MD Simulations on Formamidopyrimidine-DNA Glycosylase Protein

Haoyue Guo

Stony Brook University at Stony Brook, New York 11790.

Abstract

DNA oxidization is one of the deleterious mutations in organisms, among which 8-oxoguanine (8-oxoG)[1] is commonly found. Since the mutation is detrimental, most organisms have enzymes that can target and repair it. The Formamidopyrimidine-DNA Glycosylase (Fpg) protein is one of the 8-oxoG repair protein. The crystal structures of Fpg-DNA complex have been achieved in previous study.[1,2,3] But the process of how the Fpg protein recognizes 8-oxoG and binds to it is still unclear. In this paper, the Fpg protein was studied with molecular dynamics (MD) simulations. The TIP3P explicit solvent model was implemented in these simulations. During the simulations, the conformation of the active sites of the Fpg protein changed and the hydrogen bonds between the protein and DNA were broken. These results prove the assumptions that during the recognition of 8-oxoG, there are a series of conformation changes of the Fpg protein, and they form a complex in an energetic preferred pathway.

Introduction

One of the by-products of aerobic respiration is oxidized DNA, among which 8-oxoguanine (8-oxoG)[1], the product of oxidized guanine (G), is commonly found and widely studied. Compared with G, oxoG has C=O at 8-position instead of C-H, and N-H at 7-position instead of N:, in the five-atom heterocylic ring. The oxoG will base pair with A, while G will base pair with C. Thus, the oxidation of G will cause a G-C to T-A mutation in genome, as shown in Fig 1. Since the mutation is deleterious, most organisms have enzymes that can target oxoG and repair it. One of the oxoG repair protein is Fpg protein, also known as MutM[2, 3]. Fpg is a DNA glycosylase, which can pair with opposite C in DNA and catalyze the elimination of oxoG lesion nucleotides. The previous study suggests that Fpg perform extrahelical base excision on single base lesions.[4] The structure study shows that Fpg eliminates damages nucleotide from the helix and inserts it into an active-site pocket on the enzyme.[5] Crystal structures only show the result of the process, but how Fpg recognizes oxoG in vast excess of normal guanine bases in the whole genome is still unclear. It is believed that during the dynamic process, there are a series of conformation changes, and the Fpg-DNA complex is supposed to get the lowest-energy conformation.

Recent structural studies suggest that during the repair process, the oxoG nucleotide is rotated out of the DNA helix.[6] A segment of the Fpg protein near the active site is ordered without DNA, but is disordered when bind to normal DNA, which may conduce to the lesion base recognition.[7]

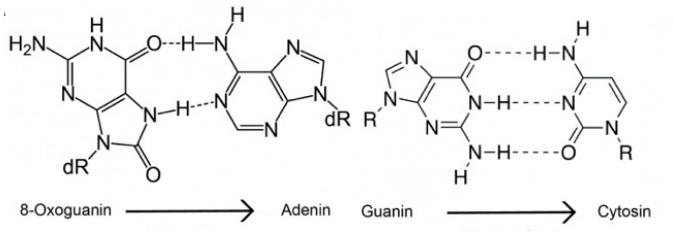

Fig 1 oxoG-A base pair and G-C base pair

The Fpg-DNA complex has two kinds of crystal structures. PDB:1R2Y[5], which is with extrahelical endpoint; PDB: 2F5O[4], which is with intrahelical endpoint in Fig 2. From Fig 3, there are five hydrogen bonds between Fpg protein and DNA in the 1R2Y molecule: OP2 in deoxyguanosine – N in asparagine; O6 in oxoguanine – N in valine; O6 in oxoguanine – N in arginine; O6 in oxoguanine – N in tyrosine; N7 in oxoguanine – O in serine.

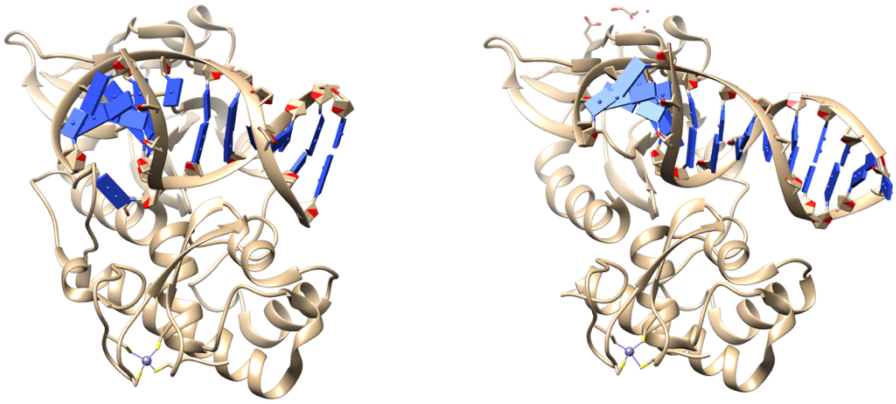

Fig 2 PDB: 1R2Y[5] and 2F50[4]

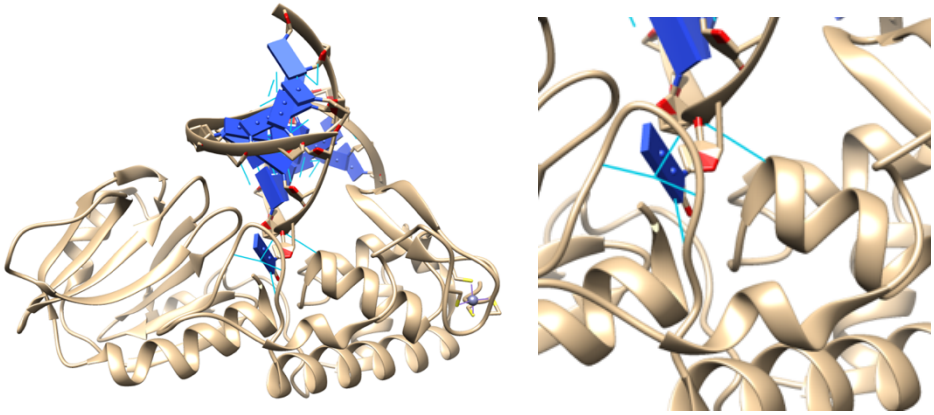

Fig 3 Hydrogen bonds (blue) between protein and DNA in PDB: 1R2Y[5].

Methods

In this paper, all simulations were run in the AMBER ff14SB, with TIP3P explicit solvent model employed. The LEaP module was used to build the initial structure, with the TIP3P solvent model for the solvent structure. The PMEMD module was utilized for MD simulations.

Preparations.

Download PDB files from the PDB bank.

Load known coordinates from PDB. There is a zinc (2+) ion in the protein, which is not in the active site. It is assumed to be a structure ion. Therefore, I keep it before the simulation. Remove the water molecules. Remove the DNA and save the protein as PDB file.

Delete all the headers in the PDB file.

LEaP.

Use LEaP module to build the initial structure. LEaP is a program that can build basic model and AMBER coordinate parameters. Use TIP3P solvent model to build the solvent structure with the box size of 10.0 for MD simulation in explicit solvent. Save the parameters and coordinates for gas structure and solvent structure.

Input file:

```
set default PBradii mbondi2
source oldff/leaprc.ff99SB
loadoff ./rizzo_amber7.ionparms/ions.lib
loadamberparams ./rizzo_amber7.ionparms/ions.frcmod
```

loadamberparams ./rizzo_amber7.ionparms/heme.frcmod

loadamberprep ./rizzo_amber7.ionparms/heme.prep

loadoff ./rizzo_amber7.ionparms/y2p.off

loadamberparams ./rizzo_amber7.ionparms/y2p.frcmod

REC = loadpdb 1r2y.rec.pdb

saveamberparm REC 1r2y.rec.gas.prm7 1r2y.rec.gas.rst7

solvateBox REC TIP3PBOX 10.0

saveamberparm REC 1r2y.rec.wat.prm7 1r2y.rec.wat.rst7

charge REC

quit

/

PMEMD.

MD simulations in solvent.

Do the initial equilibration and adjust the structure to the simulation environment.

Use PMEMD module to minimize the structure with explicit solvent model and do the simulation.

01mi.in: equilibration

A 1000-step minimization with 5.0 kcal·mol^{-1}· Å^{-2} restraint on all heavy atoms except the hydrogens.

Input file:

01mi.in: equilibration

```
 &cntrl
 imin = 1, maxcyc = 1000, ntmin = 2,
 ntx = 1, ntc = 1, ntf = 1,
 ntb = 1, ntp = 0,
 ntwx = 1000, ntwe = 0, ntpr = 1000,
 cut = 8.0,
 ntr = 1,
 restraintmask = ':2-275 & !@H=',
 restraint_wt = 5.0,
/
```

02md.in: equilibration

A 50 ps MD simulation at 298.15K (initial temperature = 298.15K, themostat temperature = 298.15K) with 5.0 kcal·mol^{-1}· Å^{-2} restraint on all heavy atoms except the hydrogens to bring the system to a more natural state with lower energy.

03mi.in: equilibration

04mi.in: equilibration

05mi.in: equilibration

Three 1000-step minimizations with 2.0, 1.0, 0.05 kcal·mol^{-1}· Å^{-2} restraint on all heavy atoms except the hydrogens.

06md.in: equilibration

07md.in: equilibration

Two 50 ps simulations at 298.15K with 1.0, 0.5 kcal·mol^{-1}· Å^{-2} restraint on all

heavy atoms except the hydrogens

08md.in: equilibration

09md.in: equilibration

Two 50 ps simulations at 298.15K removes the restraint from the backbone and ligand.

10.md.in: production

11.md.in: production

Two 500 ps simulations at 298.15K.

Input file:

11md.in: production

```
 &cntrl
    imin = 0, ntx = 5, irest = 1, nstlim = 500000,
    temp0 = 298.15, tempi = 298.15, ig = 71287,
    ntc = 2, ntf = 1, ntt = 1, dt = 0.002,
    ntb = 2, ntp = 1, tautp = 0.5, taup = 0.5,
    ntwx = 500, ntwe = 0, ntwr = 500, ntpr = 500,
    cut = 8.0, iwrap = 1,
    ntr = 1, nscm = 100,
    restraintmask = ':2-274 & @ZIN,CA,C,N', restraint_wt = 0.1,
 /
```

12.md.in: production

A 2000 ps fully unrestraint simulation at 298.15K to get a production.

Input file:

12md.in: production

```
 &cntrl
    imin = 0, ntx = 5, irest = 1, nstlim = 2000000,
    temp0 = 298.15, tempi = 298.15, ig = 71287,
    ntc = 2, ntf = 1, ntt = 1, dt = 0.002,
    ntb = 2, ntp = 1, tautp = 0.5, taup = 0.5,
    ntwx = 500, ntwe = 0, ntwr = 500, ntpr = 500,
    cut = 8.0, iwrap = 1,
    nscm = 100,
 /
```

MD simulations without solvent.

A 1000 steps energy minimization.

Input file:

energy minimization

```
 &cntrl
        imin = 1, maxcyc=1000,
        ntx = 1,
        ntwr = 100, ntpr = 100,
        cut = 999.0,
        ntb=0, igb = 8, gbsa = 0,
            saltcon = 0.0,
```

/

A 100million steps of fully unrestrained MD simulation, each of length 0.002ps at, with initial temperature at 200K and thermostat temperature at 300K. Save the snapshots to trajectory file every 50000 steps.

Input file:

langevin dynamics simulations

&cntrl

ntx = 1, irest=0,

imin = 0, nstlim = 100000000, dt = 0.002,

ntt = 3, gamma_ln=1., temp0 = 300.0, tempi=200.,

ntc= 2, ntf = 2,

igb=8, ntb = 0, saltcon=0.,

ntwx = 50000, ntwe = 0, ntwr = 50000, ntpr = 50000,

cut = 1000.0,

nscm = 500,

/

Results

LEaP results.

The initial gas phase structure was built from LEaP. Loading the molecules in VMD, the structure built from LEaP (Orange in Fig 4) overlaps well with the XRD structure (Pink in Fig 4). Note that there was a loop in initial structure instead of a β-sheet in the XRD structure.

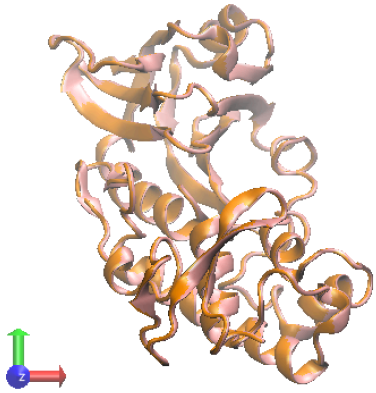

Fig 4 The initial gas phase structure (orange) compared with XRD structure (pink).

The initial solvent phase structure was built from LEaP with TIP3P explicit solvent model for the MD simulations. The size of the solvent model should range between 8-12, which depends on the structure size. If the size is too small, the MD simulation will produce errors. From Fig 5, setting the box size to 10.0, the whole protein was surrounded with water molecules and the protein was almost in the center of the model. Loading the molecules in VMD, the structure built from LEaP (yellow in Fig 5) overlaps well with the XRD structure (pink in Fig 5).

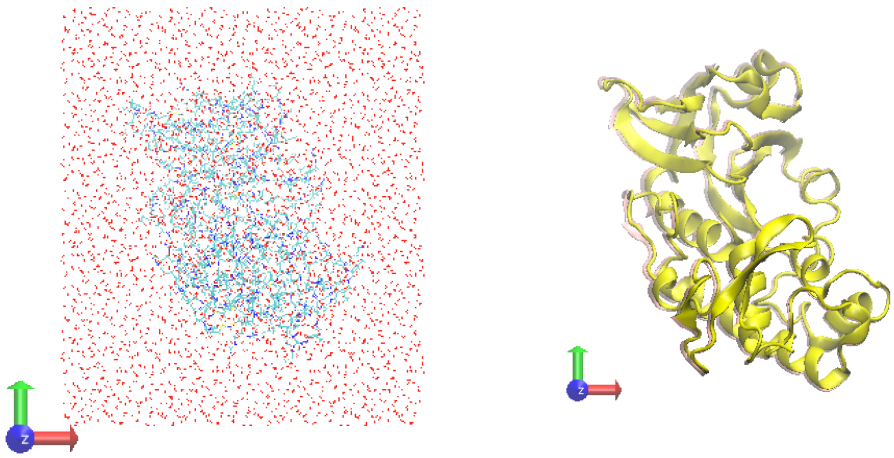

Fig 5 The initial solvent phase structure (yellow) compared with XRD structure (pink).

Results of simulations in solvent.

Minimization results.

From Fig 6, the energy of the initial structure built from LEaP was pretty high, especially in the van der Waals (VDW) energy. It is much likely that some atoms were too close to each other, which resulted in large forces and instability. At 1000 steps, the total energy dropped from -1.0629E+5 kcal·mol^{-1} to -1.4719E+5 kcal·mol^{-1}. However, the VDW energy increased from 1.0860E+4 kcal·mol^{-1} to 2.0732E+4 kcal·mol^{-1}. From Fig 7, after the final minimization, the total energy dropped from 1.0629E+5 kcal·mol^{-1} to -1.6617E+5 kcal·mol^{-1}, but the VDW energy was still pretty high.

```
  NSTEP       ENERGY          RMS            GMAX         NAME    NUMBER
     1      -1.0629E+05     1.4434E+01     1.2817E+03     HD2      1235

BOND    =        79.7161  ANGLE   =      599.9368  DIHED      =     2365.4263
VDWAALS =     10860.4022  EEL     =  -129762.6795  HBOND      =        0.0000
1-4 VDW =      1277.6338  1-4 EEL =     8293.3571  RESTRAINT  =        0.0000


  NSTEP       ENERGY          RMS            GMAX         NAME    NUMBER
  1000      -1.4719E+05     7.5580E-01     9.2000E+01     CG       2679

BOND    =     10006.3422  ANGLE   =      603.6735  DIHED      =     2474.1243
VDWAALS =     20732.5728  EEL     =  -189918.1196  HBOND      =        0.0000
1-4 VDW =       916.1731  1-4 EEL =     7802.9167  RESTRAINT  =      190.7699
EAMBER  =   -147382.3171
```

Fig 6 First minimization result.

```
                FINAL RESULTS


   NSTEP       ENERGY          RMS            GMAX         NAME    NUMBER
    1000      -1.6617E+05     2.8506E-01     4.1060E+01     NE2      3273

BOND    =    12062.3257  ANGLE   =      623.5815  DIHED      =     2416.0157
VDWAALS =    24944.4428  EEL     =  -214690.4037  HBOND      =        0.0000
1-4 VDW =      845.0962  1-4 EEL =     7627.1143  RESTRAINT  =        0.3585
EAMBER  =  -166171.8275
```

Fig 7 Final minimization result.

Loading the structure after the minimizations, the Fpg protein was surrounded with water molecules (Fig 8) and overlapped well with the XRD structure (pink in Fig 8). Compared with the initial structure (yellow in Fig 9), even the energy changed a lot, the structure did not change much (cyan in Fig 9).

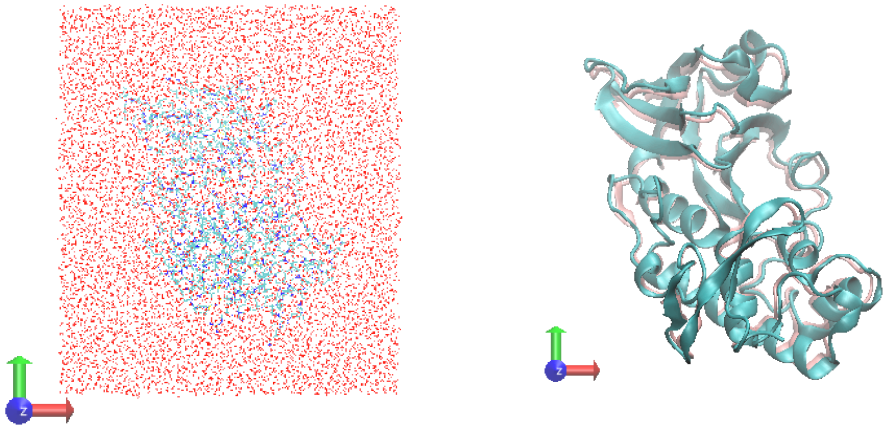

Fig 8 Minimization result (cyan) compared with XRD structure (pink).

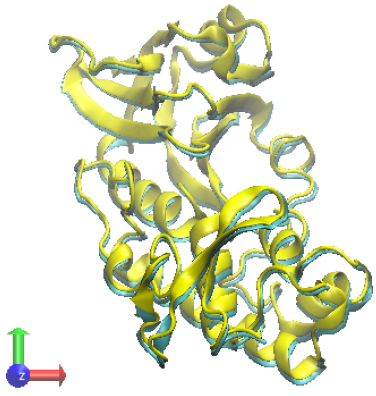

Fig 9 Minimization result (cyan) compared with initial structure (yellow).

MD Simulation results.

Six MD simulations were run with different restraints and one MD simulation was run with complete unrestraint.

Loading the results from restrained MD simulation in VMD, the structure (cyan in Fig 10) coincided well with the XRD structure (pink in Fig 10). The backbone rmsd values of the restrained MD simulation results relative to the native crystal structure were pretty low (Fig 11).

There were hundreds of hydrogen bonds and large VDW forces in the Fpg protein, which can stabilize the protein. Even with a restraint weight of 0.1, the conformation of the protein did not change much. To find the series of conformational changes in active site, a fully unrestrained simulation was run follow-by.

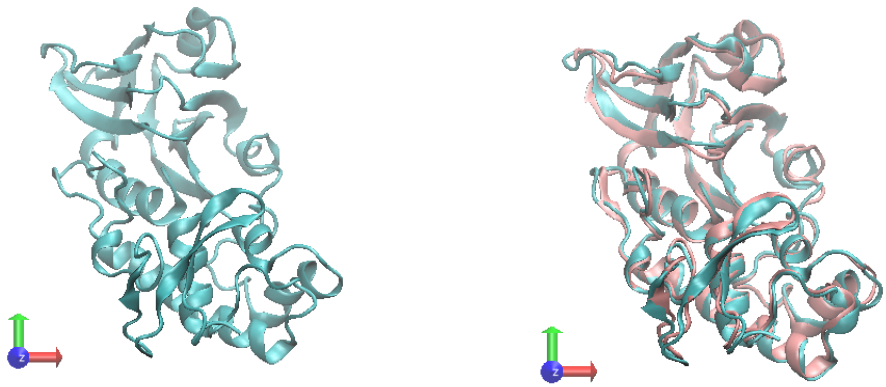

Fig 10 Restrained simulation results(cyan) compared with XRD structure(pink).

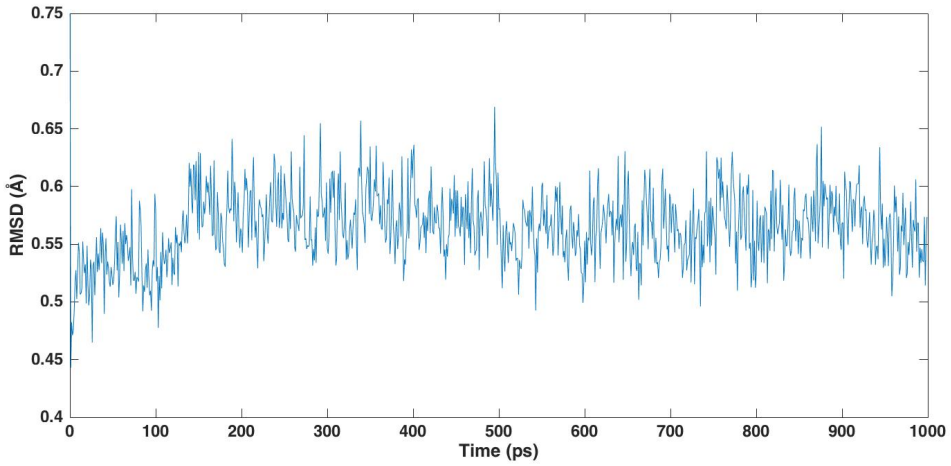

Fig 11 Backbone rmsd values of restrained simulation results relative to XRD structure.

In unrestrained MD simulation, the Fpg protein was much more flexible. It almost ran out of the water box. The best fit structure from the simulation, with the backbone rmsd value of 1.1089, overlapped well with the native crystal structure (Fig 12). However, comparing the conformation with the backbone rmsd value of 2.6405 with native crystal structure, we found that there was structural shift in the loops of the active site (Fig 14). There were no hydrogen bonds found between the protein and DNA. Comparing the worst fit conformation with the backbone rmsd value of 3.4829 with crystal structure (Fig13), we found the active site was in a more open state. The loops that formed hydrogen bonds with DNA in the crystal structure were shifted to a farther place. It was reasonable to assume that in this process, there were breaks of hydrogen bonds. But more simulations on Fpg-DNA complex were needed to prove this. The structure with more hydrogen bonds tend to be more stable. These results also indicated that the Fpg protein recognized and the bonded to DNA in an energetic preferred pathway. But more simulations are needed to make a conclusion.

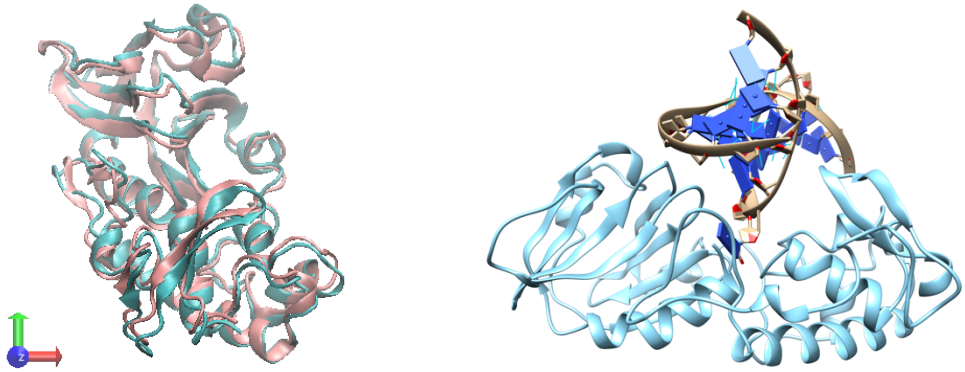

Fig 12 Unrestrained simulation results(cyan) (1271-1.1089) compared with XRD structure(pink).

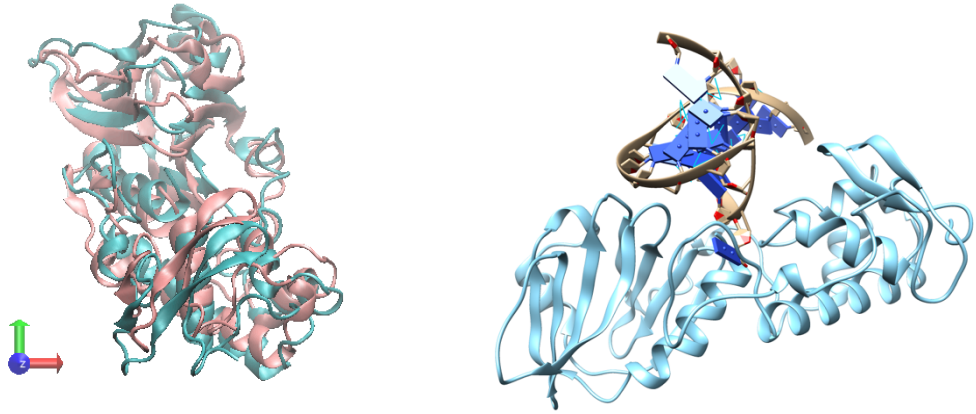

Fig 13 Unrestrained simulation results(cyan) (3871-3.4829) compared with XRD structure(pink).

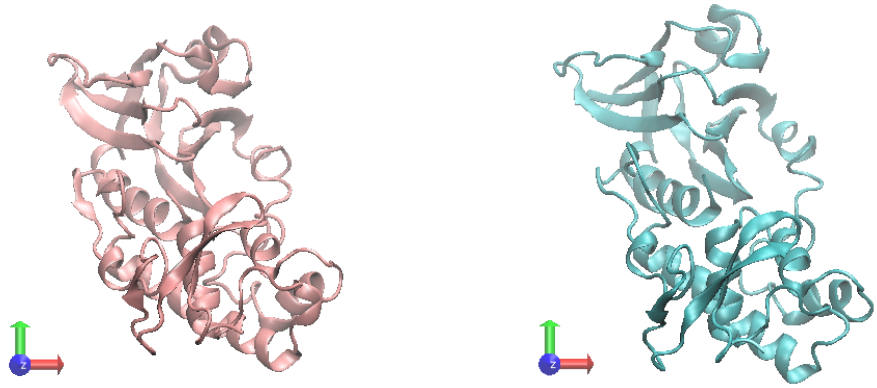

Fig 14 Simulation results(cyan) (2841-2.6405) compared with XRD structure(pink).

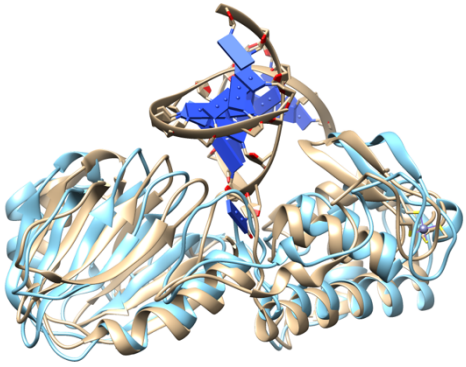

Fig 15 Simulation results(cyan) (2841-2.6405) interact with DNA.

From Fig 16, The backbone rmsd values of the simulation results relative to the XRD structure kept raising from about 0.5 to 3.5.

The total energy of the Fpg protein as a function of time during unrestrained MD simulation was shown in Fig 17. The potential energy of the Fpg protein as a function of time during MD simulation is shown in Fig 18. The temperature as a function of time during MD simulation is shown in Fig 20. The temperature was fluctuant around 298K.

To find the relation between the potential energy and the conformations of the Fpg protein, the potential energy as a function of backbone rmsd values were plotted in Fig 19. However, there was no clear correlation between them.

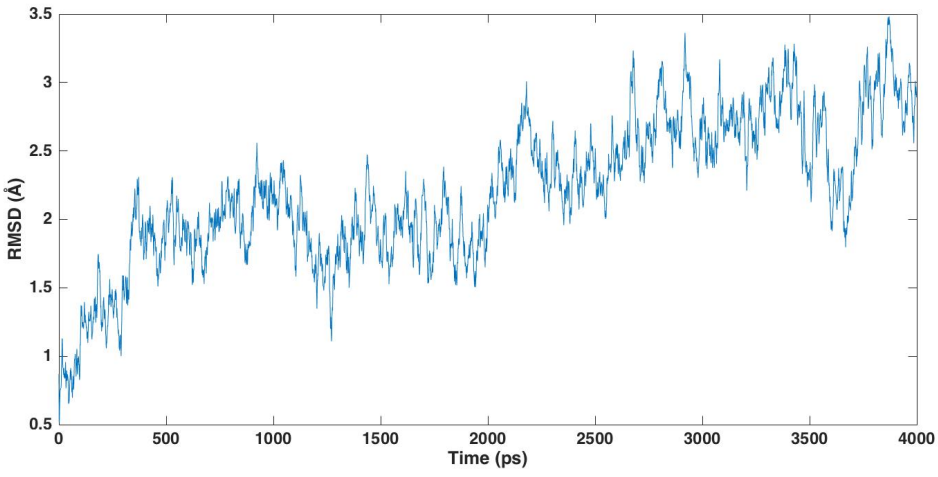

Fig 16 Backbone rmsd values of unrestrained simulation results relative to XRD structure.

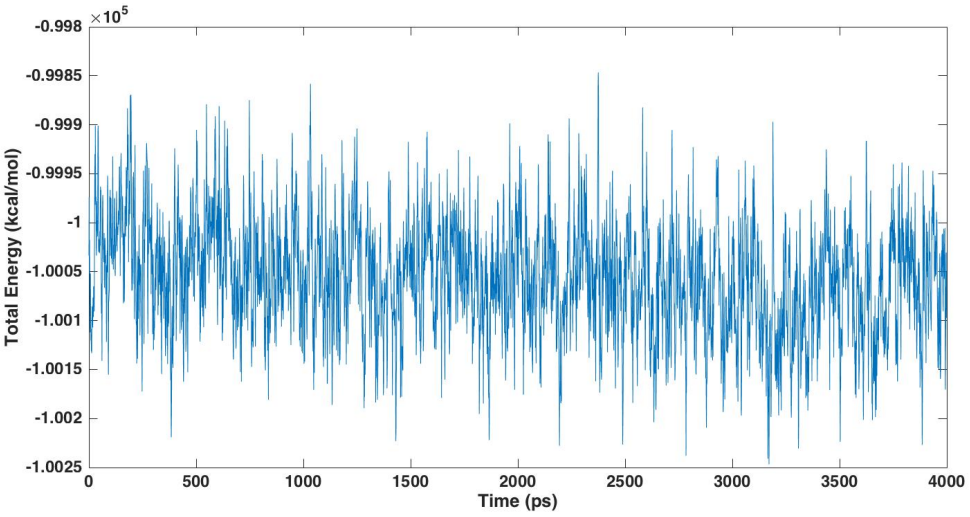

Fig 17 Total energy of the Fpg protein as a function of time during unrestrained MD simulation.

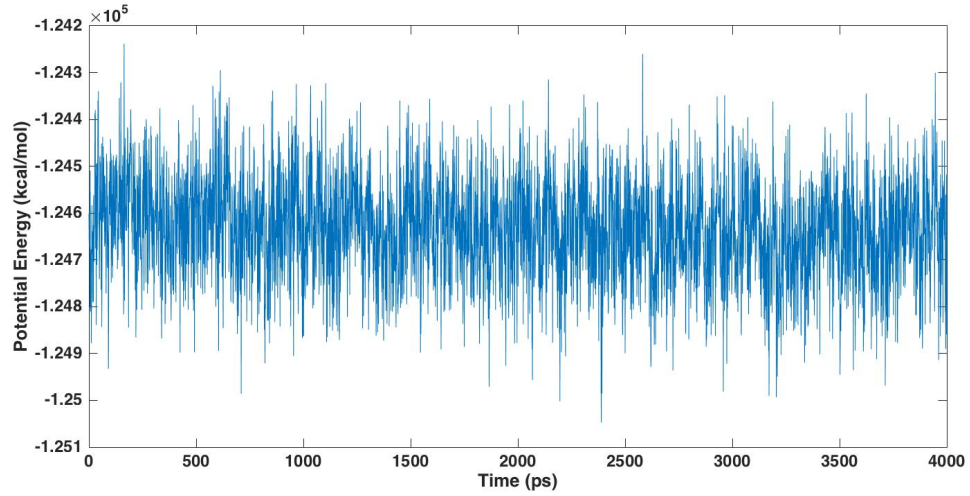

Fig 18 Potential energy of the Fpg protein as a function of time during unrestrained MD simulation.

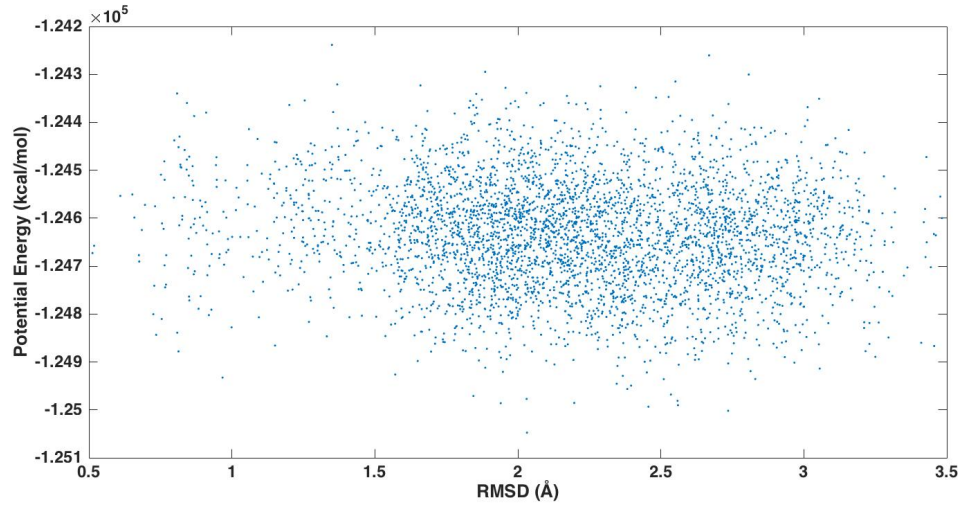

Fig 19 Potential energy of the Fpg protein relative to rmsd values during unrestrained MD simulation.

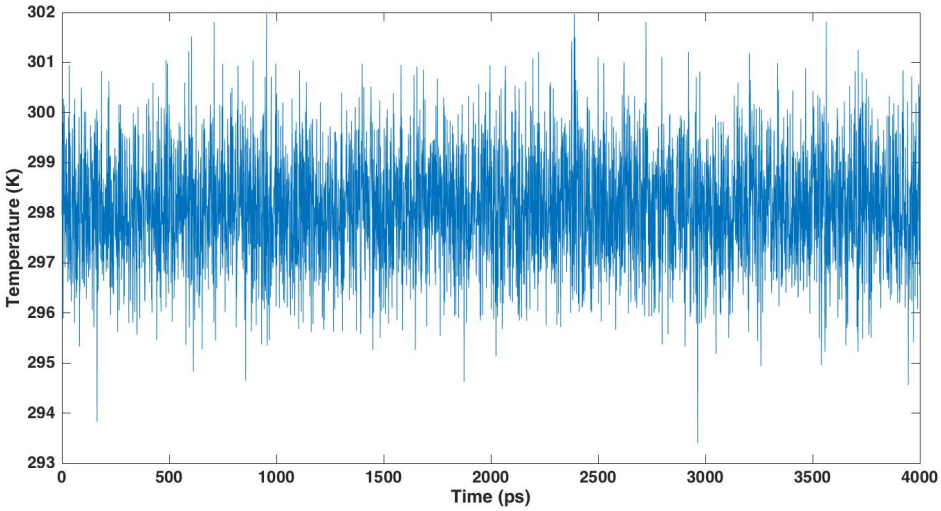

Fig 20 Temperature during unrestrained MD simulation as a function of time.

Results of simulation without solvent.

During the MD simulations, I came across with some errors. The structure after the MD simulation was a mass of threads and the rmsd of the simulation results relative to the XRD structure kept rising from 2.5 to 25. To find out if there were problems with the structure built from the LEaP module, I ran a fully unrestrained MD simulation without solvent.

Minimization results.

The energy of the structure built from LEaP was minimized. Loading the structure after minimization (yellow in Fig 21) in VMD, it overlapped well with the XRD structure (pink in Fig 21).

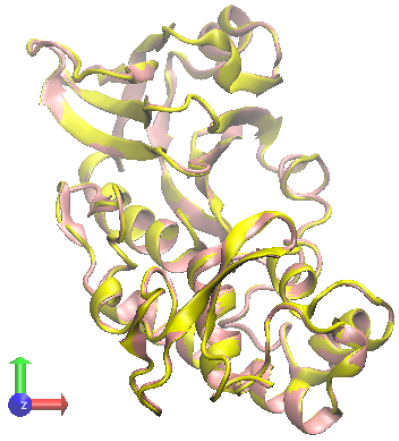

Fig 21 Minimization result (yellow) compared with XRD structure (pink).

Simulation results.

Loading the results from MD simulation in VMD, the structure shifted much compared with the XRD structure (Fig 22). The rmsd values of the simulation results relative to the XRD structure raised from about 1.0 to 8.0 (Fig 9). The Fpg protein in the simulation without solvent was apparently much more flexible.

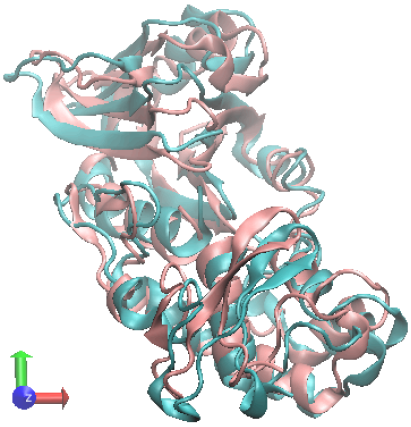

Fig 22 Simulation results (cyan) compared with XRD structure(pink).

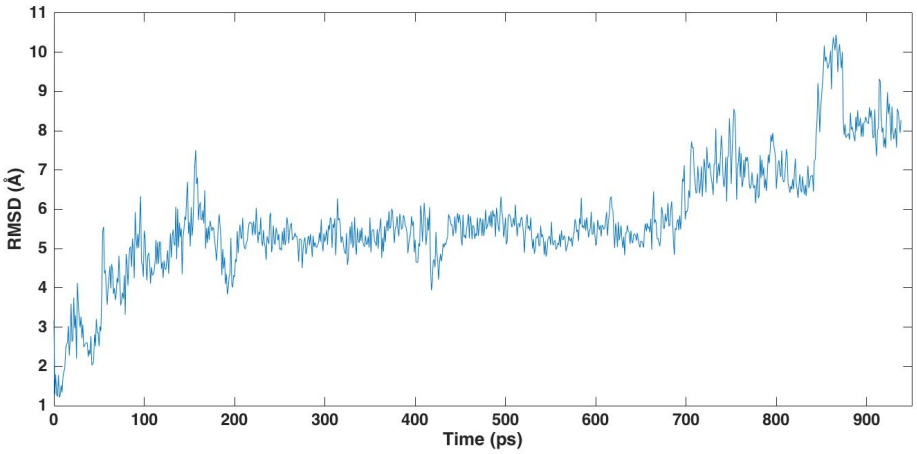

Fig 23 Rmsd values of simulation without solvent results relative to XRD structure.

Discussions

After the minimizations, the total energy of the structure dropped from 1.0629E+5 kcal·mol^{-1} to -1.6617E+5 kcal·mol^{-1}. However, the VDW energy increased from 1.0860E+4 kcal·mol^{-1} to 2.0732E+4 kcal·mol^{-1}. It was a little bit confusing because the VDW energy was expected to decrease, even to a negative number, after the minimizations. This may because when the nucleotides were removed from the complex, there were some bulks in the protein.

After the nucleotides were removed, the Fpg protein was much more flexible in the unrestrained simulations. There were structural shifts in the simulations and no hydrogen bonds were found between the simulation structures and lesion DNA. This did not work with my assumption that during the simulations there were breaks and formations of hydrogen bonds. It is supposed to be more accurate if MD simulations are run with Fpg-DNA complex. More work will be done about this in the future. During the recognition and binding to lesion DNA, the Fpg protein is supposed to have a series of conformational changes. This process is likely to be in an energetically preferred pathway. The energy changes will be studied in the future.

Besides, it was tricky to align the simulated structures with the crystal structures. The backbone rmsd with all αC works well in most cases. But if we want to focus on the changes in the active site, we may choose the α-helix or the β-sheets or the loops to align for more accurate analysis.

To figure out how the Fpg protein recognizes and binds to DNA, more simulations on the Fpg-DNA complex are needed.

Conclusions

The Fpg protein was studied with MD simulations in TIP3P explicit solvent model. During the dynamic process, there were a series of conformation changes as expected. The structure of the simulation results shifted when comparing with the crystal structure. During the unrestrained simulations, the conformation of the active site changed between slightly open and closed state. There were no hydrogen bonds found between the structures from the simulation results with lesion DNA. It indicated that the Fpg-DNA complex had lower energy because it formed hydrogen bonds in between. To get more accurate results, more simulations on the Fpg-DNA complex are needed in the future work. The energy change in this process is also to be studied in the future.

Acknowledgements This work was achieved under the guidance of Prof. Carlos Simmerling and Prof. Robert C. Rizzo. Great respects are devoted to Prof. Carlos Simmerling and Prof. Robert C. Rizzo. for their tutoring in computational biology. Gratitude is expressed to Chuan Tian and He Huang for their help in the MD simulations. Appreciation is given to Jiaye Guo for her suggestions on this project.

References

[1] Michaels, M. L.; Miller, J. H. The GO system protects organisms from the mutagenic effect of the spontaneous lesion 8-hydroxyguanine (7, 8-dihydro-8-oxoguanine). *Journal of bacteriology*, **1992**, *174(20)*, 6321.

[2] Grollman, A. P.; Moriya, M. Mutagenesis by 8-oxoguanine: an enemy within. *Trends in genetics*, **1993**, *9(7)*, 246-249.

[3] Fowler, R. G.; White, S. J.; Koyama, C.; Moore, S. C.; Dunn, R. L.; Schaaper, R. M. Interactions among the Escherichia coli mutT, mutM, and mutY damage prevention pathways. *DNA repair*, **2003**, *2(2)*, 159-173.

[4] Banerjee, A.; Santos, W. L.; Verdine, G. L. Structure of a DNA glycosylase searching for lesions. *Science*, **2006**, *311(5764)*, 1153-1157.

[5] Fromme, J. C.; Verdine, G. L. DNA lesion recognition by the bacterial repair enzyme MutM. *Journal of Biological Chemistry*, **2003**, *278(51)*, 51543-51548.

[6] Gilboa, R.; Zharkov, D. O.; Golan, G.; Fernandes, A. S.; Gerchman, S. E.; Matz, E.; Shoham, G. Structure of formamidopyrimidine-DNA glycosylase covalently complexed to DNA. *Journal of Biological Chemistry*, **2002**, *277(22)*, 19811-19816.

[7] Fromme, J. C.; Verdine, G. L. Structural insights into lesion recognition and repair by the bacterial 8-oxoguanine DNA glycosylase MutM. *Nature Structural & Molecular Biology*, **2002**, *9(7)*, 544-552.