Learning Note of OpenMM 7.0

1 previous work using OpenMM dynamics simulation toolkit

Rajpurkar, A.R. et al., *Nature Communication* , 2021

https://doi.org/10.1038/s41467-021-23831-4


ARTICLE

https://doi.org/10.1038/s41467-021-23831-4 OPEN

Check for updates

Deep learning connects DNA traces to transcription to reveal predictive features beyond enhancer-promoter contact

Aparna R. Rajpurkar[1,2], Leslie J. Mateo[2], Sedona E. Murphy[1,2] & Alistair N. Boettiger[2]✉

首先，就是之前调研的这篇深度学习的工作。（以距离矩阵为输入，转录开关为学习对象。）其中，使用了OpenMM来模拟一些简单的带enhancer & promoter的polymer（较为简单的转录开关机制），来判断一些判断方法的实用性。模拟中，使用了下文中对于OpenMM的改进，即加入了force calculation functions，使得本文模拟中，52个结合位点间没有吸引，只是靠着density提供吸引力（用来模拟核约束），而在位点过于接近时候有一个斥力。

Nuebler, J. et al.,*Proceedings of the National Academy of Sciences* , 2018

https://www.pnas.org/doi/suppl/10.1073/pnas.1717730115

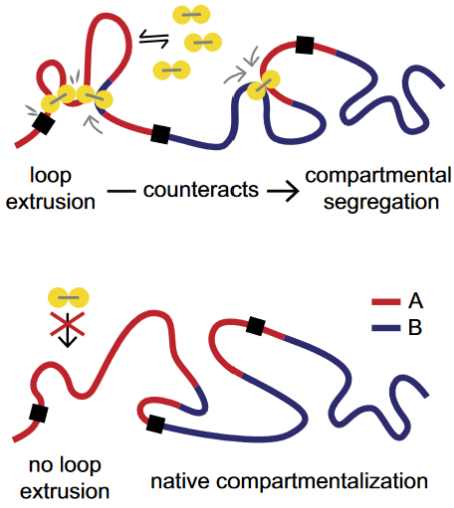

Chromatin organization by an interplay of loop extrusion and compartmental segregation

Johannes Nuebler[a], Geoffrey Fudenberg[b], Maxim Imakaev[a], Nezar Abdennur[a], and Leonid A. Mirny[a,1]

[a]Department of Physics, Institute for Medical Engineering and Science, Massachusetts Institute of Technology, Cambridge, MA 02139; and [b]Gladstone Institutes of Data Science and Biotechnology, San Francisco, CA 94158

Edited by Robert H. Singer, Albert Einstein College of Medicine, Bronx, NY, and approved June 4, 2018 (received for review October 10, 2017)

这篇是上一篇的基础，也使用这一系统。本文主要研究哺乳动物染色质上TADs，并使用分子链模拟这一结构的形成，以及一些生物分子浓度改变时，TADs会发生何种变化。



Supplemental Information中详细描述了模拟过程。聚合物表示为具有一系列单体，用谐波势能连接，有体积排斥势，并且在两个typeB的单体间的微小吸引力

2 basic functions of OpenMM 7.0

- 读取输入文件（Amber, CHARMM, Gromacs, and Desmond）
- 编辑分子模型
- 定义作用在分子系统上的力（明确指定，或从文件加载力场）
- 计算力和能量
- 运行模拟
- 输出结果

其中，我们最关心的肯定是自定义力场部分，这里OpenMM提供的有：

Table 1. Custom forces supported by OpenMM 7.0.

Custom Force Class	Description
CustomBondForce	Applies forces to pairs of bonded atoms based on the distance between them.
CustomAngleForce	Applies forces to triplets of bonded atoms based on the angle between them.
CustomTorsionForce	Applies forces to sets of four bonded atoms based on the dihedral between them.
CustomExternalForce	Applies forces to individual atoms based on their positions.
CustomCompoundBondForce	Applies forces to sets of arbitrarily many bonded atoms based on any combination of their positions, distances, angles, and dihedrals.
CustomNonbondedForce	Applies forces to pairs of non-bonded atoms based on the distance between them.
CustomGBForce	Supports multi-stage computations of non-bonded interactions, such as generalized Born implicit solvent models.
CustomCentroidBondForce	Similar to CustomCompoundBondForce, but the interaction is based on the centroids of groups of atoms rather than individual atoms.
CustomManyParticleForce	Supports non-bonded interactions that depend on the positions of arbitrarily many atoms at once.
CustomHbondForce	Supports a variety of hydrogen bonding models.

3 simulation test

首先测试教程自带的模拟程序。通过这个程序我们可以对模拟环境进行调整。

首先是开始import的功能包：

```
from openmm.app import *
from openmm import *
from openmm.unit import *
from sys import stdout
```

运行需要`cd`到对应母文件夹，使用`python simulatePdb.py`命令启动。

3.1 A pdb example

之后导入已经写好分子信息，这里使用的是pdb文件，实际上还有其他的种类，如(Amber和Gromacs)，包含了分子拓扑结构和原子位置

```
pdb = PDBFile('input.pdb')
```

之后就是引入力场，力场信息使用.xml文件存储，这里引入两个

```
forcefield = ForceField('amber14-all.xml', 'amber14/tip3pfb.xml')
```

然后就是设置模拟系统，主要是加入分子信息和力场，其中有三个关键参数：远距离静电作用为particle mesh Ewald，直接空间作用（应该是碰撞体积）为1nm，约束键长度为氢分子的键长。

```
system = forcefield.createSystem(pdb.topology, nonbondedMethod=PME,
        nonbondedCutoff=1*nanometer, constraints=HBonds)
```

下一步调整的是积分器，这里的示例是郎之万中值积分器（不同的模型应当选取不同的积分器），同样三个参数：温度，摩擦因数和**步长**：

```
integrator = LangevinMiddleIntegrator(300*kelvin, 1/picosecond,
0.004*picoseconds)
```

之后就是启动模拟了，需要三个部分：分子信息，系统信息和积分器，之后就是设置初始位置，然后设置形成一个附近能量极小值，一般来说开始模拟的时候要加上这一步：

```
simulation = Simulation(pdb.topology, system, integrator)
simulation.context.setPositions(pdb.positions)
simulation.minimizeEnergy()
```

最后就是模拟的输出了，首先会输出一个叫output的文件，每1000步写入一次信息。同时，而有一个在cmd（指定stdout作为输出文件即可）中print的内容（=True的内容会被打印），最后指定模拟步数。

```
simulation.reporters.append(PDBReporter('output.pdb', 1000))
simulation.reporters.append(StateDataReporter(stdout,
1000,step=True,potentialEnergy=True, temperature=True))
simulation.step(10000)
```

最终效果为:

```
(base) PS C:\Users\zhouquan\anaconda3\Library\share\openmm\examples> python simulatePdb.py
#"Step","Potential Energy (kJ/mole)","Temperature (K)"
1000,-142593.43593246874,291.04135922789004
2000,-140716.68593246874,300.08191406953165
3000,-140684.49843246874,298.6672816483564
4000,-140737.81093246874,298.7236562131145
5000,-140317.40468246874,300.57526043166905
6000,-140552.77968246874,301.15573974592377
7000,-140760.37343246874,298.84600272638164
8000,-140126.31093246874,302.27987843215664
9000,-140887.99843246874,293.40392116105085
10000,-141277.06093246874,296.5417573081118
(base) PS C:\Users\zhouquan\anaconda3\Library\share\openmm\examples>
```

3.2 Using AMBER Files

Amber系统需要导入两个文件，prmtop文件包括了分子拓扑形式和力场参数，inpcrd是原子位置:

```
prmtop = AmberPrmtopFile('input.prmtop')
inpcrd = AmberInpcrdFile('input.inpcrd')
```

同时，在设置部分，还要额外从inpcrd设置Box Vector

```
if inpcrd.boxVectors is not None:
    simulation.context.setPeriodicBoxVectors(*inpcrd.boxVectors)
```

3.3 Using Gromacs Files

与上部分基本相同，只是replacing `AmberInpcrdFile` and `AmberPrmtopFile` with `GromacsGroFile` and `GromacsTopFile`

3.4 Using CHARMM Files

同样，导入信息也有一定区别。

3.5 The OpenMM-Setup Application

实际上，这一运行脚本可以自动生成，通过openMM setup。安装和启动可以使用以下命令:

```
conda install -c conda-forge openmm-setup
openmm-setup
```

会生成一个网页，从而根据输入文件生成脚本

3.6 模拟参数 Simulation Parameters

3.6.1 平台

平台使用的应该是OpenCL, CUDA, CPU, or Reference

3.6.2 力场

这里只大部分请参考文档，只记录一小部分：

amber14提供了一系列标准力场

File	Parameters
amber14/protein.ff14SB.xml	Protein (recommended)
amber14/protein.ff15ipq.xml	Protein (alternative)
amber14/DNA.OL15.xml	DNA (recommended)
amber14/DNA.bsc1.xml	DNA (alternative)
amber14/RNA.OL3.xml	RNA
amber14/lipid17.xml	Lipid
amber14/GLYCAM_06j-1.xml	Carbohydrates and glycosylated proteins[2]
amber14/tip3p.xml	TIP3P water model[3] and ions
amber14/tip3pfb.xml	TIP3P-FB water model[4] and ions
amber14/tip4pew.xml	TIP4P-Ew water model[5] and ions
amber14/tip4pfb.xml	TIP4P-FB water model[4] and ions
amber14/spce.xml	SPC/E water model[6] and ions

James A. Maier, Carmenza Martinez, Koushik Kasavajhala, Lauren Wickstrom, Kevin E. Hauser, and Carlos Simmerling. Ff14sb: improving the accuracy of protein side chain and backbone parameters from ff99sb. *Journal of Chemical Theory and Computation*, 11(8):3696–3713, 2015.

这里大部分力场文件收录在工具包中，可以使用导入

```
forcefield = ForceField('amber14-all.xml', 'amber14/tip3pfb.xml')
```

3.6.3 Constraints

有四个值可以使用

Value	Meaning
`None`	No constraints are applied. This is the default value.
`HBonds`	The lengths of all bonds that involve a hydrogen atom are constrained.
`AllBonds`	The lengths of all bonds are constrained.
`HAngles`	The lengths of all bonds are constrained. In addition, all angles of the form H-X-H or H-O-X (where X is an arbitrary atom) are constrained.

3.6.4 积分器 Integrators

其中值得注意的是，有一个布朗积分器，可以用于模拟扩散动力学系统

4 Polymer Dynamics

4.1 之前引用文献的模拟模型

Nuebler, J. et al.,*Proceedings of the National Academy of Sciences* , 2018

https://www.pnas.org/doi/suppl/10.1073/pnas.1717730115

Chromatin organization by an interplay of loop extrusion and compartmental segregation

Johannes Nuebler[a], Geoffrey Fudenberg[b], Maxim Imakaev[a], Nezar Abdennur[a], and Leonid A. Mirny[a,1]

[a]Department of Physics, Institute for Medical Engineering and Science, Massachusetts Institute of Technology, Cambridge, MA 02139; and [b]Gladstone Institutes of Data Science and Biotechnology, San Francisco, CA 94158

Edited by Robert H. Singer, Albert Einstein College of Medicine, Bronx, NY, and approved June 4, 2018 (received for review October 10, 2017)

首先我们来看本文的模型，这里取DNA是一条有谐波势连接的单体链，这里主要分析一下他的具体模拟包——polychrom包。

（以下超链接可以跳转到网页详细说明）在这个包中，力场在 `polychrom.forces` 和 `polychrom.forcekits` 中定义，接触关系由 `polychrom.contactmaps` 文件定义，加载和存储则在 `polychrom.polymerutils` 中，最终由 `polychrom.simulation` 统合，分析则有 `polychrom.polymer_analyses` 脚本负责。

4.2 amber14中的DNA模型

Maier JA et al.,*Journal of chemical theory and computation* ,2015

这一分子模型为原子尺度，我们难以使用。

4.3 polychrom

4.3.1 安装

环境安装:

- [x] conda安装
- [ ] CUDA安装

前置包安装:

- [x] six
- [x] cython
- [x] numpy>=1.9
- [x] scipy>=0.16
- [x] h5py>=2.5
- [x] pandas>=0.19
- [x] joblib
- [x] pyknotid
- [x] pytest

4.3.2 测试

详见滴答清单上操作流程

4.3.3 正式模拟

E-P1系列:

polymer参数设置

```
N=400

platform="CUDA",
```

```
integrator="variableLangevin",
error_tol=0.003,#<0.01
GPU="0",#0 represent for the first GPU, and 1 for the second. If only 1 GPU is available, it will be selected automatically
collision_rate=0.03,#0.01-0.05
N=N,
save_decimals=2,
PBCbox=False,
reporters=[reporter],
temperature=293,#devault value is 300K
mass=1518500,#estimated by dsDNA 607.4 per bp and 2.5Kb per monomer

polymer = starting_conformations.grow_cubic(400, 1000)#(m, n)represents a m-long polymer in a cube of n*n*n
```

Force设置

首先，限制区域的力为：spherical_confinement，Constrain particles to be within a sphere.

在Harmonic Bonds中，两个参数分别为键长(bL bondLength)和键长波动(bWD bondWiggleDistance)（自由能导致，Bond energy equal to 1kT at bondWiggleDistance）

在angle force中，k代表了聚合物的僵硬程度——越小越软，越大越僵硬，推荐值为1.5。（i-th triplet Potential is k * alpha^2 * 0.5 * kT）

最后，排斥力的参数代表了r=0处的力场势能(It has the value of trunc at zero, stays flat until 0.6-0.7 and then drops to zero together with its first derivative at r=1.0.)

当然，根据我们的现象，需要一个吸引力场，形式可以尝试（可能是本次重点）

最终数据的参数如下表，每次进行100次模拟，模拟长度为5000步长

Number	步长	键力bL/bWD	角度力k	排斥力trunc	吸引力场
1	auto	1.0/0.05	1.5	3.0	No
2		0.5/0.025			
3					

步长实际上也是我们讨论的重点，理论上variableLangevin会自定义步长

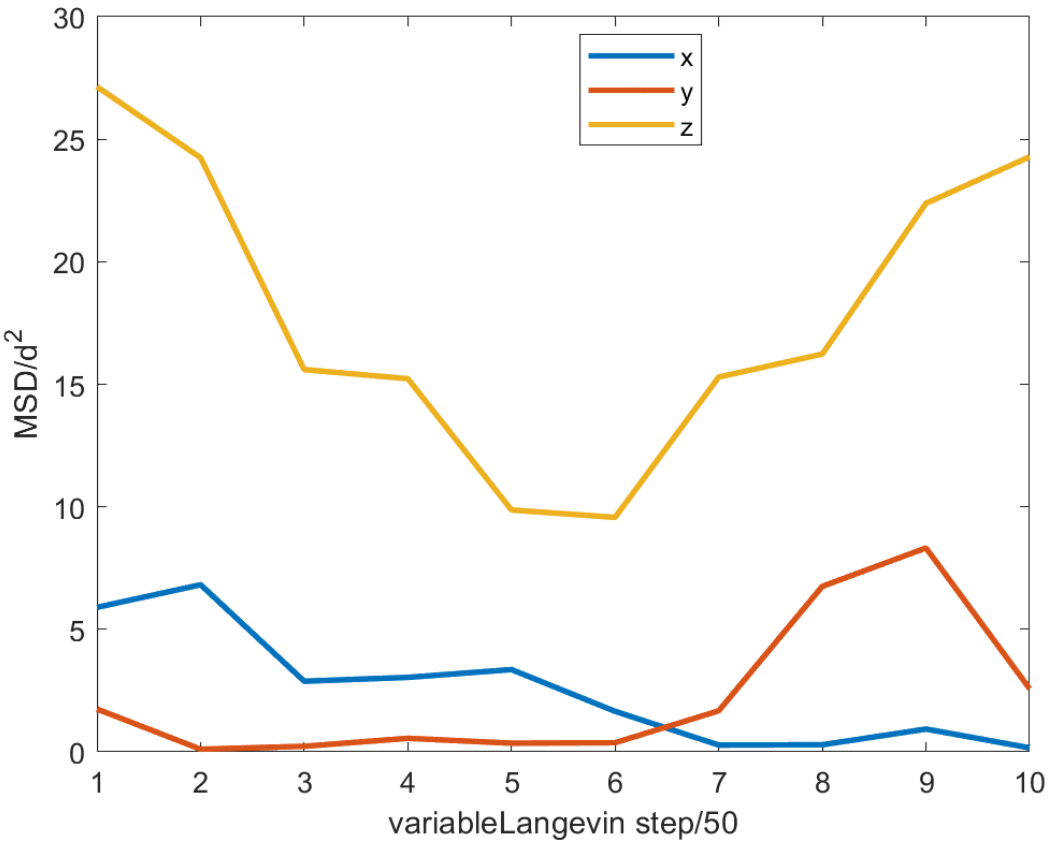
4.3.4 数据处理：多次实验

但我们发现了一些问题：程序自带的reporter并不能记录所有时刻的monomer位置，仅能汇报最后时刻。所以，目前有两种改进方法，本部分记录第一种。

特别地：**这几次模拟使用的monomers间距为0.5单位**

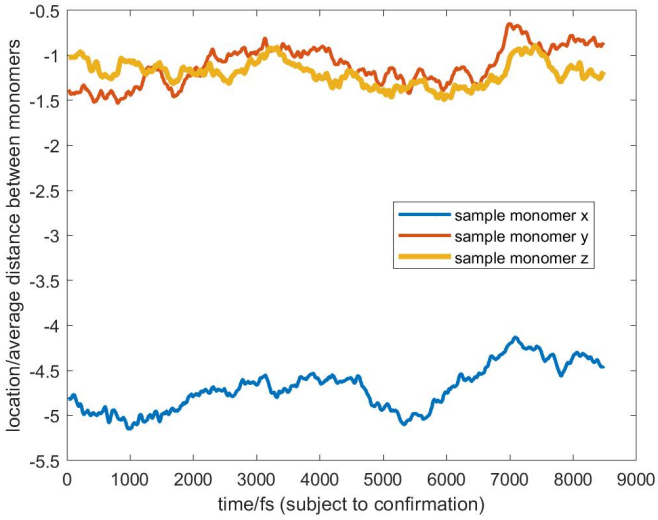
进行10次持续时间不同的模拟，每一次模拟50个分子链，得到不同时间的数据。按照统计规律来说，此数据得到的MSD随时间分布，应当与实际记录分子链各个时刻数据统计上一致。

进行数据处理后，实验效果不佳。MSD随着时间分布如下：

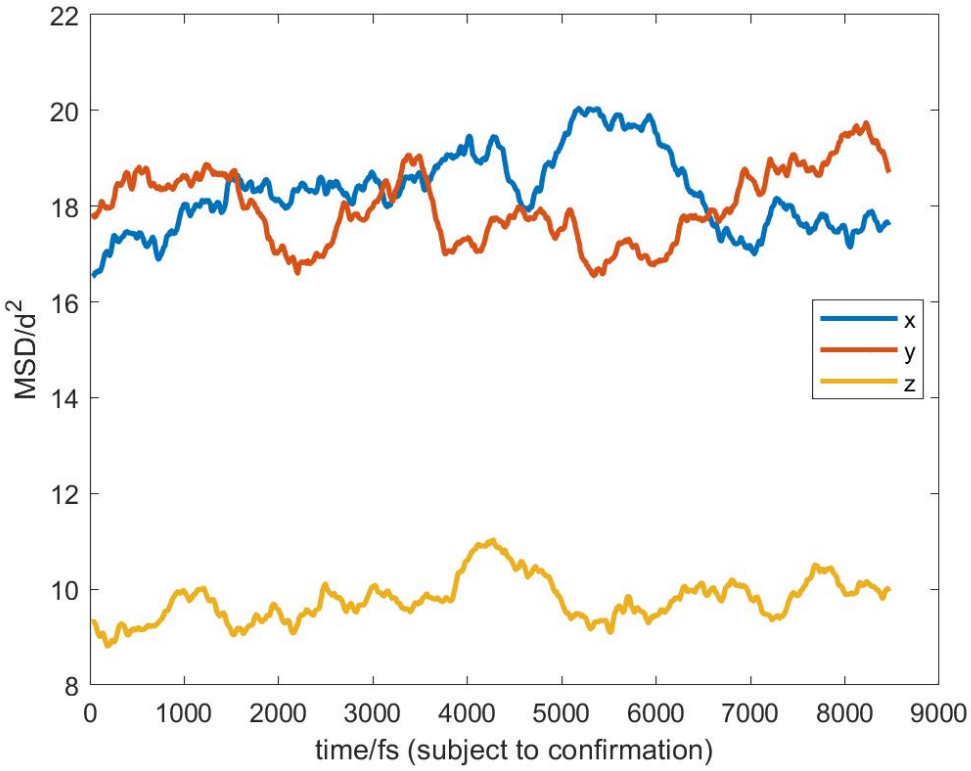


4.3.5 数据处理：单次模拟

而正常的单次模拟给出的结果如下。我们可以看到，如果不重新建立一个simulation，仅仅`do_block`其不会进行初始化，以下为验证：



最终给出的MSD如下图，为500长度分子链中第250个和350个之间相对位移的MSD



其中d为两个monomers之间的平均距离。目前有几个问题：

- 时间尺度太小（待使用郎之万积分法中给出的确定）

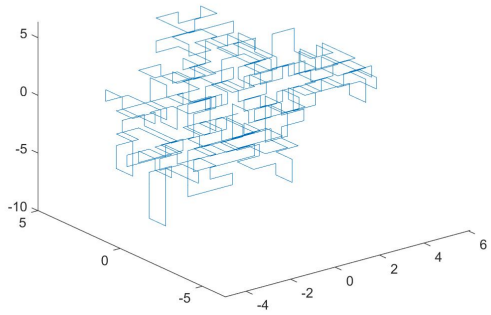
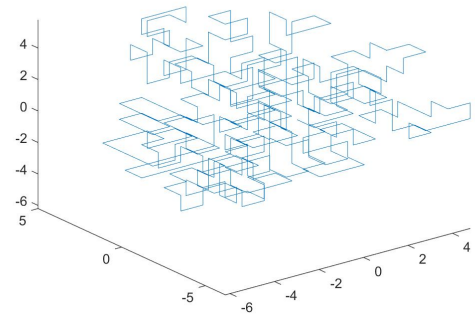
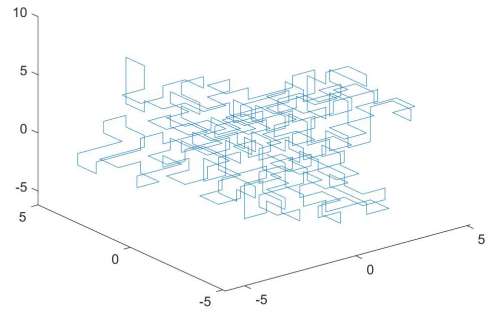
- 为何z方向上差别如此大?
- 100个monomers之间为何距离如此接近？可能和外加的总体限制立场有关（有待调整）
- 波动太大（这组数据仅为10次实验）

使用VMD三维作图

4.3.6 确定初始状态

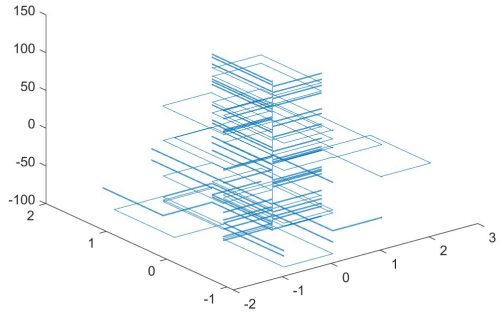
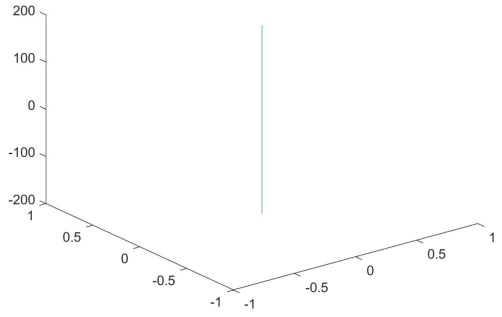
4.3.6.1 cubic standard mode

grows a ring starting with a 4-monomer ring in the middle, on a cubic lattice in the cubic box of size boxSize.

cubic size	figure
500	
1000	
2000	

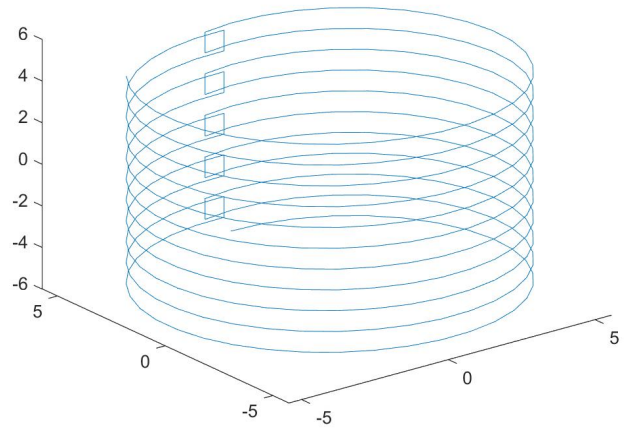
4.3.6.2 cubic linear mode

it grows a linearly organized chain from 0 to size.

box size	
200	
400	

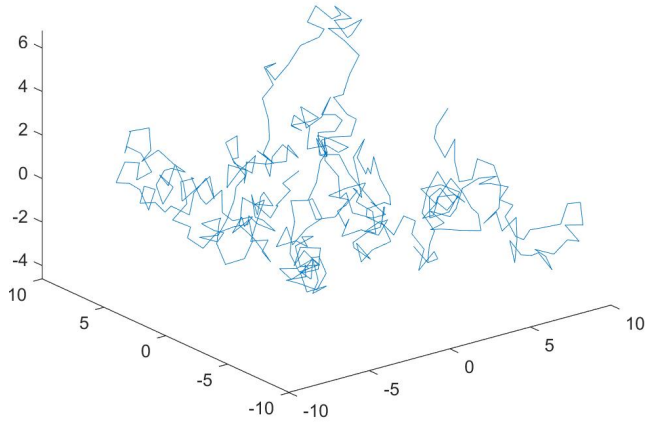
4.3.6.3 spiral

Creates a "propagating spiral", often used as an easy mitotic-like



4.3.6.4 random walk

Creates a freely joined chain of length N with step step_size



P.S. constrained_random_walk ?