README

Molecular Dynamics datasets

The following directories contain data produced from Molecular Dynamics simulations performed using the LAMMPS (Ver 3Mar2020) software:

```
equilibrium_trajectories.zip
kinetics_1_trajectories.zip
kinetics_2_trajectories.zip
```

- `equilibrium`: contains datasets related to trajectories done at T_f, used for equilibrium results
- `kinetics_x`: contains datasets related to trajectories that starts from an unfolded configuration and are simulated at $T<T_f$. Divided into two sets.

Metadata and input files examples

`datasets_list_description.csv` describe briefly each "dataset directory" (e.g. `6Nov22`) that collects a certain number of trajectories. Specifically, it describe the temperature used. In depth details about how each trajectory it is setup can be found the article, in the example input files and in the log relative to each "dataset directory".

`example_input_files.zip`: contains an example of LAMMPS input files. All files are generate similarly to these, where only starting position and temperature are changed.

```
example_input_files
    1srl_0.lam
    1srl_input.data
    1srl_input.pair_coeff
```

where the `.lam` is the input file (in which temperature, integrator etc. are specified), `.data` describes the protein topology (bound interactions, backbone, angles etc.) and `.pair_coeff` is a programmatically generated file that describe the non-bounded interaction between residues. The latter is obtained from the native contact map of each protein.

Each dataset (e.g. `31Oct22`) has the following structure:

```
31Oct22
    Data_1srl_0
    Data_1srl_1
    Data_1srl_10
    Data_1srl_11
    Data_1srl_12
```

```
...
    logs
```

- `Data_1srl_0` contains the actual trajectory with the following files:

  ```
      1srl_1000000000.restart
      1srl.dcd
      1srl_freq_1.restart
      1srl_freq_2.restart
      1srl_init.data
  ```

  where the `.dcd` is the trajectory file, `.data` is the LAMMPS topology file (can be converted to other file using the Python library `MDAnalysis` or using a VMD plugin) and the `.restart` are LAMMPS restart files
- `logs`: contains LAMMPS log file for each trajectory. Each log contains information about simulation details, simulation length and etc.

Analysis files

`entangled_folding.zip` is a directory, structured as a minimal python library (even though it is NOT, see the readme inside it for further details), that collects scripts and Jupyter notebooks (in Python) used to analyse Molecular Dynamics trajectories. I contains plotting function too, together with user-implemented functions and libraries used for the analysis. Its structure is the following:

```
entangled_folding
    analysis
        classificationTraj
        contact_fit.py
        correlation_contact_maps.py
        foldingTime
        __init__.py
        intermediate
        pathways
        phaseSpace
        traj
        utils
        visualization
    README.md
    setup.cfg
    setup.py
```

Each section has a readme file to present the content of each subsection. Moreover, each script/notebook is highly commented.