

Практикум 4

Задание 1. Prody и B-факторы, часть 1.

В этом задании предлагается использовать python пакет prody для анализа распределения B-факторов и поиска разных закономерностей.

PDB ID: 7b9o

Нужно было найти аминокислотные остатки с максимальным/минимальным средним значением B-факторов.

```
==== МАКСИМАЛЬНЫЙ СРЕДНИЙ B-ФАКТОР ====
```

```
Остаток, средний B-фактор атомов которого максимален: ASP444, его B-фактор равен 96.73  
Атом с максимальным значением B-фактора в остатке: OD1, его B-фактор равен 101.28  
Атом с минимальным значением B-фактора в остатке: N, его B-фактор равен 92.74  
Стандартное отклонение значений B-фактора для остатка: 2.79
```

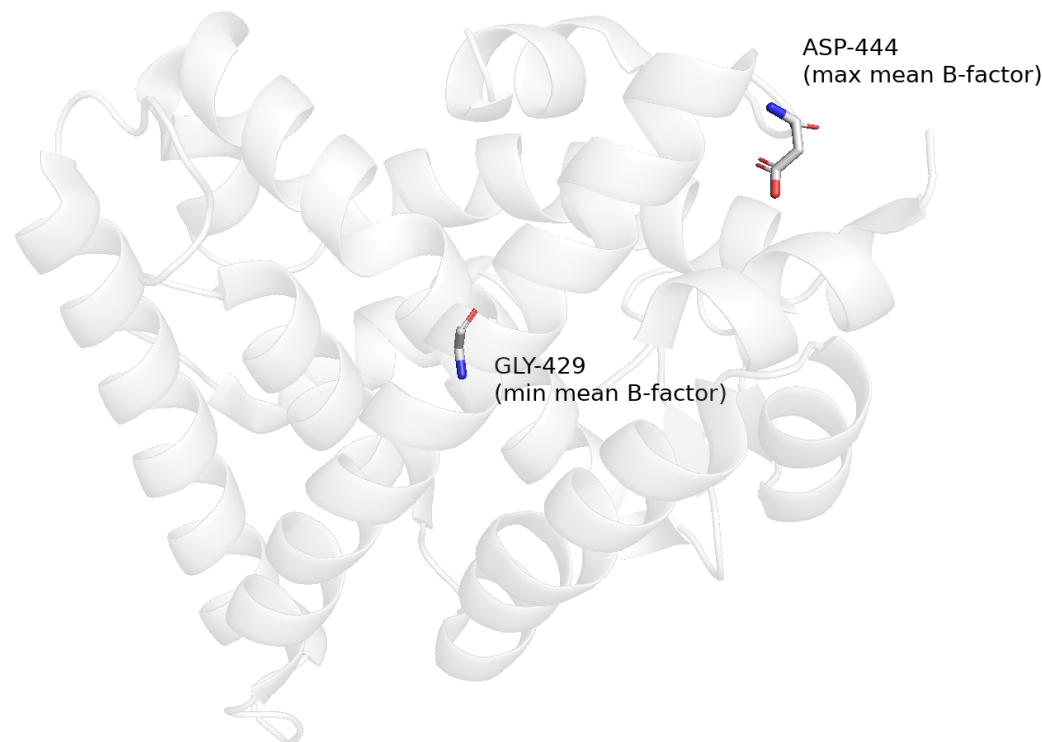
```
==== МИНИМАЛЬНЫЙ СРЕДНИЙ B-ФАКТОР ====
```

```
Остаток, средний B-фактор атомов которого минимален: GLY429, его B-фактор равен 46.58  
Атом с максимальным значением B-фактора в остатке: N, его B-фактор равен 47.61  
Атом с минимальным значением B-фактора в остатке: C, его B-фактор равен 45.59  
Стандартное отклонение значений B-фактора для остатка: 0.83
```

Видно, что разброс значений B-фактора в остатке с максимальным средним значением B-фактора намного больше, чем с минимальным. Это может объясняться тем, что все атомы несложного по строению глицина примерно одинаково статичны в составе остова белка, в то время как аспартат подвижен в целом почти полностью.

Остаток с максимальным значением Beta-фактора

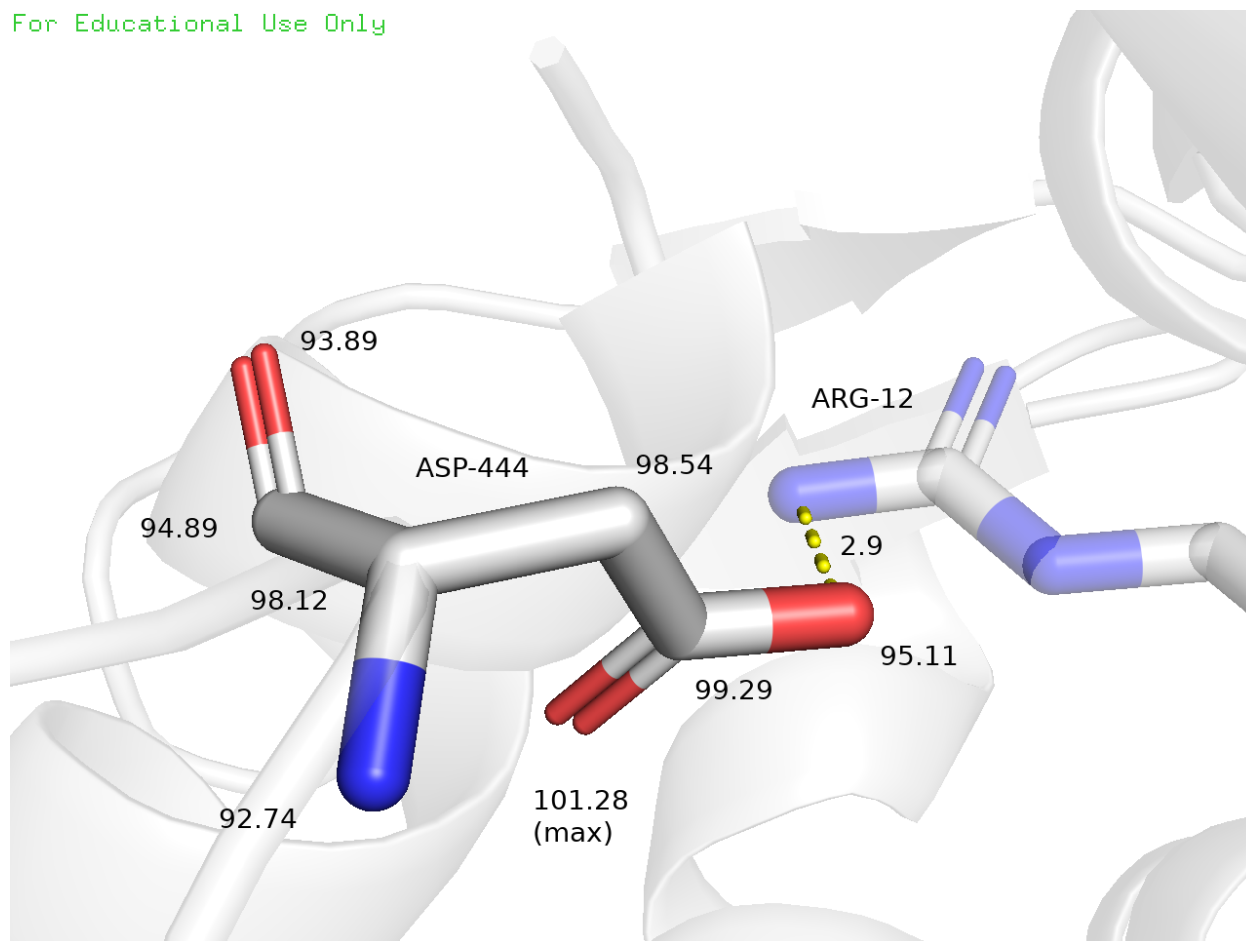
For Educational Use Only



Структура белка с выделенными аминокислотами с минимальным и максимальным средним значением B-фактора.

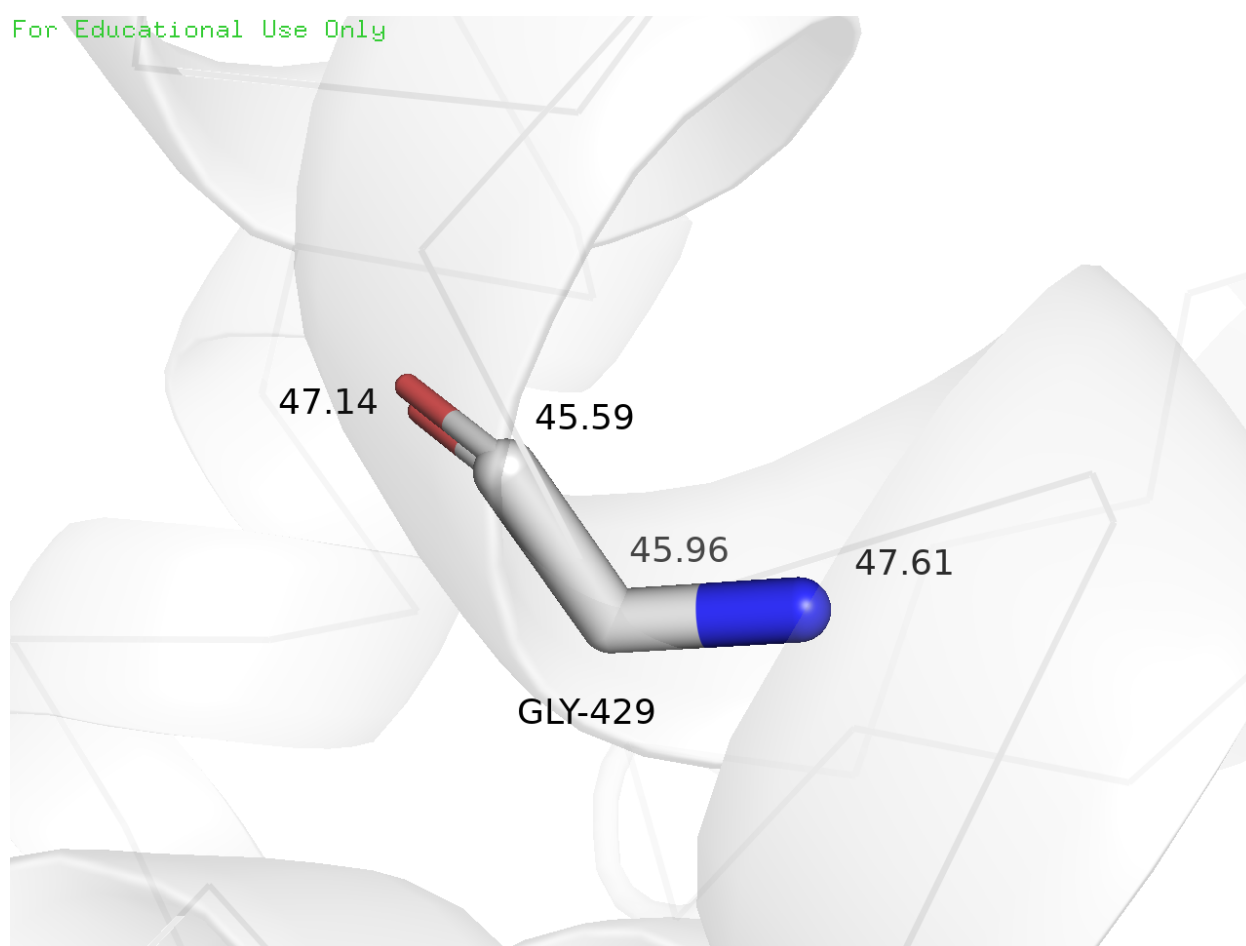
Ожидается, что аминокислота с максимальным средним значением B-фактора находится на поверхности и обращена вовне белка, чем и объясняется ее высокая подвижность. Аминокислота с минимальным средним значением B-фактора находится внутри альфа-спирали.

For Educational Use Only



Аминокислота с максимальным средним значением В-фактора, возможно, все же формирует одну водородную связь, но мы не можем быть уверены, потому что значение В-фактора для возможно контактирующего атома велико, и он может находиться дальше от остатка аргинина. (числами подписаны значения В-факторов атомов)

For Educational Use Only

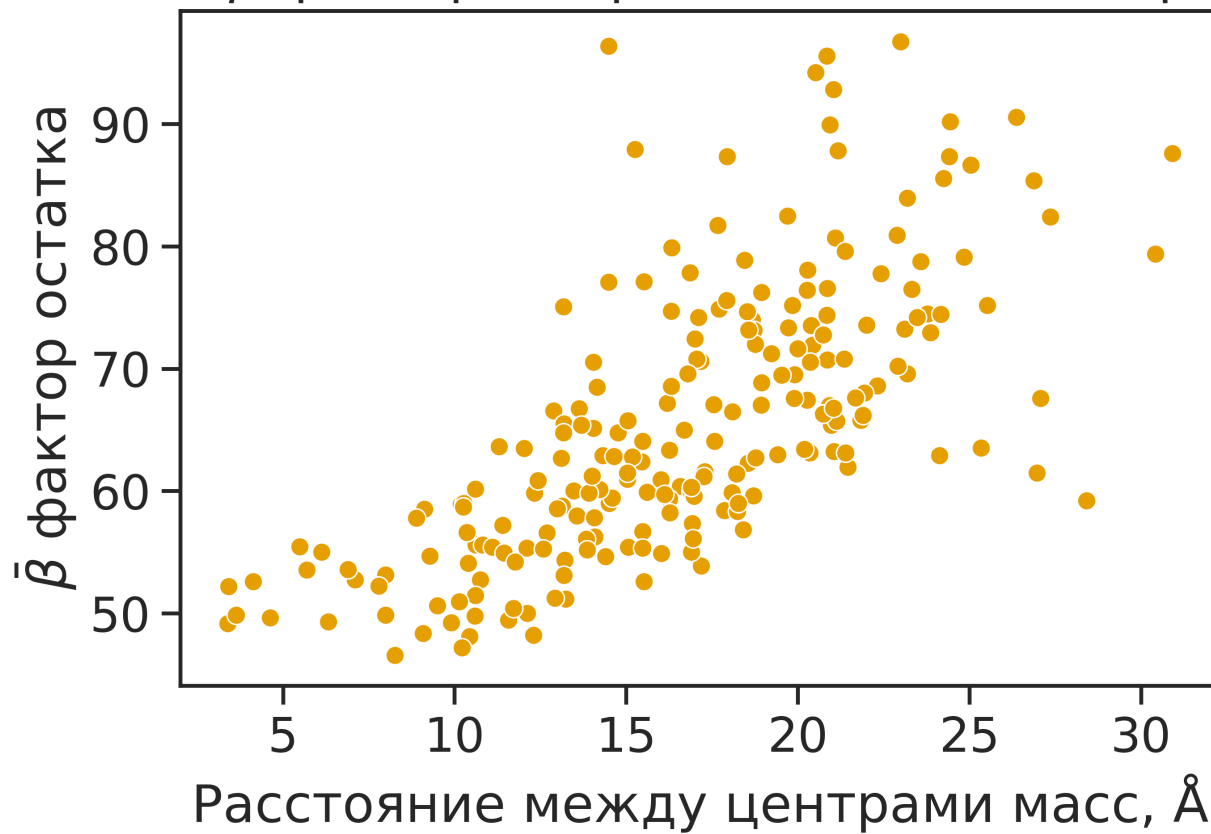


Глицин - простая неполярная аминокислота, находится в составе вторичной структуры (альфа-спирали) и подвижность ее ожидаемо очень сильно ограничена. (числами подписаны значения В-факторов атомов)

Задание 2. Prody и В-факторы, часть 2.

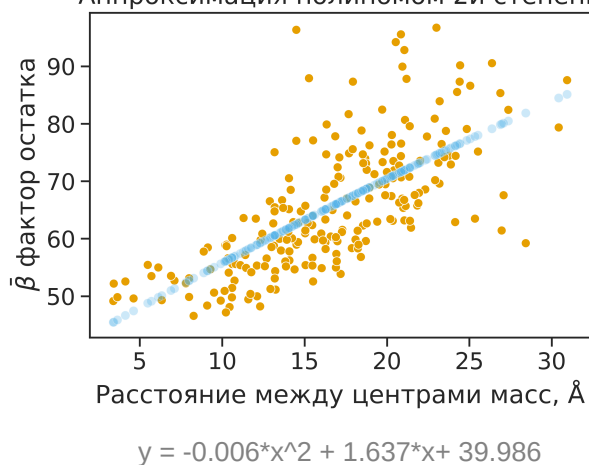
В этом задании предлагалось с помощью пакета prody сначала рассчитать положение центра масс белка, а потом посчитать расстояния от него до центра масс каждой аминокислоты.

Зависимость β фактора от расстояния до центра масс белка

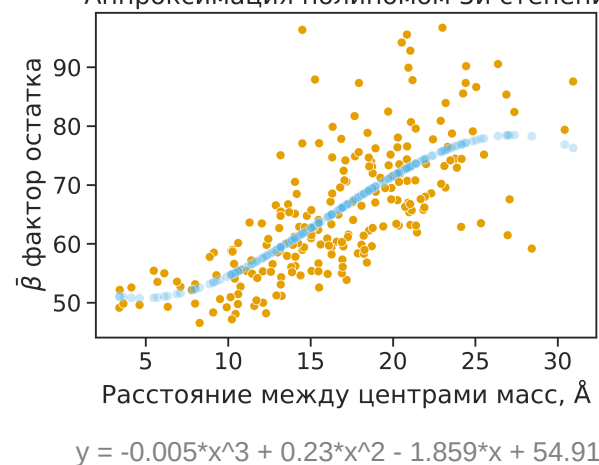


Видно, что чем больше расстояние, тем больше среднее значение B-фактора для аминокислоты. Чем ближе к центру масс структуры, тем остатки менее подвижны из-за большего количества стабилизирующих взаимодействий и в целом, как мы видели на примере с глицином, включённости в состав вторичной структуры. А, следовательно, они обладают меньшим B-фактором.

Зависимость β фактора от расстояния до центра масс белка
Аппроксимация полиномом 2й степени



Зависимость β фактора от расстояния до центра масс белка
Аппроксимация полиномом 3й степени



Аппроксимации экспонентой или логарифмом совсем не дали результатов.

	2rd order polynomial	3rd order polynomial
explained_variance_score	0.485717	0.502998
mean_squared_error	63.070481	60.951108
mean_absolute_error	6.070759	5.935210

Как мы видим, результаты довольно средние. С одной стороны, разброс точек велик, с другой стороны, полином третьей степени хорошо "ведёт себя" на начальном участке, но на участке 25-30 Å точек мало, и скорее всего он показывает неправильную зависимость: по логике, бета-фактор должен расти как бы экспоненциально и на этом участке взмывать вверх (экспоненциальным распределением аппроксимировать зависимость вычислительно не удалось).

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

В этом задании предлагается провести небольшую стимуляцию восстановления функции электронной плотности по экспериментальным данным.

Упрощения, которые вводятся в этом практикуме:

- Мы условно рассматриваем одномерную вселенную, в которой атомы находятся на одном отрезке длиной 30 Å.
- Электронная плотность атома описывается гауссовой кривой. За высоту колокола принимаем количество электронов этого атома. Максимум колокола находится в центре атома.
- В молекулах атомы располагаются на расстоянии 1-1.5 Å (ковалентная связь), молекулы на расстоянии 3-5 Å.

Задание функции

Сначала сгенерируем произвольную функцию электронной плотности при помощи скрипта `compile-func.py` (автор: А.В. Алексеевский). Функция имеет вид суммы нескольких гауссовых кривых и в нашем случае задана на интервале [0, 30] Å.

Отдельная кривая имеет вид:

```
gauss = lambda*x*exp(-(beta^2)*(x-gamma)^2)
```

Lambda задает высоту гауссиана, через которую мы имитируем число электронов у атома.

Beta задает ширину гауссиана, значение 3 соответствует релевантной нам ширине колокола плотности около 1 Å, что похоже на реальную ситуацию.

Gamma задает положение центра атома.

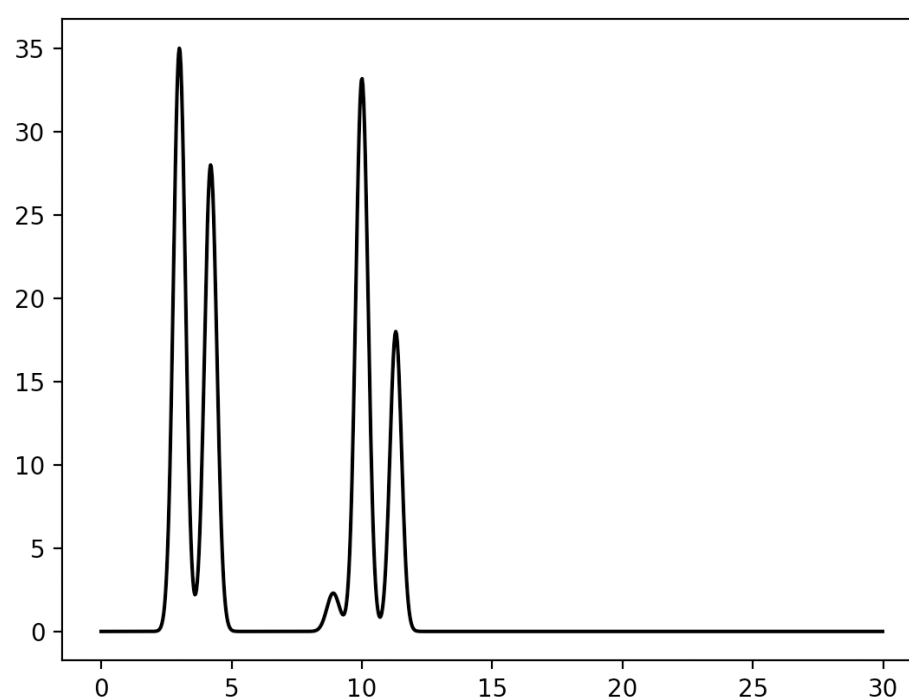
Зададим 2 молекулы, первая имеет **два** атома (расположены друг от друга на расстоянии 1.2 Å), вторая - **три** (друг от друга на расстоянии 1.1 и 1.3). Расстояние между молекулами - 4.7 Å.

Принцип генерации параметров для скрипта: первая цифра - более-менее произвольная, маленькое значение (2.3) имитирует атом водорода, большие значение - большие атомы.

Второе значение берется как 3 с небольшой дисперсией, что соответствует релевантной ширине колокола.

Третье значение соответствует заданным мной расстояниям, положим, что отсчет начинается с 3 Å.

```
python compile-func.py -g 35,3.1,3+28,3,4.2+2.3,2.9,8.9+33.2,3,10+18,3.2,11.3
```



Вид смоделированной функции электронной плотности.

Расчет амплитуд и фаз сигналов, моделирующих экспериментальные данные

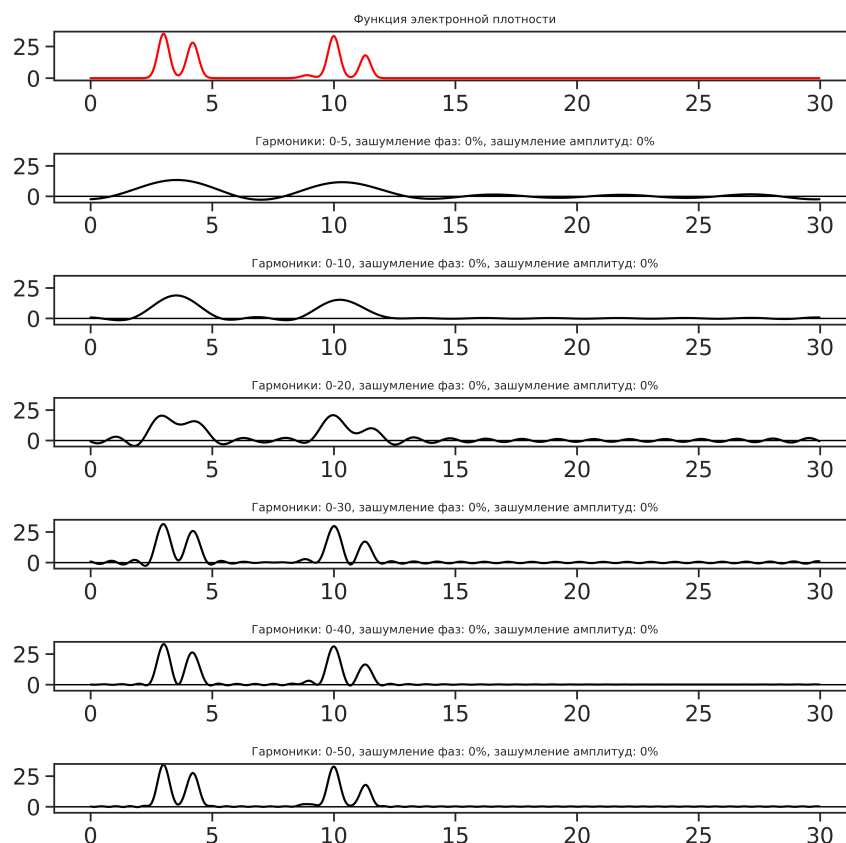
В реальном эксперименте мы получаем только амплитуды, но в этой стимуляции мы считаем, что знаем функцию электронной плотности и по ней можем однозначно определить и амплитуды, и фазы.

Несовершенности эксперимента:

- определяются амплитуды не для всех сигналов

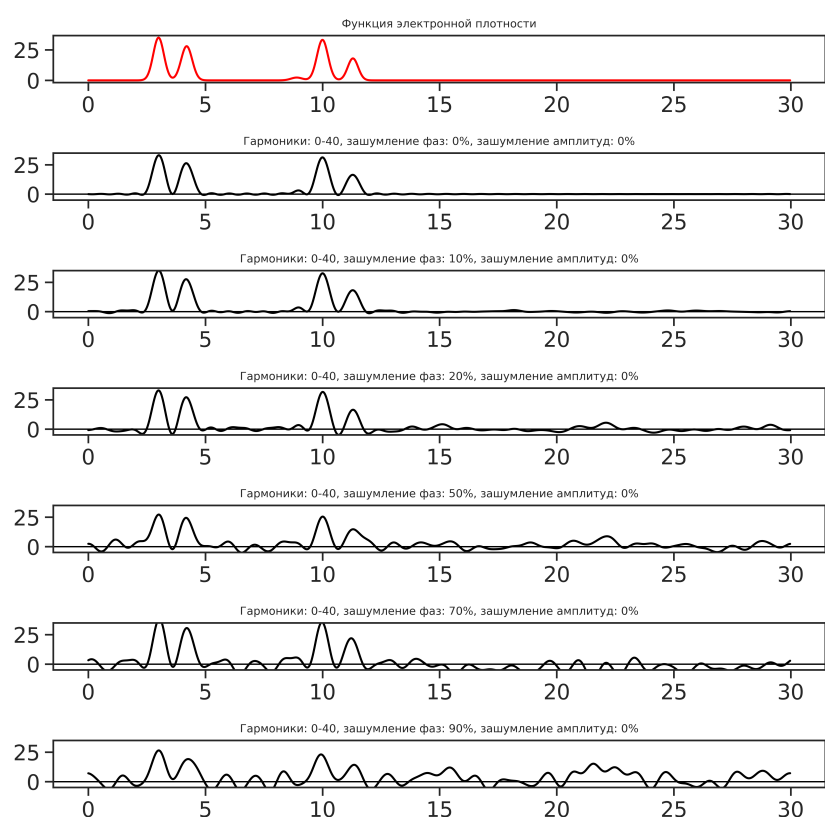
- ошибки в определении интенсивности сигналов
- ошибки в определении фаз

Эксперимент по восстановлению функции электронной плотности

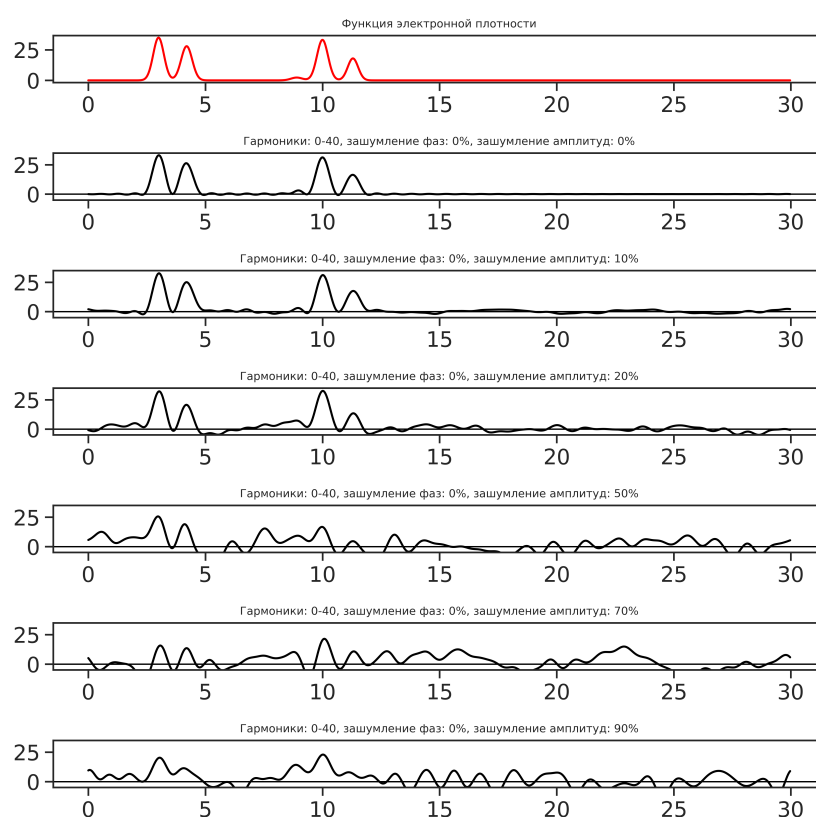


Восстановление функции по первым гармоникам, без зашумления фаз и амплитуд

Начальные гармоники, отвечающие низкому разрешению, дают плохое качество восстановления, однако, два "горба", отвечающие двум молекулам, видны.

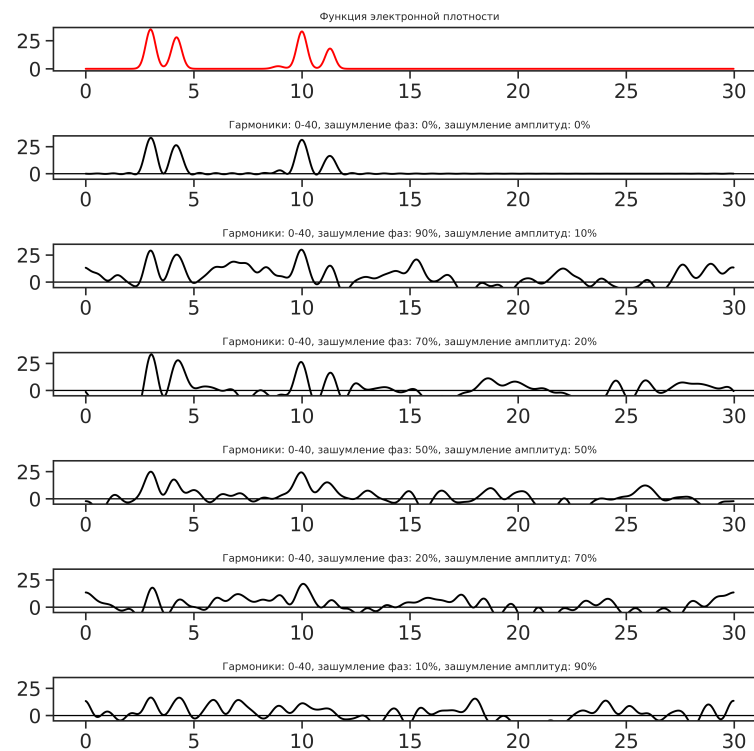


Восстановление функции по первым 40 гармоникам с зашумлением фаз, без зашумления амплитуд



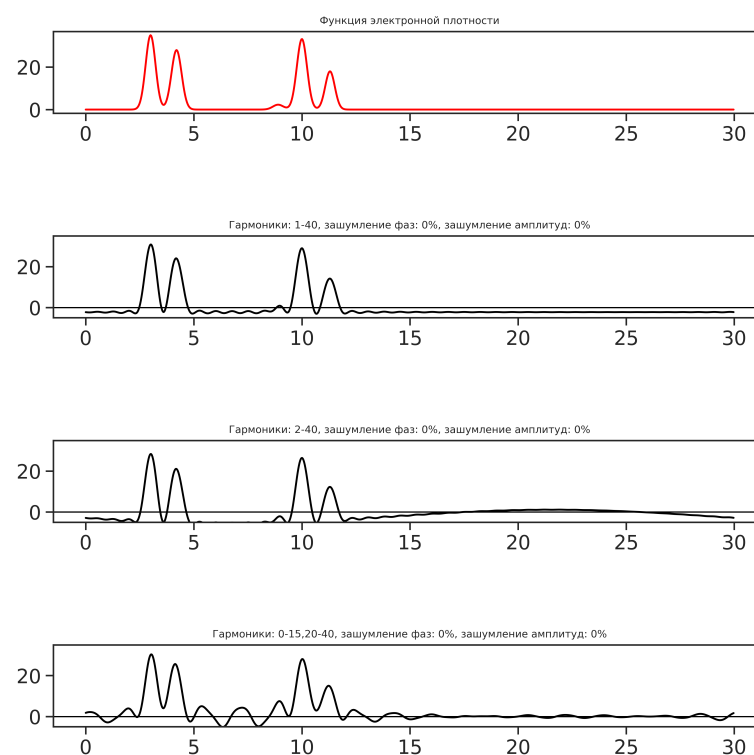
Восстановление функции по первым 40 гармоникам с зашумлением амплитуд, без зашумления фаз

Видно, что зашумление амплитуд сильнее снижает качество восстановления функции, чем зашумление фаз.



Восстановление функции по первым 40 гармоникам с зашумлением амплитуд и фаз

При высоком зашумлении фаз и низком зашумлении амплитуд функция ЭП слабо узнаваема: видны горбики, отвечающие нашим молекулам, но их можно вычлнить из шума только если иметь перед глазами исходную функцию электронной плотности. В противоположном случае, при высоком зашумлении амплитуд и низком зашумлении фаз функция ЭП становится совсем неузнаваемой.



Восстановление функции по неполному набору гармоник

Потеря первых нескольких гармоник не препятствует хорошему восстановлению функции. Потеря гармоник где-то посередине же приводит к потере информации.

Эксперимент	Набор гармоник	Разрешение	Шум амплитуды	Шум фазы	Качество восстановления
Первые гармоники	0-5	6.00	0	0	Плохое
Первые гармоники	0-10	3.00	0	0	Плохое
Первые гармоники	0-20	1.50	0	0	Среднее
Первые гармоники	0-30	1.00	0	0	Хорошее
Первые гармоники	0-40	0.75	0	0	Отличное
Первые гармоники	0-50	0.60	0	0	Отличное
Зашумление фаз	0-40	0.75	0	0	Отличное
Зашумление фаз	0-40	0.75	10	0	Отличное
Зашумление фаз	0-40	0.75	20	0	Хорошее
Зашумление фаз	0-40	0.75	50	0	Среднее
Зашумление фаз	0-40	0.75	70	0	Плохое
Зашумление фаз	0-40	0.75	90	0	Плохое
Зашумление амплитуд	0-40	0.75	0	0	Отличное
Зашумление амплитуд	0-40	0.75	0	10	Отличное
Зашумление амплитуд	0-40	0.75	0	20	Хорошее
Зашумление амплитуд	0-40	0.75	0	50	Плохое
Зашумление амплитуд	0-40	0.75	0	70	Плохое
Зашумление амплитуд	0-40	0.75	0	90	Плохое
Зашумление амплитуд и фаз	0-40	0.75	0	0	Отличное
Зашумление амплитуд и фаз	0-40	0.75	90	10	Среднее
Зашумление амплитуд и фаз	0-40	0.75	70	20	Среднее
Зашумление амплитуд и фаз	0-40	0.75	50	50	Плохое
Зашумление амплитуд и фаз	0-40	0.75	20	70	Плохое
Зашумление амплитуд и фаз	0-40	0.75	10	90	Плохое
Неполный набор	1-40	0.75	0	0	Отличное
Неполный набор	2-40	0.75	0	0	Хорошее
Неполный набор	0-15,17-40	0.75	0	0	Среднее

Таблица с результатами