

pygamd_v_me_50_meal: 一款用于 pygamd 分子动力学模拟软件的脚本。

由于 PYGAMD 未提供适用于生物大分子凝聚体模拟的数据处理软件，因此我整合了我所使用过的脚本，集成一个脚本，并提供了丰富的选项以供选择。

本软件分为两部分：分子动力学模拟前的文件生成、分子动力学模拟后的文件处理。在拿到 PDB 文件后，可以使用本软件进行模拟体系的构建、运行脚本的生成、分子动力学模拟的运行。再得到结果后，可以继续使用本软件进行模拟所得到的 XML 文件的处理。

安装

可直接使用 pip 安装：

```
pip install git+https://gitee.com/lyxlyxlyxxx/pygamd_v_me_50_meal.git
```

或者

```
git clone https://gitee.com/lyxlyxlyxxx/pygamd_v_me_50_meal.git
cd pygamd_v_me_50_meal
pip install -e .
```

分子动力学模拟前的文件生成

pygamd_v_me_50_meal/simulate_creation/xml_generator.py

XMLGenerator 类用于从 PDB 文件生成模拟所需要的粗粒化 XML 文件，目前支持蛋白质与 DNA 粗粒化模型的生成。

使用

```
v50 -pdb2xml /path/to/pdb_file.pdb
v50 -p /path/to/system -pdb2xml filename.pdb
```

例如，要将 /home/protein 文件夹下的 **1kx5.pdb** 转化为 **xml** 文件，输入

```
v50 -pdb2xml /home/protein/1kx5.pdb
或 v50 -p /home/protein -pdb2xml 1kx5.pdb
```

即可生成 **1kx5.xml** 文件。

下列功能直接在运行过程中以交互式选项呈现

设置弹性网络、选择 DNA 粗粒化模型、生成模拟脚本、设置盒子大小（盒子为正方体盒子，大小默认为 100nm，如需要设置其他长度，请使用-box_size选项）

蛋白质粗粒化模型：一个氨基酸粗粒化为一个粒子：Dignon G L, Zheng W, Kim Y C, Best R B, Mittal J. PLOS Computational Biology, 2018, 14(1): e1005941

DNA 粗粒化模型：目前支持两种模型

1. 3SPN 模型

Knotts T A, Rathore N, Schwartz D C, De Pablo J J. A coarse grain model for DNA[J]. *The Journal of Chemical Physics*, **2007**, 126(8): 084901.

2. Mittal 2 Bead 模型：

Kapoor U, Kim Y C, Mittal J. Coarse-Grained Models to Study Protein–DNA Interactions and Liquid–Liquid Phase Separation[J]. *Journal of Chemical Theory and Computation*, **2023**: acs.jctc.3c00525.

注意：目前“Mittal 2 Bead 模型”还未能成功在 PYGAMD 中实现。

分子动力学模拟后的文件处理

GetSequence

GetSequence 类用于提取 xml 或 pdb 文件的序列，根据文件后缀自动检测文件类型。

使用

你可以直接使用 -get_seq 选项：

```
v50 -get_seq /path/to/file.xml
v50 -get_seq /path/to/file.pdb
```

也可以通过 -p 提供路径，然后使用 -get_seq 提供文件名：

```
v50 -p /path/to/system -get_seq filename.xml
v50 -p /path/to/system -get_seq filename.pdb
```

CoordinatesProcessor

CoordinatesProcessor类用于进行坐标提取，本脚本所有分析均基于提取后的坐标文件。

使用

```
v50 -p /path/to/system -xyz t
```

注意:

1. 在提取坐标之前, 请确保提供路径的最后一级为"数字+分子名称"或有多个连接的格式, 并且按照 xml 文件中的顺序排好。
如体系中含有 40 个 A 分子、20 个 B 分子、30 个 C 分子, 可将体系命名为 "40A20B30C" 或 "40A+20B+30C"。
当需要 nohup 运行时, 可以修改 Data 类中的 length_dict 字典, 或将路径改为 40A-100+20B-635+30C-395。
2. 请将分子动力学模拟得到的 XML 文件放置在提供的目录下或者在该目录下的 xml 文件夹内。
3. 在进行数据处理之前请进行坐标提取, 我们的程序依赖于坐标文件, 请确保提供的路径下有相应的坐标文件。

可选参数

-remove_enm : 是否移除弹性键 (如果加入了弹性网络)

-remove_condensate_pbc : 是否移除凝聚体的 PBC。若设置为 True, 则会根据凝聚体去除 PBC。

TODO: 还是有 bug, 待修复。

ContactMapCalculator

ContactMapCalculator 类用于从提取的坐标中计算各个分子之间以及某个分子自身的 Contact Map。

注意, 在计算自身的 Contact Map 时, 已经排除了自己与自己的相互作用。

使用

```
v50 -p /path/to/system -cm t
```

在默认的交互界面, 我们提供了以下选项:

是否已经完成坐标提取、选择需要计算的分子对、选择轨迹切片, 可以选择平衡后的体系进行计算

由于脚本运行较慢, 我们提供了 -cm_choice 选项, 选项后跟着两个由 / 分割的两部分, 第一个部分为要计算的分子对, 第二个部分为要计算的轨迹范围。

使用方法:

```
v50 -p /path/to/system -cm t -cm_choice 1-2,2-2/1000,2000
```

指要计算的分子对为[1, 2]和[2, 2], 选取的轨迹为第 1000-2000 帧

此时无需交互即可完成计算, 因此可以使用nohup:

```
nohup v50 -xyz f -cm t -cm_choice 1-2,2-2/1000,2000 &
```

我们还提供了以下选项:

1. -r_cut : 设置计算 Contact Map 的 r_{cut} , 默认为 $1.12 (2^{1/6} \text{ \AA})$ 。

```
v50 -p /path/to/system -cm t -r_cut 4.0
```

2. -draw : 用于无需计算的情况下绘图。

```
v50 -p /path/to/system -cm t -draw t
```

RgRMSDRMSFCalculator

RgRMSDRMSFCalculator类用于计算 Rg、RMSD、RMSF。

使用

```
v50 -rg t  
v50 -rmsd t -ref reference_file.xml  
v50 -rmsf t -ref reference_file.xml
```

在运行过程中会询问

1. 需要计算的分子（ref 提供的那个就计算哪个）
2. 若只计算结构域，请输入计算结构域的氨基酸残基范围
3. 若要截取平衡结构，请输入截取的轨迹的范围

MassDensityDistributionCalculator

MassDensityDistributionCalculator类用于计算体系的质量数密度分布（其实是数密度分布）

注意：在使用前请先进行对凝聚体PBC的去除。

使用

```
v50 -mass_density t
```

在运行过程中会询问

1. 选择需要计算的分子对
2. 选择轨迹切片，可以选择平衡后的体系进行计算