Pytraj: Interactive data analysis for molecular dynamics simulations

Hai Nguyen,[1,3] Daniel R. Roe,[2] Jason Swails,[1] David A. Case[1]

[1] BioMaPS Institute, Rutgers University

[2] Department of Medicinal Chemistry, The University of Utah

[3] Email: hainm.comp@gmail.com

Introduction

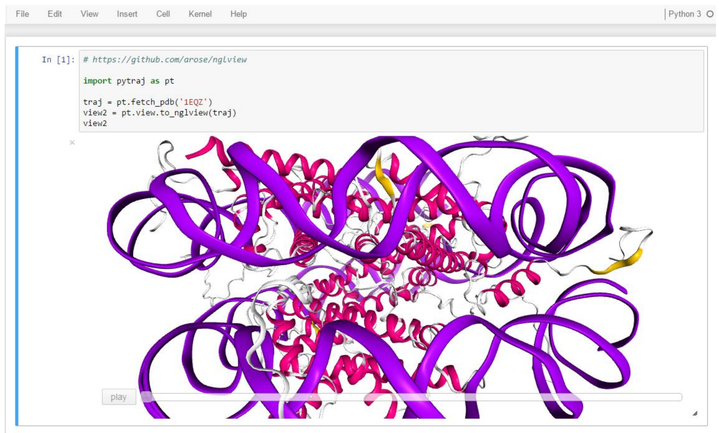
We describe here the overall design of pytraj, a Python front-end for CPPTRAJ, which is a popular data analysis program in Amber. Its source code is freely available via GitHub:https://github.com/Amber-MD/pytraj. Pytraj exposes the flexibility and functionality of cpptraj (more than 100 types of analysis such as coordinate RMSD, radius of gyration, automatic imaging, etc) to the Python ecosystem (NumPy (high performance array), Jupyter Notebook (a web application that allows sharing of documents with live code), pandas (a high-performance data analysis tool), and many others). In addition to supporting analysis via CPPTRAJ functions, pytraj enables users to perform trajectory analysis in parallel with different parallelization schemes (OpenMP from CPPTRAJ, multiprocessing in Python or MPI from mpi4py). Pytraj also supports trajectory visualization in Jupyter Notebook by using NGL Viewer as backend.

Features

- support more than 80 types of data analyses (rmsd, radgyr, autoimage, pca, clustering,...)
- read/write various file formats (.nc, .dcd, .trr, .pdb, .mol2, ...)
- fast (core codes were written in C++ and Cython) support parallel calculation with trivial installation (openmp, multiprocessing, mpi, ...)
- interactive analysis with large trajectory data that does not fit to memory
- many more with comprehensive tutorials: http://amber-md.github.io/pytraj/latest/tutorials/

```
# load pytraj
In [1]: import pytraj as pt

# load amber netcdf file
In [2]: traj = pt.iterload('tz2.nc', 'tz2.parm7')

In [3]: traj
Out[3]:
pytraj.TrajectoryIterator, 101 frames:
Size: 0.000503 (GB)
<Topology: 223 atoms, 13 residues, 1 mols, non-PBC>


# perform analysis
In [4]: data = pt.molsurf(traj, mask='!@H=')

In [5]: data
Out[5]:
array([ 1308.20650442,  1282.93883304,  1266.44646555, ...,  1168.2443778 ,
        1178.70291104,  1180.73997721])

# plotting
# write to disk in DCD format
In [6]: pt.write_traj('charmming.dcd', traj, overwrite=True)
```

Integrate with NGL trajectory viewer in Jupyter notebook



Integrate with Python's ecosystem

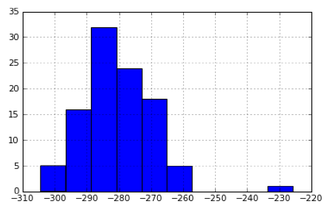
File Edit View Insert Cell Kernel Help

```
In [1]: import pandas as pd
        pd.set_option('precision', 3)

        import pytraj as pt

        traj = pt.iterload('tz2.nc', 'tz2.parm7')
        traj
Out[1]: pytraj.TrajectoryIterator, 101 frames:
        Size: 0.000503 (GB)
        <Topology: 223 atoms, 13 residues, 1 mols, non-PBC>

In [2]: energy_df = pt.energy_decomposition(traj, igb=8, dtype='dataframe')
        energy_df[['bond', 'angle', 'dihedral']].head(8)
```

Out[2]:

	bond	angle	dihedral
0	0.015	128.545	111.611
1	0.014	105.065	105.392
2	0.013	103.520	93.031
3	0.016	94.561	105.522
4	0.013	99.508	105.850
5	0.016	95.671	109.626
6	0.014	100.938	91.888
7	0.013	95.189	99.322

```
In [3]: %matplotlib inline

        energy_df['tot'].hist()
Out[3]: <matplotlib.axes._subplots.AxesSubplot at 0x2aaadddd6b00>
```



Easy parallel calculation

```
In [7]: import pytraj as pt
        import numpy as np
        np.set_printoptions(precision=2)

        traj = pt.iterload('tz2.nc', 'tz2.parm7')
        traj
Out[7]: pytraj.TrajectoryIterator, 101 frames:
        Size: 0.000503 (GB)
        <Topology: 223 atoms, 13 residues, 1 mols, non-PBC>


In [8]: # Serial
        pt.rmsd(traj, ref=0, mask='@CA')
Out[8]: array([  1.95e-07,   2.55e+00,   4.22e+00, ...,   4.97e+00,   5.54e+00,
                4.83e+00])

In [9]: # Parallel
        pt.pmap(pt.rmsd, traj, ref=0, mask='@CA', n_cores=8)
Out[9]: OrderedDict([('RMSD_00001',
                      array([  1.95e-07,   2.55e+00,   4.22e+00, ...,   4.97e+00,   5.54e+00,
                               4.83e+00]))])
```

Software availability

- **pytraj**: https://github.com/Amber-MD/pytraj
  - pytraj will be available in AMBER16
  - doc: http://amber-md.github.io/pytraj/
- **cpptraj**: https://github.com/Amber-MD/cpptraj
  - also in AMBER
- **nglview**: https://github.com/arose/nglview

References

1. Daniel R. Roe and Thomas E. Cheatham, III, "PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data". J. Chem. Theory Comput., 2013, 9 (7), pp 3084-3095.
2. Rose, A. S.; Hildebrand, P. W., NGL Viewer: a web application for molecular visualization. *Nucleic Acids Research* **2015**.
3. http://jupyter.org/