



Рис. 1:

Предположим, что молекула ДНК имеет трехмерную организацию спирали с постоянным расстоянием между витками (смотри Рис.1). Пусть в данной молекуле ДНК содержится 500 миллионов пар нуклеотидов (500 megabases, 500Mb) и имеет длину порядка 170 мм, а расстояние между соседними витками составляет 2 мм.

#### Задача

1. Параметризовать эту кривую, то есть записать ее уравнение в виде  $(x(t), y(t))$  (или написать скрипт, который рисует такую кривую).
2. Разбить хромосому на 1000 равных бинов по 500 Kb, и построить матрицу расстояний между центрами бинов.
3. Пользуясь данными по первой хромосоме человека (250 Mb, 85 мм) из <ftp://cooler.csail.mit.edu/coolers/hg19/Rao2014-IMR90-MboI-allreps-filtered.500kb.cool> сгенерируйте матрицу Hi-C для данной хромосомы. Предлагается считать, что среднее покрытие такое же, как в предлагаемых данных; для простоты можно считать, что все предлагаемые данные линейные (то есть, что в тренировочных данных два бина на расстоянии  $N$  бинов друг от друга вдоль хромосомы имеют расстояние в 3D пропорциональное  $N$ ).
4. (бонус, если вдруг будете успевать) попробовать восстановить 3D структуру хромосомы из смоделированной в предыдущем пункте Hi-C матрицы с помощью тула 3DMax <https://github.com/BDM-Lab/3DMax>.