Molecular Dynamics - simulating and analyzing protein (P53) + (53PB1)

[1]Arya HajiTaheri

[2] Research faculty: Dr. Ebalunode

Abstract. It is difficult to solve and predict the motion of molecules analytically. Molecular Dynamics (MD) systems are used to simulate biomolecules and to study the behavior of atoms in controlled environments. In cancer treatment, MD is used to examine the binding of tumor antigen P53 with tumor suppressor 53PB1. The data can be used to produce a protein that is stable and requires less energy to bind with other biomolecules. Using numeric integrators and visualization tools, such as OpenMM, the stability and the binding energy of the protein can be determined; and therefore, can be used to develop proteins that are less prone to mutations and more likely to prevent cancer.

Keywords: Molecular Dynamics, P53, Tumor Suppressor, 53PB1, Protein visualization, Binding Energy

Introduction

Molecular dynamics is a method for studying the physical movements of atoms and molecules. The simulation combines concepts from quantum and statistical mechanics. Molecular systems typically consist of a vast number of particles, it is impossible solve such complex systems analytically; MD simulation circumvents this problem by using numerical methods. Furthermore,
MD simulations are chaotic; small differences in the initial conditions quickly lead to very different trajectories. Using these calculations, the efficiency of the protein can be measured and be used to combat cancer.

$$\hat{H}\Psi = E\Psi$$

$$\left(\hat{T} + \hat{V}\right)\Psi = E\Psi$$

$$\left(-\sum_i \frac{\nabla_i^2}{2} - \sum_{i,I} \frac{Z_I}{|\mathbf{R}_I - \hat{\mathbf{r}}_i|} + \sum_{i \neq j} \frac{1}{|\hat{\mathbf{r}}_i - \hat{\mathbf{r}}_j|}\right)\Psi(\mathbf{r}_i; \mathbf{R}_I) = E(\mathbf{R}_I)\Psi(\mathbf{r}_i; \mathbf{R}_I)$$

Kinetic energy operator — Electron nuclear attraction — Electron electron repulsion — Electron wavefunction — Electronic energy

on the time-independent Schrödinger's equation. The time-dependent Schrödinger equation predicts that wave functions can form standing waves, called molecular orbitals.

Time independent nonrelativistic Schrödinger equation is used for a single particle moving in an electric field.

Methods

All ground-state quantum chemistry is based

Discretization:

Schrödinger's equation is nearly impossible to solve, so approximate methods are used. The method used to approximate is using empirical functions. To find the total, kinetic, and potential energy of the protein solution, the Verlet integration was used,

$$x(t + \Delta t) = x(t) + v(t)\Delta t + \frac{1}{2}a(t)\Delta t^2 + O(\Delta t^4)$$

$$v(t + \Delta t) = v(t) + \frac{1}{2}(a(t) + a(t + \Delta t))\Delta t + O(\Delta t^2)$$

Where:

- t is the time,
- x is the ensemble of the position vector of N objects,
- V is the scalar potential function,
- a is the acerbation of every particle

Time step is limited by fast degrees of freedom (bond vibrations). Verlet is used to calculate trajectories of particles.

[3]Experiment Environment

To visualize the protein, the solution [4]KZY was obtained in .pdb (protein database) format. Using ipython and OpenMM, the solution was integrated and visualized using the following pseudo code:

Begin

    Import openmm libraries

        Open write file ('chainname.csv')
load pdbfile = ;chainname.pdb'
create forcefield ('amber99sb.xml', 'tip3p.xml')

Initialize new model
= (pdb.topology, pdb.position)
model.addHydrogen(forcefield)
model.addssolvent(forcefield)
system =
forcefield.createsystem(model)

Integrate =
intergratorfunction(temperature,time)
Create new simulation =
simulate(model.topology, system, integrator)
set model.positoin
simulaton.minmizeEnergy()
writefile
simulation('filename',timestep)
simulation.step = 20000

End

While running on the following system:

```
-----------------
System Information
-----------------
       Time of this report: 8/10/2017, 00:40:23
              Machine name: ARYA
          Operating System: Windows 10 Home 64-bit (10.0, Build 15063)
                  Language: English (Regional Setting: English)
       System Manufacturer: Dell Inc.
              System Model: Inspiron 7559
                      BIOS: 1.1.7
                 Processor: Intel(R) Core(TM) i7-6700HQ CPU @ 2.60GHz (8 CPUs), ~2.6GHz
                    Memory: 16384MB RAM
       Available OS Memory: 16244MB RAM
                 Page File: 7312MB used, 11364MB available
               Windows Dir: C:\WINDOWS
           DirectX Version: DirectX 12
---------------
Display Devices
---------------
                 Card name: NVIDIA GeForce GTX 960M
              Manufacturer: NVIDIA
                 Chip type: GeForce GTX 960M
                  DAC type: Integrated RAMDAC
               Device Type: Render-Only Device
            Display Memory: 12188 MB
          Dedicated Memory: 4065 MB
             Shared Memory: 8122 MB

                 Card name: Intel(R) HD Graphics 530
              Manufacturer: Intel Corporation
                 Chip type: Intel(R) HD Graphics Family
                  DAC type: Internal
               Device Type: Full Device
       Driver Problem Code: Unknown
            Display Memory: 8250 MB
```
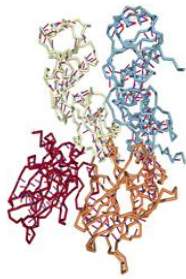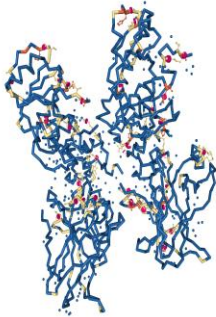
Results:

Using the code above, the four chains in the solution 1KZY were simulated and the following was visualized:
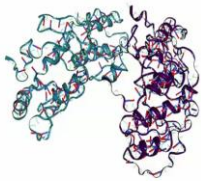
Protein 1KZY

Where P53(red and orange) and 53PB1 (white and blue) were inserted in a system with 300 kelvins, around room neutralized with temperature. The protein was calcium ions and water

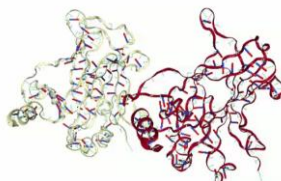and the solution became like so:

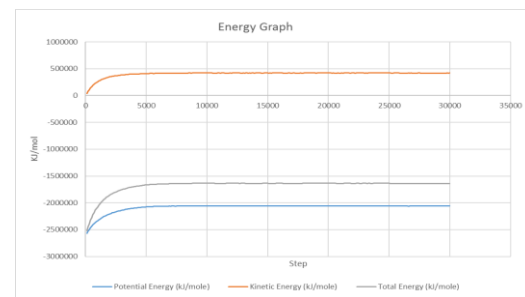Afterwards the chains were separated into AB (P53) and CD (53PB1) and the same procedures were done for those sub chains.
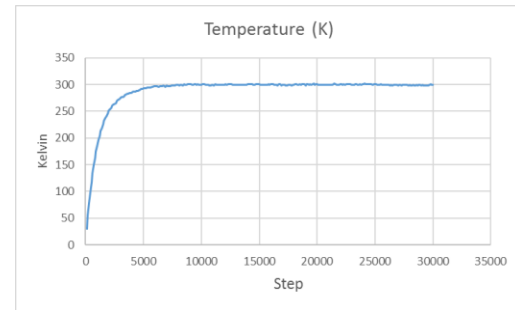
Chain AB:

Chain CD

Using the Verlet intergrator, the energy and the temperature were calculated after 30000 timesteps.

The solution became stable after 8000 steps of running time. Furthermore, the protein released 300 Kelvin to the environment. To measure the binding energy, the energy data from sub chains AB and CD was needed. The same integrator was used on chains AB and CD and similar patterns were observed. The sub chains converge into equlibrium after 8000 time steps, and 300 Kelvin was relased into the enviornment.

Chain AB:
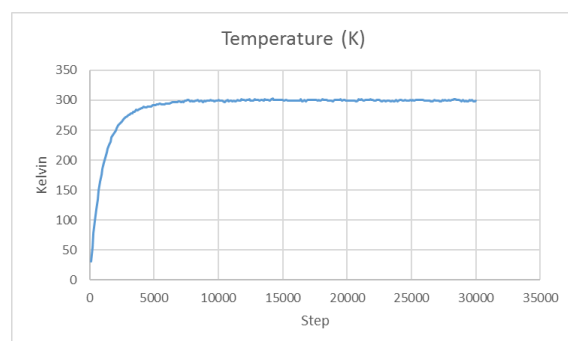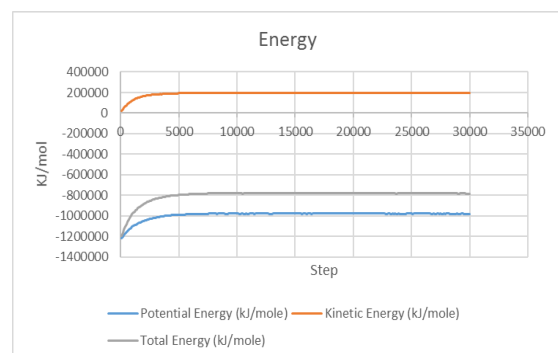
Chain CD:
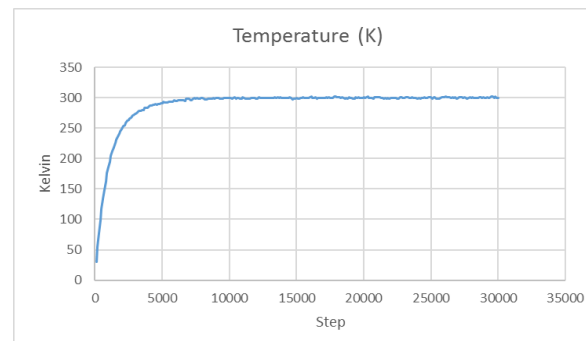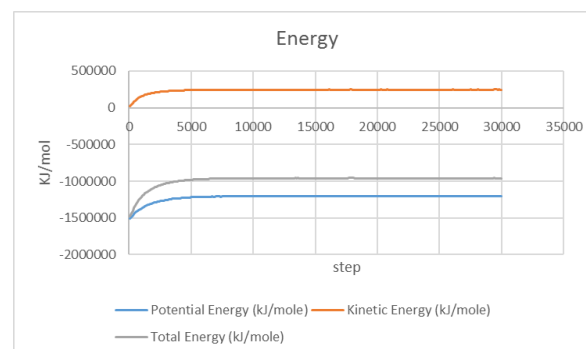
The total energy values from the previous graphs are collected to calculate the binding energy (B.E.) using the following:

BE = Energy for (P53) + Energy for (53PB1) - Energy for complex (53BP1 - P53)

The binding energy at 300 Kelvin is – 103038 KJ. Binding energy is negative because if it was positive, or zero, then the atoms would separate and escape into space. The lower the B.E. results in a better protein for cancer treatment since less energy and effort is needed to bind to the target biomolecules.

Conclusion:

The solution 1KZY encodes a tumor suppressor protein containing transcriptional activation, DNA binding, and oligomerization domains. Mutations in this gene are associated with a variety of human cancers.

The TP53 gene is the most frequently mutated gene (>50%) in human cancer. TP53 gene encodes proteins that bind to DNA and regulate gene expression to prevent mutations of the genome. By using MD and visualization, we can change the structures of the protein and/or obtain a better substitute suppressor that is more stable and requires less energy to bind to other molecules. The new protein bind that has the most stable state and the lowest binding energy can be used in creating a new drug, and perhaps a possible solution in treating cancer and other lethal genetic mutation.

References

[1] Center for Advanced Computing and Data Systems, Vistas in Advanced Computing, Arya HajiTaheri, University of Houston: ahajitaheri2@uh.edu

[2] Center for Advanced Computing and Data Systems, Research Faculty, Jerry Ebalunode, University of Houston: jebalunode@uh.edu

[3] System information: https://github.com/arya-ht/MolecularDynamics---Protein-Simulation-1KZY/blob/master/sysinfo.txt

[4] 1KZY solution .pdb: http://www.rcsb.org/pdb/explore.do?structureId=1kzy