

## Практикум № 4. Python для структур и не ТОЛЬКО

В данном практикуме я рассматриваю белок Тирозил-тРНК синтазу *E.coli*. Белок состоит из двух субъединиц, одинаковых в природе. Каждая субъединица состоит из центральной области с активным центром, в котором связывается тирозин, а также подвижного "шарика" который в свободном состоянии связан с центральной областью через высокоподвижный линкерный пептид. При связывании молекула тРНК своими листками как раз заякоривается на этот шарик, а ее 3'-конец оказывается в активном центре (рис. 2А).

В рассматриваемой структуре, один из "шариков" оказывается фиксирован (у него низкий b-фактор и хорошо разрешена электронная плотность), а второй шарик менее упорядочен в кристалле (у него высокий b-фактор и плохо разрешена электронная плотность). Цепи в структуре имеют специально введенные точечные мутации и различны, возможно авторы сделали это специально чтобы получить один из "шариков" фиксированным. Возможно также, что это результат кристаллизации.

Далее в ходе практикума я буду пользоваться описанными тут особенностями структуры.

### 1 Prody и B-факторы часть 1

Средний B-фактор максимален у остатка PHE'397 цепи A. Значения у отдельных атомов варьируются от 96.44 до 119.84, составляя в среднем 110.02.

Остаток находится как раз в незафиксированном подвижном "шарике" молекулы белка, причем остаток фенилаланина экспонирован в раствор. Не удивительно, что это очень подвижный остаток с большим количеством возможных положений радикала. Разброс значений для атомов - 25 единиц, то есть как остов, так и боковой радикал подвижны.

Средний B-фактор минимален у остатка GLY'208 цепи B. Значения у отдельных атомов варьируются от 19.94 до 22.19, составляя в среднем 20.675.

Этот остаток находится в коровой части белка близь активного центра. Во-первых он входит в состав альфа-спирали, поэтому его остов зафиксирован водородными связями, нужными для образования структуры. Во вторых у глицина нет бокового радикала, и все его атомы плотно уложены

во вторичную структуру. Разброс значений фактора для них - 2 единицы, в 10 раз меньше, чем для предыдущего остатка, что не удивительно - его остов плотно зафиксирован.

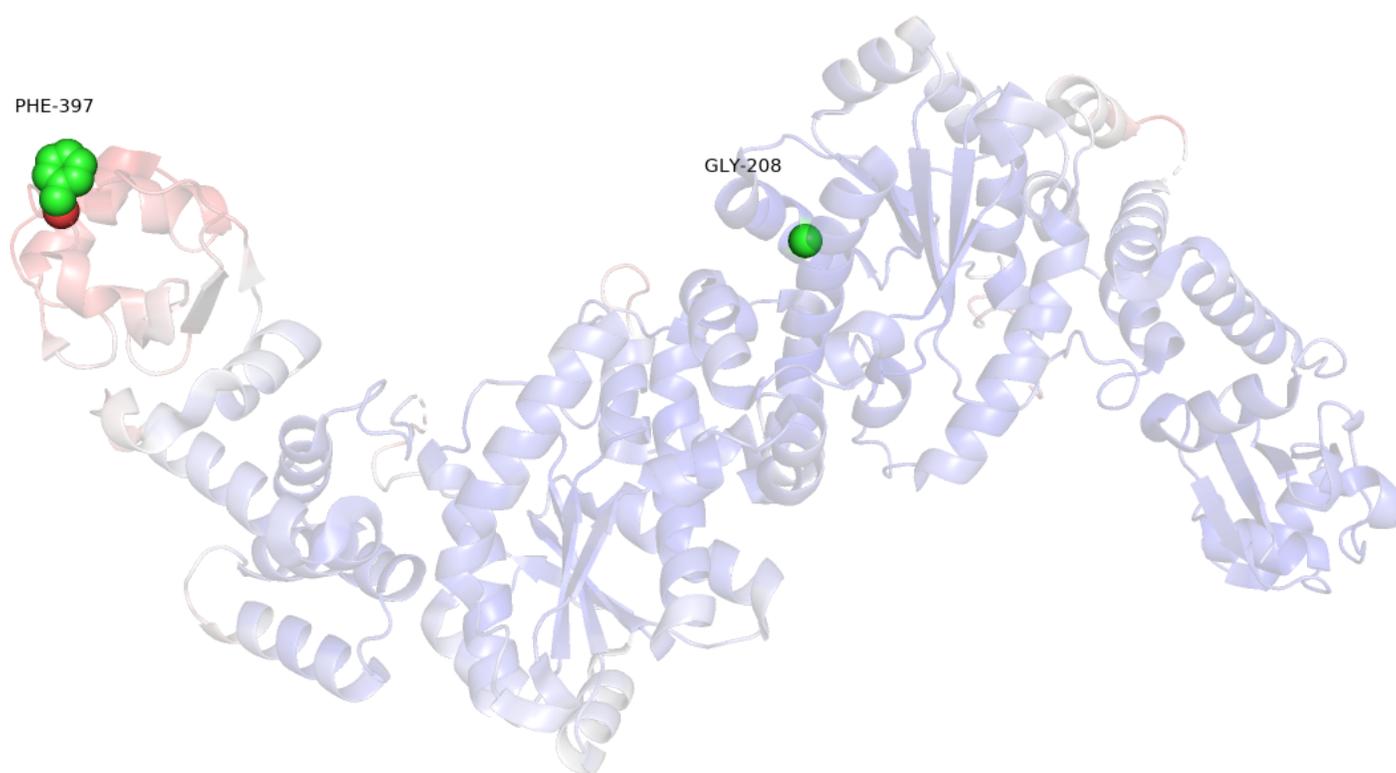


Рис. 1: XXX

## 2 Prody и B-факторы часть 2

Здесь я предлагаю вспомнить описание белка и разделить его на 6 частей.

Красные и розовые - подвижные "шарики" зеленый и желтый - кор с активными центрами, рыжее и синее - промежуточные области. По сути мы разделили наш белок на подобия доменов.

Для каждого домена вычислим свой центр масс и будем измерять зависимость подвижностей остатков от расстояния до центра масс домена. Это имеет смысл, как мы теперь знаем из лекции про домены, так как домен как раз можно определить как самостоятельную подвижную единицу белка, иногда со своим собственным гидрофобным ядром.

Чем аппроксимировать зависимость не очень понятно, однако посмотрев на scatterplot для доменов A3 и B3 - то есть те, где точек больше всего, я решила, что больше всего это похоже либо на экспоненту, либо на какую то степень.

Я думаю, что квадратичная зависимость тут была бы уместна: если считать, что мы измеряем подвижность относительно некоторой центральной точки, которая в пространстве неподвижна. При этом мы измеряем подвижность точки, связанной отрезком с этой центральной неподвижной точкой. Понятно, что это очень сильное упрощение - у нас не точка, а остаток, и остаток совсем сложно связан с центром домена, но можно предположить, что в жестком гидрофобной ядре центр остатка на фиксированном расстоянии от центра домена. Получается что мы рассматриваем модельку кручения домена вокруг своего центра, и забываем про подвижные боковые группы на границе.

Получается что у нас есть точка, движущаяся по сфере, а площадь сферы зависит как квадрат от радиуса сферы.

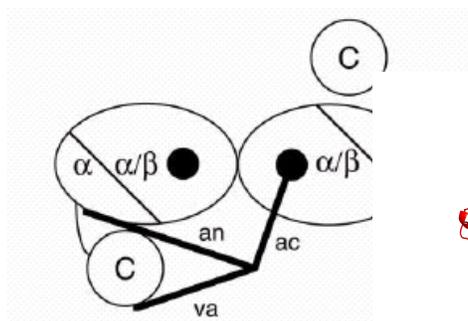
Уравнение:

$$B_{factor} = A + B * dist^2$$

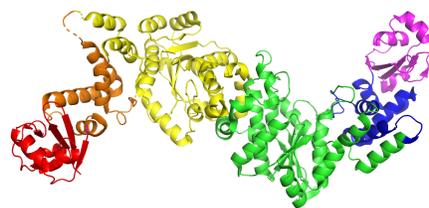
Коэффициенты для доменов:

область	цвет	A	B
A1	pink	81.02	0.0256
A2	blue	35.98	0.0523
A3	green	29.76	0.0237
B1	red	23.41	0.0644
B2	orange	25.14	0.0779
B3	yellow	22.09	0.0410

Самая плохая аппроксимация для домена A1. Оно и понятно - для него практически нет высокоточной ЭП (см практикум 2), точек немного, и самый большой тут дополнительный член B - и вообще значения B-фактора тут самые высокие. Для остальных доменов член A на удивление схож - где то около 25 (домен A2 ближайший к A1, думаю 35 у A2 тут обусловлено подвижностью A1). Член B немного отличается, хотя тоже, как мне кажется, сходимость у него достаточно хорошая - его значения от 0.02 до 0.08, так что можно считать модельку приемлимой. Цветовой код:



(a) Упрощенная модель белка. Черные точки - активные центры. Черной жирной линией показана тРНК



(b) Цвета остатков при разделении на группы

Рис. 2: Разделение остатков белка на "домены"

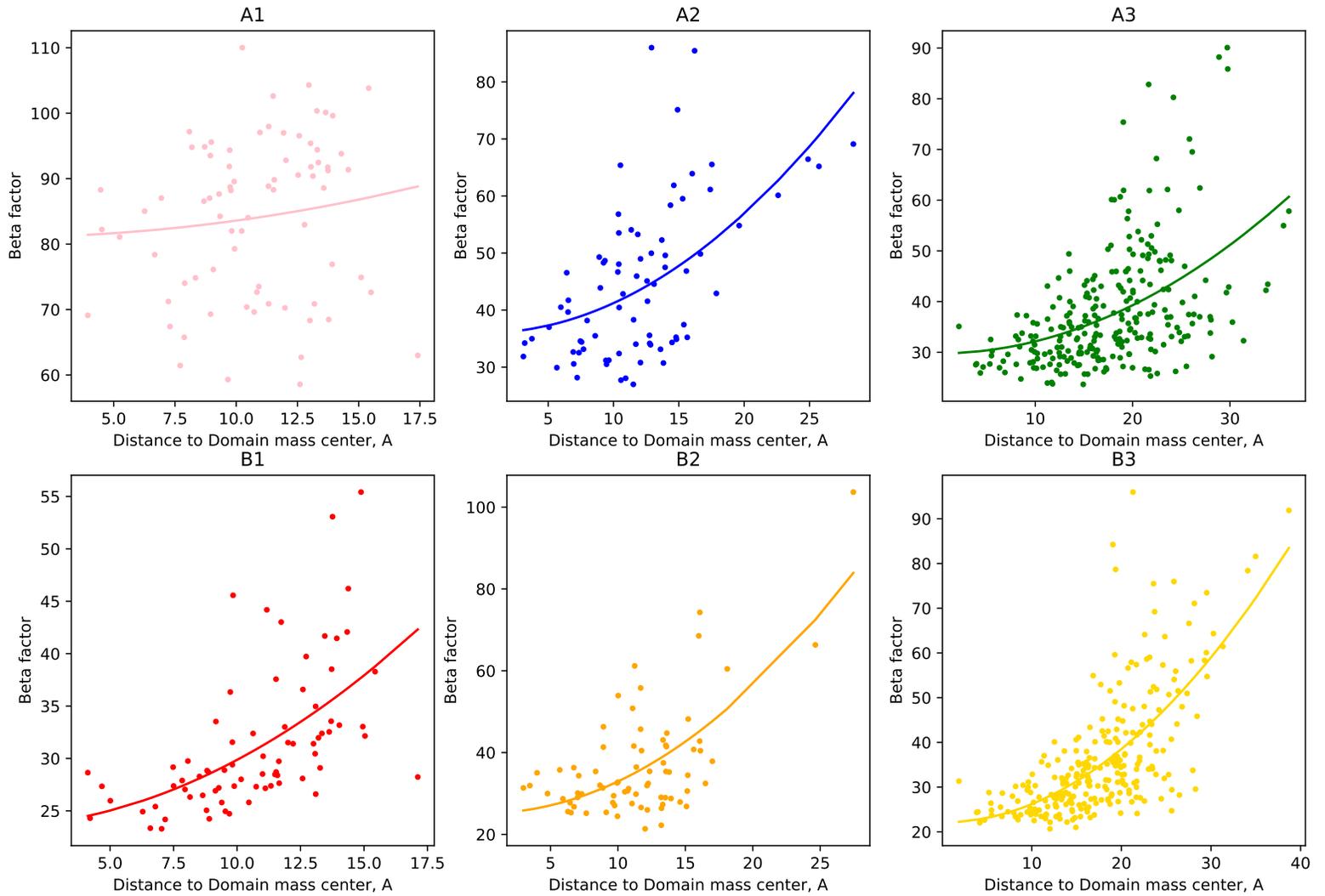


Рис. 3: Линейные аппроксимации для каждого из доменов. Цвета соответствуют рис. 2

Интересно на графиках, что высокоподвижные точки находятся сильно выше теоретической кривой. Возможно, это те самые подвижные боковые

группы на поверхности жесткой глобулы!!

Конечно, лучше всего аппроксимация видна для доменов №3, в которых больше всего точек, для остальных не так очевидно, что происходит.

### **3 Восстановление функции электронной плотности по экспериментальным данным**

Я решила симитировать молекулу воды и C=O группу. Между атомами Н и О расстояние в 1 ангстрем, между C=O - полтора. Высоту пиков я задала равной числу электронов у атома умноженную на 2.

## 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
<b>Полный набор гармоник</b>						
0-1	30 Å	100%	0	0	Плохое, не понятно даже положение молекул	
0-3	10 Å	100%	0	0	Плохое, хотя угадывается расположение молекул	
0-10	3 Å	100%	0	0	Средне-плохое: точно определяем положение молекул, в которых есть хотя бы один тяжелый атом	
0-15	2 Å	100%	0	0	Среднее: тяжелые атомы уже различимы	
0-30	1 Å	100%	0	0	Отличное: различимы все атомы	
0-60	0.5 Å	100%	0	0	Отличное: различимы все атомы, шума вообще нет	
<b>Шум</b>						
0-10	3 Å	100%	0	20	Средне-плохое: точно определяем положение молекул, в которых есть хотя бы один тяжелый атом	Рис. 5 первый ряд
0-20	1.5 Å	100%	0	20	Среднее: различаем тяжелые атомы	
0-30	1 Å	100%	0	20	Среднее: различаем тяжелые атомы	
0-10	3 Å	100%	20	0	Средне-плохое: точно определяем положение молекул, в которых есть хотя бы один тяжелый атом	Рис. 5 второй ряд
0-20	1.5 Å	100%	20	0	Среднее: различаем тяжелые атомы	
0-30	1 Å	100%	20	0	Хорошее: можно угадать водороды	
<b>Неполный набор гармоник</b>						
2-15	3 Å	87%	5	5	Средне-плохое	Рис. 6, второй ряд
2-20	1.5 Å	90%	5	5	Средне-хорошее (можно угадать один водород)	
2-60	0.5 Å	97%	2	2	отличное	

0-6,8-15	3 Å	93%	5	5	Средне-плохое	Рис 6, третий ряд
0-9,11-15,17-20	1.5 Å	90%	5	5	среднее	
0-19,21-24,27-29,32-44,47-60	0.5 Å	90%	2	2	хорошее	
0-6,8-15,25	1.2 Å	56%	5	5	Средне-плохое	Рис 6, четвертый ряд
0-9,11-15,17-20,30	1 Å	60%	5	5	среднее	
0-19,21-24,27-29,32-44,47-60,70	0.43 Å	76%	2	2	хорошее	

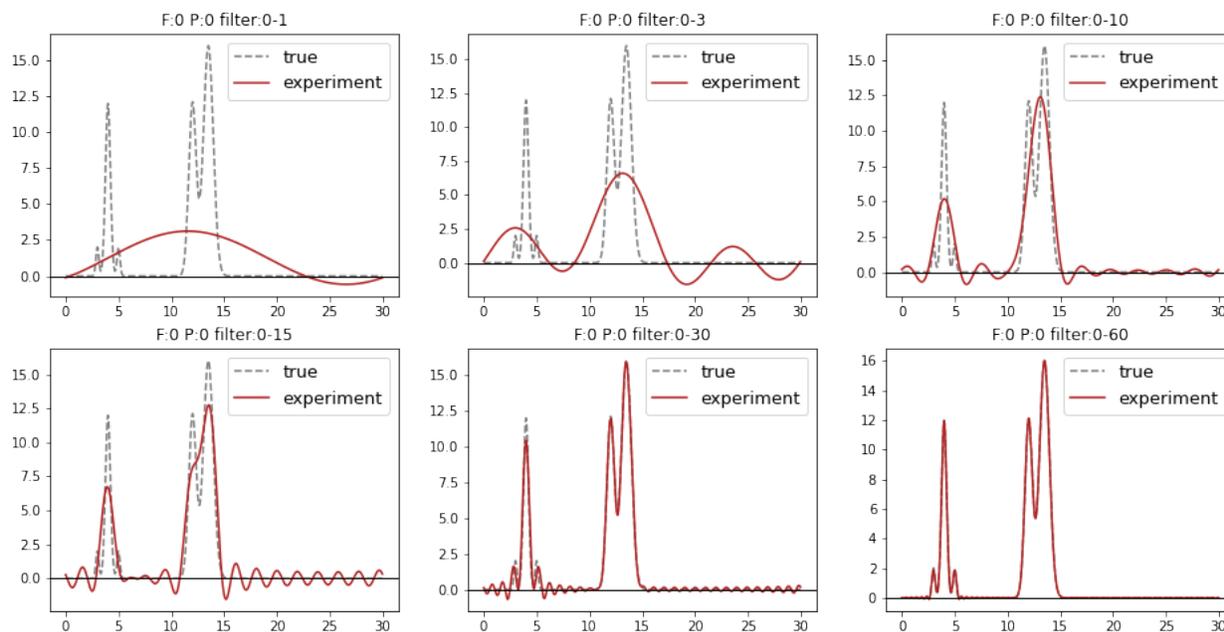


Рис. 4: Подбор наибольшей гармоники.

На рис. 4 видно, что достаточно хорошее разрешение для различения водородов - только 1 ангстрем (середина, нижний ряд). При этом предположить различие тяжелых атомы на расстоянии 1.5 ангстрем (C=O группа), становится возможным только при разрешении около 2 ангстрем (слева снизу). Отдельные молекула в принципе видны на 10 ангстремах, но с плохо отделимым шумом.

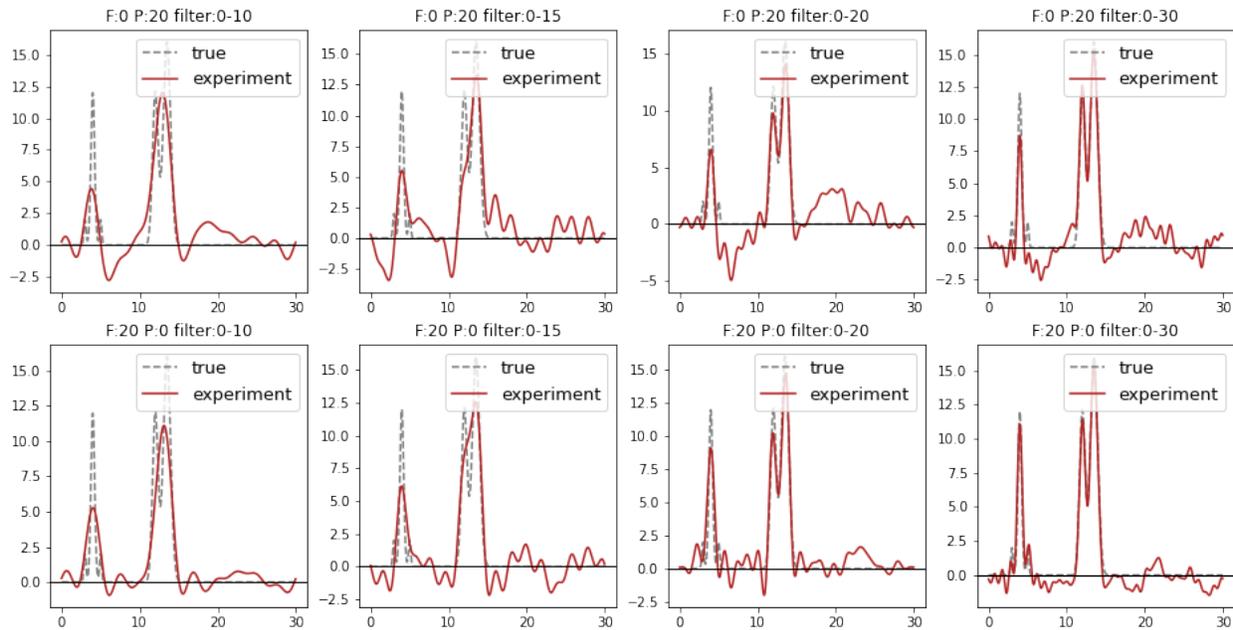


Рис. 5: XXX

На рис. 5 сверху находятся зашумленные фазы, а снизу - зашумленные амплитуды. Я предлагаю обратить внимание на третью колонку, соответствующую разрешению в 1.5А. При зашумлении амплитуд (снизу) водороды еще можно угадать, а вот при зашумлении фаз угадать водороды становится невозможно. Таким образом, можно представить, насколько вообще важны фазы для РСА.

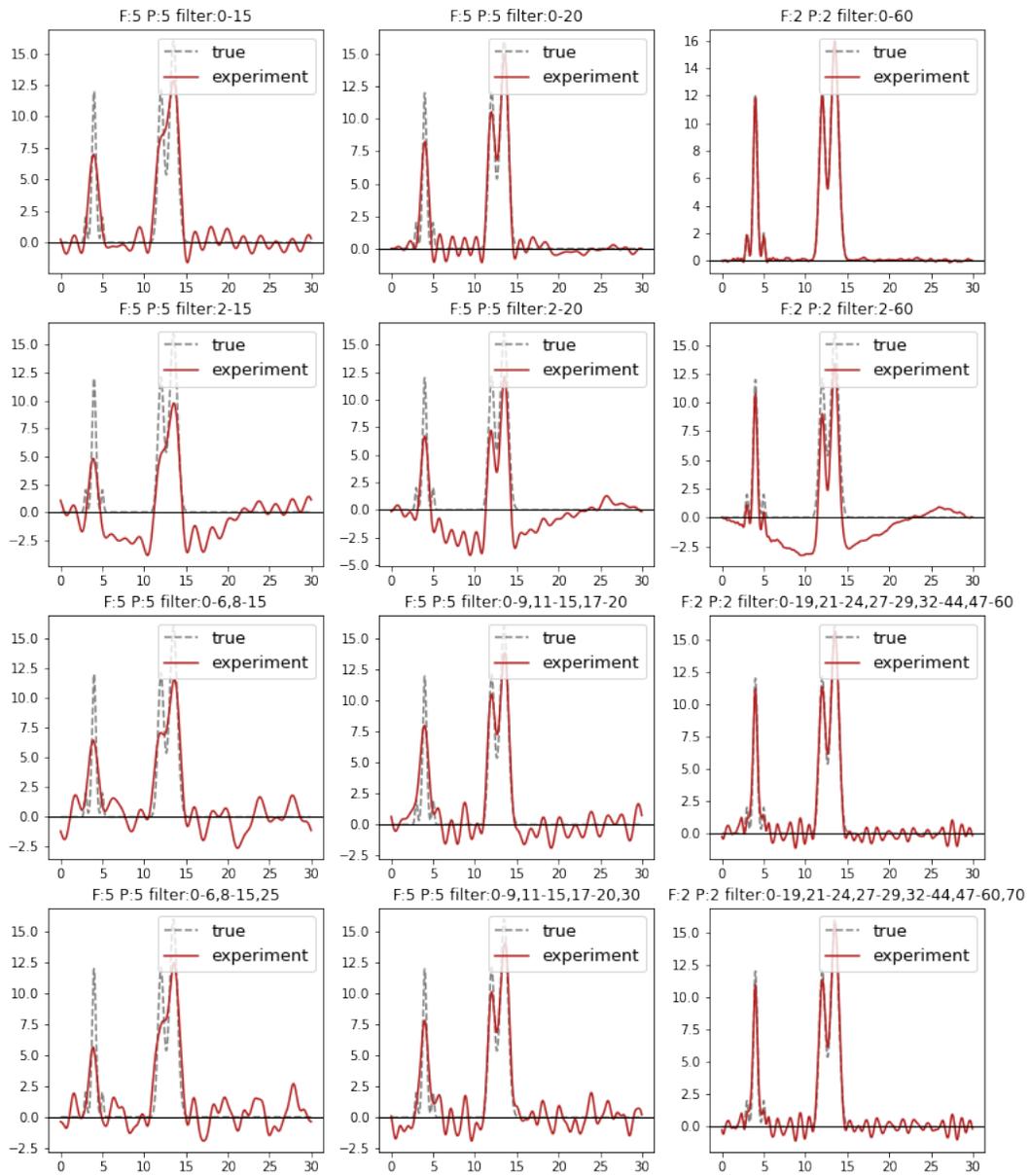


Рис. 6: XXX

На рис. 6 в верхнем ряду все фазы для разрешения в 2, 1.5 и 0.5 ангстрем с небольшими зашумлениями. Во втором ряду убраны первые две гармоники. Видно, что это придает графику странный вид, но при этом различить отдельные тяжелые атомы на среднем рисунке или водороды на третьем вполне можно. Так что не очень страшно для хорошего разрешения, что самые большие гармоники не были зарегистрированы.

В третьем ряду из середины исключены некоторые гармоники. Видно, что водороды на третьем рисунке можно только угадать, так что качество ухудшается при пропаже гармоник из середины и ухудшем предсказании.

В последнем ряду к предыдущему эксперименту добавлена гармоника с высоким номером. Видно, что ряд 3 от ряда 4 практически не отличаются. В третьей колонке водороды по-прежнему очень плохо угадываются. И это при формальном разрешении в 0.4 ангстрем!

Итого, даже если была зарегистрирована какая-нибудь очень высокая гармоника, но при этом следующая от нее отстоит далеко, то это еще ничего не значит, и показатель полноты данных очень важен при анализе качества.

Итого, отвечая на вопрос, "как определять разрешение для набора гармоник Фурье, по которым восстанавливается функция?" я бы предложила отсечку по процентам полноты - например, 90%, и исходя из этой отсечки называла бы последнюю гармонику.