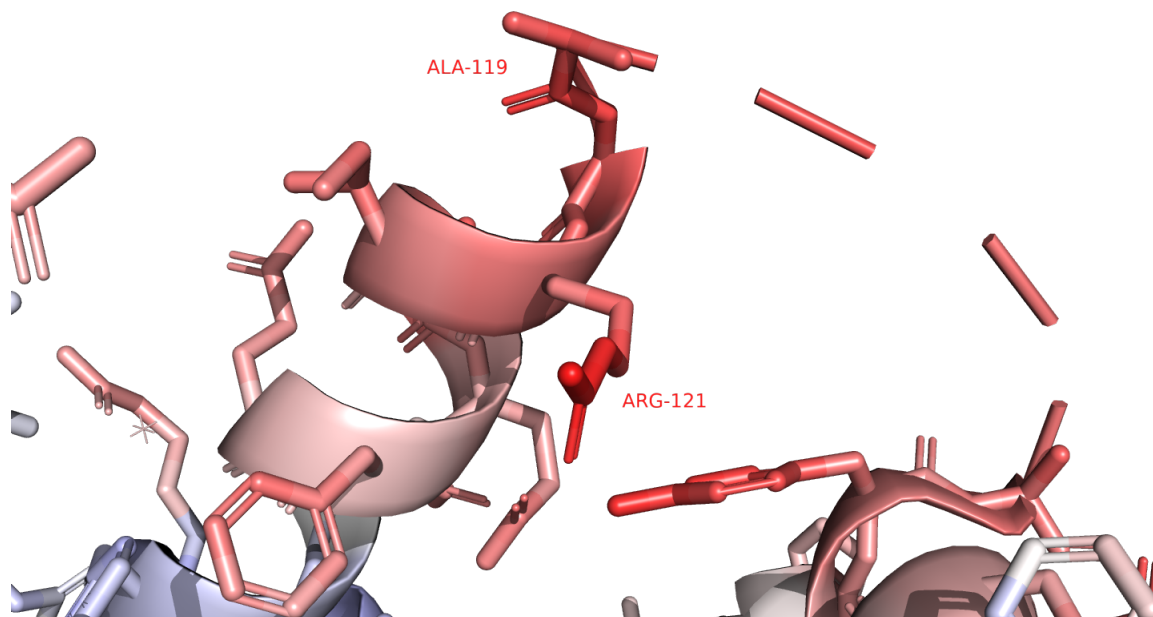


Рис. 2: B-факторы атомов структуры в районе остатка Ala-119. Чем выше B-фактор, тем краснее цвет. Пунктирной линией отмечена неразрешенная часть остова

Видим, что эта часть структуры плохо разрешена (вероятно, как раз, вследствие ее крайне высокой подвижности). Информация о том, что средний B-фактор выше всего для Ala-119 оказывается довольно бесполезной в этом случае: она лишь указывает на то, что Ala-119 находится в подвижной части остова (но не характеризует остаток как таковой). Приведем еще на всякий случай значения B-факторов для атомов Ala-119: [99.72, 99.88, 102.85, 107.24, 108.54] (см. также скриншот соответствующего `.cif` файла на рис 3)

3804	АТОМ	2483	N	N	ALA	B	1	109	?	1.54222	-41.41378	-3.23009	1.000	91.26091	?	109	ALA	B	N	1
3805	АТОМ	2484	C	CA	ALA	B	1	109	?	1.65102	-42.58527	-4.09965	1.000	94.78299	?	109	ALA	B	CA	1
3806	АТОМ	2485	C	C	ALA	B	1	109	?	2.89960	-42.50455	-4.97926	1.000	91.08615	?	109	ALA	B	C	1
3807	АТОМ	2486	O	O	ALA	B	1	109	?	2.91922	-43.00868	-6.10577	1.000	89.60429	?	109	ALA	B	O	1
3808	АТОМ	2487	C	CB	ALA	B	1	109	?	0.40380	-42.72089	-4.96500	1.000	100.84659	?	109	ALA	B	CB	1
3809	АТОМ	2488	N	N	ALA	B	1	119	?	-11.39853	-36.28934	4.27254	1.000	99.87807	?	119	ALA	B	N	1
3810	АТОМ	2489	C	CA	ALA	B	1	119	?	-10.68176	-37.46231	3.78571	1.000	102.85315	?	119	ALA	B	CA	1
3811	АТОМ	2490	C	C	ALA	B	1	119	?	-9.36257	-37.67246	4.52925	1.000	107.23913	?	119	ALA	B	C	1
3812	АТОМ	2491	O	O	ALA	B	1	119	?	-9.09280	-37.02848	5.54764	1.000	108.53769	?	119	ALA	B	O	1
3813	АТОМ	2492	C	CB	ALA	B	1	119	?	-10.42760	-37.34296	2.29008	1.000	99.72011	?	119	ALA	B	CB	1

Рис. 3: Распределение величин В-фактора для атомов остатка Ala-119

Найдем теперь, наоборот, остаток с минимальным средним по атомам В-фактором. Этим остатком оказался Gly-155 цепи А. Вы только взгляните на него, как он уютно и прочно встроился в свою альфа-спираль (см. рис 4)! Разброс значений В-фактора для его атомов невелик. Этот результат, напротив, является достаточно ожидаемым: глицин – это самая маленькая аминокислота, подвижность ее атомов определяется лишь подвижностью соответствующей части белкового остова.

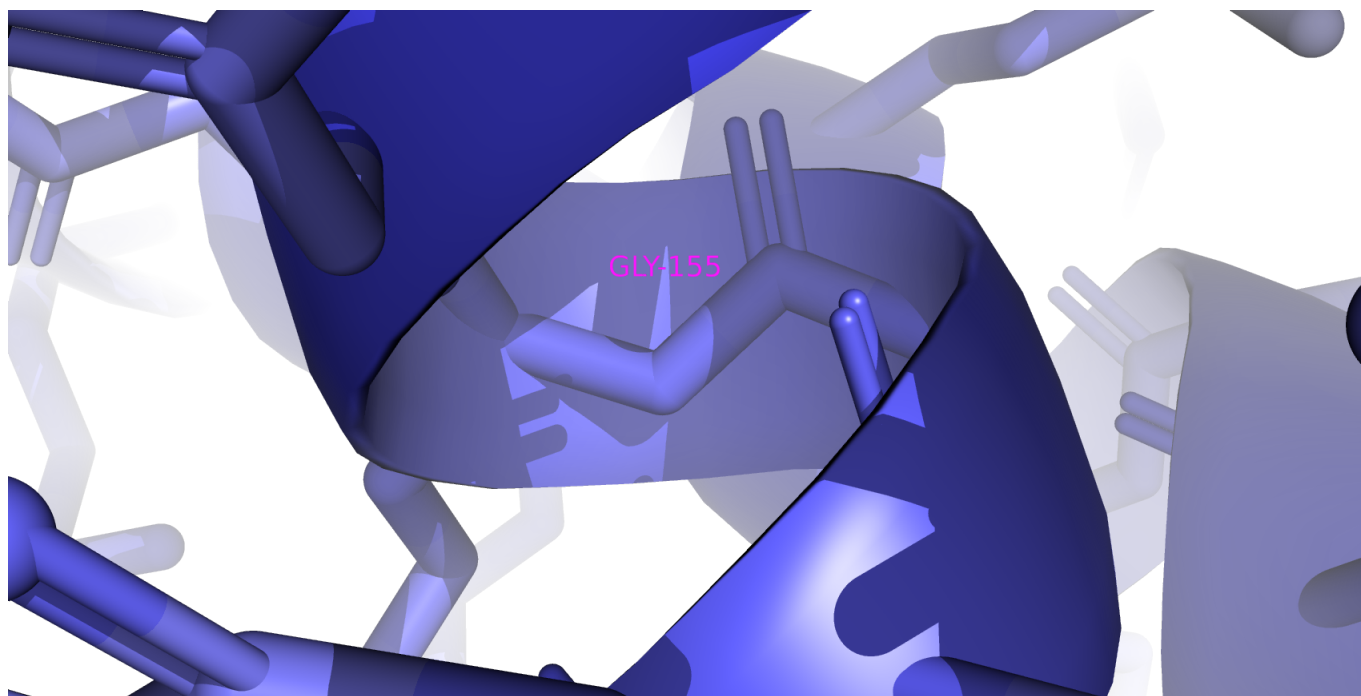


Рис. 4: Распределение величин В-фактора для атомов остатка Gly-155

Рассмотрим теперь эти два остатка в контексте структуры в целом (см. рис. 5).

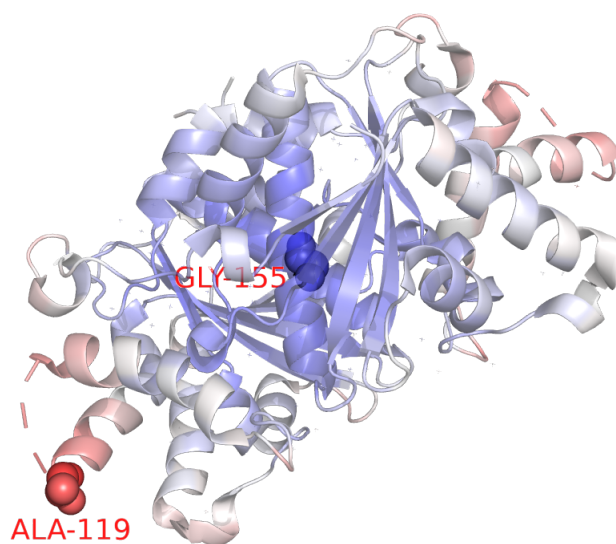


Рис. 5: Распределение величин В-фактора структуры 7dpi. Крупно показаны остатки Gly-155 (остаток находится в неподвижной, центральной области белка) и Ala-119 (этот остаток, напротив, расположен в, вероятно, очень подвижной и плохо разрешенной части структуры)

Здесь наблюдается следующее явление: подвижный остаток находится в подвижной части белка, неподвижный – в неподвижной (надо признаться, я сам был поражен, когда в первый раз это осознал). Кстати, тут, видимо, самое время рассказать, что белок является гомодимером. Центральная его часть, судя по цветовой схеме распределения значений В-фактора, является достаточно неподвижной, в отличие от периферийных участков.

3 Задание 2. Prody и В-факторы часть 2

В этом задании мы рассмотрели зависимость среднего (по атомам) В-фактора остатка от расстояния до центра масс нашего белка (см. рис 6).

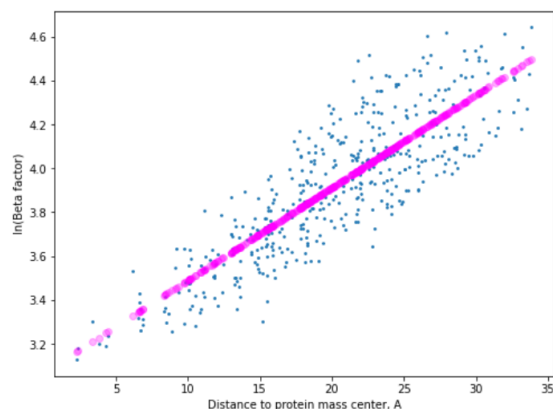


Рис. 6: Зависимость (логарифма) среднего В-фактора остатка от расстояния до центра масс белка. Розовая кривая – линейное приближение.

Мы видим, что в нашем белке средняя величина В-фактора остатка положительно скоррелирована с расстоянием до центра масс. Более того, эта зависимость хорошо аппроксимируется экспонентой (см. рис 6). Кроме того, чем больше расстояние, тем больше дисперсия у средних значений В-фактора. Качественно эта зависимость хорошо согласуется с тем, что мы наблюдаем глазами в *rubol* (см. рисунок 5). Можно сделать вывод, что центральная, массивная часть нашего белка является достаточно жесткой, а вот периферийные участки подвижны. То, что возрастание В-фактора при удалении от центра масс имеет именно экспоненциальную природу, в принципе, можно было бы объяснить с помощью совсем тривиальной модели. Пусть белок – это плоский объект, состоящий из скрепленных друг с другом жестких стержней одинаковой длины; соседние стержни могут поворачиваться друг относительно друга на углы в каком-то диапазоне (т.е. белок – это ломаная прямая на плоскости, состоящая из соединенных друг с другом отрезков длины 1, между двумя соседними отрезками угол должен лежать в диапазоне $[\pi - \alpha : \pi + \alpha]$ с каким-то небольшим фиксированным значением α). Пусть мы закрепили левый конец этой ломаной, и первые несколько ее отрезков, а остальным дали свободно варьировать в допустимом диапазоне углов между соседними отрезками. Тогда конец правого подвижного отрезка будет бегать по дуге 2α , правый конец следующего за ним отрезка, соответственно, по дуге 4α , ..., правый конец n -го подвижного отрезка будет бегать, соответственно, по дуге $2^n\alpha$. В-фактор для каждой вершины будет пропорционален длине дуги, по которой она бегаёт. Таким образом, В-фактор будет экспоненциально зависеть от расстояния до неподвижной части ломаной.

Интуитивно, кажется, что эта зависимость – это не какой-то фундаментальный закон природы, так как она определяется формой и свойствами белка, которые должны сильно варьировать. Можно предположить, что по-

хожую зависимость мы будем видеть для таких белков, у которых в центре масс структура является достаточно ригидной, а при удалении от центра будет иметь какую-то \pm - константную "гибкость" (дальше, понятно, можно попытаться более строго эти свойства структуры определить).

4 Задание 3. Как работает восстановление функции электронной плотности по экспериментальным данным

Как это кратко и точно изложено в задании к практикуму: "В обычном мире, мы делаем кристалл, который состоит из молекул имеющих электроны, которые распределены так, что мы можем описать их электронной плотностью. Эту электронную плотность мы хотим узнать в ходе рентгеноструктурного анализа. Проводим эксперимент, получаем амплитуды, берем откуда-то фазы, проводим синтез Фурье и восстанавливаем электронную плотность, надеясь, что эта полученная "восстановленная" электронная плотность максимально близка к реальной". В этом задании мы решили слегка изменить "обычный мир": здесь мы упражнялись в работе со скриптами, написанными А.А.Алексеевским. Эти скрипты моделируют одномерную вселенную, укладываемуюся в отрезок от 0 до 30 ангстрем. Вселенная заполнена пятью атомами, в нашем случае: Н(3), С(4.5), О(9), С(10.5), О(16), где в скобках указаны координаты атомов на отрезке. В роли природы (/окружающего мира/бога и тд) выступают, собственно, плоды творчества А.А.Алексеевского: они генерируют распределение "настоящей" (см. assumptions нашей модели в задании к практикуму) электронной плотности для нашей вселенной, а также вносят в наш эксперимент погрешности измерения. В роли кристаллографа выступаем мы.

Сперва мы предположили, что в эксперименте нам удалось получить полный набор гармоник (с нулевой по i -ю). Мы варьировали значение i – номер последней разрешенной гармоники, и следили за тем, как меняется точность предсказания нашей модели. На рис. 7 и в таблице 1 приведены результаты. Видим, что для отличного качества (по графику восстановленной функции можно определить положение максимума всех гауссовых слагаемых функции ("атомов")), достаточно первых тридцати гармоник.

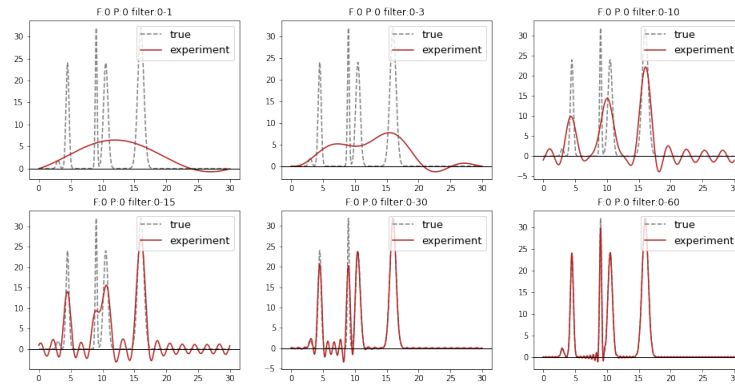


Рис. 7: График восстановленной функции по полному набору гармоник с $n = 0, 1, \dots$. Здесь и далее над каждым рисунком в поле F указана погрешность, вносимая в значение амплитуды ряда Фурье, описывающего ЭП нашей вселенной; в поле P указана погрешность, вносимая в значение фазы; в поле filter – указаны взятые нами гармоники. Здесь и далее серой пунктирной линией указана "настоящая" ЭП вселенной, красной – восстановленная по нашей модели ЭП.

Затем мы решили внести шум в данные нашего эксперимента, изменив значения фаз и модулей ряда Фурье, приближающего ЭП нашей вселенной (см. рис 8 и таблицу 1). Видим, что изменение фазы вносит больший (отрицательный) вклад в качество модели по сравнению с изменением амплитуды. Надо заметить, впрочем, что и в том, и в другом случае качество заметно ухудшилось по сравнению с незашумленными данными (возможно, мы переборщили с абсолютными значениями вносимого нами отклонения модуля / фазы от их истинных значений).

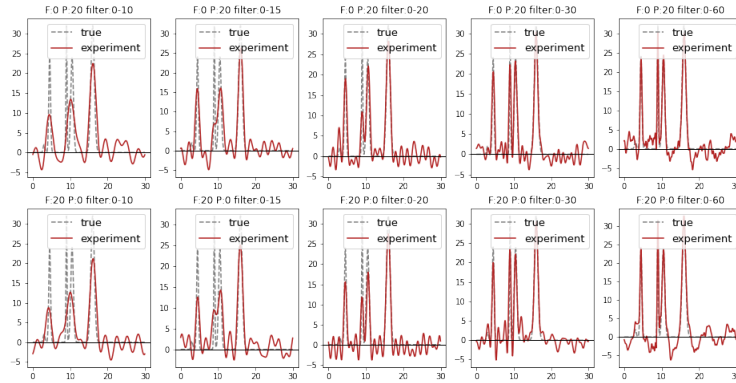


Рис. 8: График восстановленной функции по полному набору гармоник с $n = 0, 1, \dots$. Дополнительно внесен шум (на верхней панели внесена погрешность в значение фазы, на нижнем – в значение амплитуды.)

Наконец, мы провели исследование того, как влияет на качество модели наличие лишь некоторых гармоник. Мы приблизительно случайно исключали некоторые гармоники из их полного набора и затем рисовали восстановленную по нашей модели ЭП (см. рис 9 и таблицу 1). Мы видим, сильную скоррелированность качества нашей модели и полноты данных. При этом формальное "разрешение" (посчитанное по количеству имеющихся гармоник) очень слабо отражает реальное качество модели: особенно, если пропущенным оказываются гармоники с высокими номерами. Можно сделать вывод о том, что качество хорошо оценивается комбинацией параметров: полнотой данных + гармоникой с самым низким из имеющихся номером (более подробно эти моменты разбирались на коллоквиуме, так что, если можно длинные рассуждения здесь опущу).

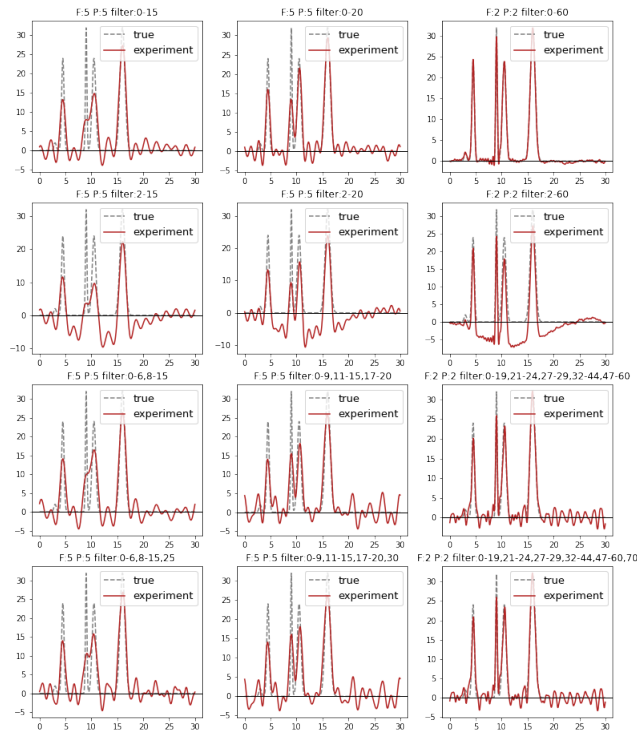


Рис. 9: График восстановленной функции по неполному набору гармоник.

Таблица 1. Восстановление функции по коэффициентам ряда Фурье.

(числа написаны для примера представления данных; в вашем конкретном случае будут другими)

Набор гармоник	Разрешение (Å)	Полнота данных (%)	Шум амплитуды (% от величины F)	Шум фазы (% от величины phi)	Качество восстановления (отличное, хорошее, среднее, плохое)	Комментарии
Полный набор гармоник						
0-1	30 Å	100%	0	0	Плохое	В скрипте func2fourier.ru идиотская опечатка на 42 строке, исправьте, пожалуйста!
0-3	10 Å	100%	0	0	Плохое	
0-10	3 Å	100%	0	0	Среднее	
0-15	2 Å	100%	0	0	Хорошее	
0-30	1 Å	100%	0	0	Отличное	
0-60	0.5 Å	100%	0	0	Отличное	
ШУМ						
0-10	3 Å	100%	0	20	Плохое	
0-15	2 Å	100%	0	20	Плохое	
0-20	1.5 Å	100%	0	20	Среднее	
0-30	1 Å	100%	0	20	Среднее	
0-60	0.5 Å	100%	0	20	Среднее	
0-10	3 Å	100%	20	0	Плохое	
0-15	2 Å	100%	20	0	Плохое	
0-20	1.5 Å	100%	20	0	Среднее	
0-30	1 Å	100%	20	0	Среднее	
0-60	0.5 Å	100%	20	0	Хорошее	
Неполный набор гармоник						
2-15	3 Å	87%	5	5	Среднее	
2-20	1.5 Å	90%	5	5	Среднее	
2-60	0.5 Å	97%	2	2	Отличное (даже водород, и тот виден!)	
0-6,8-15	3 Å	93%	5	5	Средненько (на троечку)	
0-9,11-15,17-20	1.5 Å	90%	5	5	Средненько, да и то, как-то жутковато	

0-19,21-24,27-29,32-44,47-60	0.5 Å	90%	2	2	Неплохо, при желании водород угадывается	
0-6,8-15,25	1.2 Å	56%	5	5	Средненько (на троечку)	То же самое: внесение одной малой гармоник не повлияло на качество
0-9,11-15,17-20,30	1 Å	60%	5	5	Средненько, да и то, как-то жутковато	То же самое: внесение одной малой гармоник не повлияло на качество
0-19,21-24,27-29,32-44,47-60,70	0.43 Å	76%	2	2	Неплохо, при желании угадывается водород	То же самое: внесение одной малой гармоник не повлияло на качество

5 Сессии PyMol

<https://kodomo.fbb.msu.ru/~osyafinkelberg/term7/prac4/session0.pse>

6 ССЫЛКИ

[1] Kim et al, 'Crystal structure of FMN and NADPH-dependent nitroreductase NfnB mutant Y88A derived from sphigopyxis sp. strain HMH' 2021 Journal of biological chemistry