

Mechanistic Study of the Chlorella Virus DNA-Ligase by QM/MM Method



By

Amitosh Gautam

Reg. No.: 20121087

Under the guidance of

Dr. Arnab Mukherjee

Department of Chemistry

IISER Pune, India

To

Department of Chemistry

Indian Institute of Science Education and Research (IISER)
Pune

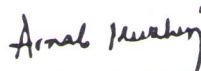
Certificate

This is to certify that this dissertation entitled "**Mechanistic Study of the Chlorella Virus DNA-Ligase by QM/MM Method**" for the fulfilment of BS-MS dual degree programme at the Indian Institute of Science Education and Research, Pune by Amitosh Gautam carried out by the candidate at the Indian Institute of Science Education and Research (IISER), Pune, under my supervision. The work presented here has not been included in any other thesis for the award or any degree from any other university or institution.



Amitosh Gautam

20121087.



Dr. Arnab Mukherjee (Research Supervisor)

Associate Professor

IISER Pune.

Date: March 20, 2017

Declaration

I hereby declare that the thesis entitled "**Mechanistic Study of the Chlorella Virus DNA-Ligase by QM/MM Method**" submitted for the fulfillment of the BS-MS dual degree programme at Indian Institute of Science Education of Research (IISER), Pune has not been submitted by me to any other University or Institution. This work was carried out at the Indian Institute of Science Education of Research (IISER), Pune, India under supervision of Dr. Arnab Mukherjee.


Amitosh Gautam

20121087.


Dr. Arnab Mukherjee (Research Supervisor)

Associate Professor

IISER Pune.

Date: March 20, 2017

Acknowledgement

First, I would like to thank my mentor, **Dr. Arnab Mukherjee** for his advice, guidance and encouragement at all times. I would always be indebted to him for keeping faith in me, keeping me motivated and helping me develop a scientific outlook. He has been a constant source of inspiration for me.

Secondly, I would like to thank to my TAC member **Dr. Srabanti Chaudhury** for providing a continuous guidance.

I would also like to thank all the lab members. I learnt a lot from them all. Especially **Reman Kr. Singh**, in gratitude for countless times he has helped me unconditionally. Thank you for being ever supportive.

It would not have had been possible without my friends at IISER. You guys provided me constant shipment for fun and joy with a pinch of salt at times. This experience would have had been incomplete without you.

Dedicated to my beloved Family

Contents

1. Abstract:	9
2. Introduction:	10
2.1. Structural aspects of <i>Chlorella</i> virus DNA Ligase.....	10
2.2. Mutational studies of <i>chlorella</i> virus DNA ligase:	12
3. Methods and Techniques:	14
3.1. System Selection and description:	14
3.2. Loop-modeling:	15
3.3. Simulation details:.....	17
3.4.Partitioning the system: QM and MM partition:	18
3.5. Nudged Elastic Band (NEB) calculations:	19
3.6. QM/MM GAMESS optimization:	21
4. Results:	22
4.1. Structural changes during the equilibration:	22
4.2. The minimum free energy path	24
4.2.1. QM Model 1: Proton transfer from water:	24
4.2.2. QM Model 2: Approaching of the active site lysine group followed by proton transfer:	29
4.2.3. QM Model 3: Proton transfer from the active lysine group:	31
4.2.4. QM Model 4: A concerted mechanism.....	33
4.2.5. QM Model 5: A constrained mechanism:.....	35
5. Discussion:.....	37
6. Conclusion:.....	38
7. References:	40

List of Figures

Fig. 1: Scheme of the three step mechanism of the DNA-ligase enzyme.

Fig. 2: Open and closed form conformation of chlorella virus DNA-ligase enzyme.

Fig. 3: Active site residues of the DNA-ligase enzyme.

Fig. 4: Crystal structure of the chlorella virus DNA-ligase enzyme.

Fig. 5: Loop modeling of the product state of the enzyme.

Fig. 6: Generating the reactant state from the modeled product state.

Fig. 7: QM region of the enzyme.

Fig. 8: Force projections on a replica in NEB calculations.

Fig. 9: Movement of the ARG298 and magnesium ions towards the active site.

Fig. 10 to Fig. 12: QM model 1

Fig. 10: Scheme for QM model 1.

Fig. 11: Energy profiles and distances plots for the scheme.

Fig. 12: Highest energy intermediate structure for the mechanism.

Fig. 13 to Fig. 15: QM model 2

Fig. 13: Scheme for QM model 2.

Fig. 14: Energy profiles and distances plots for the scheme.

Fig. 15: Highest energy intermediate structure for the mechanism.

Fig. 16 to Fig. 18: QM model 3

Fig. 16: Scheme for QM model 3.

Fig. 17: Energy profiles and distances plots for the scheme.

Fig. 18: Highest energy intermediate structure for the mechanism.

Fig. 19 to Fig. 21: QM model 4

Fig. 19: Scheme for QM model 4.

Fig. 20: Energy profiles and distances plots for the scheme.

Fig. 21: Highest energy intermediate structure for the mechanism.

Fig. 22 to Fig. 24: QM model 5

Fig. 22: Scheme for QM model 5.

Fig. 23: Energy profiles and distances plots for the scheme.

Fig. 24: Highest energy intermediate structure for the mechanism.

Fig. 25: Stabilization energy profile for the QM model 1.

1. Abstract:

DNA ligase is an enzyme which catalyzes the joining mechanism of single and double stranded breaks (Nicks) in DNA. They are required for DNA replication, repair and recombination. DNA ligase requires either ATP or NAD⁺ as the nucleotide cofactor. An overall mechanism of the enzyme was outlined based on the crystal structures of different DNA ligase enzymes at different steps. DNA ligation, joining of DNA strands, consists of three nucleotidyl transfer steps. The mechanism initiates with the formation of Ligase-AMP intermediate by the interaction of lysine from the active site of the enzyme with nucleotide cofactor (ATP or NAD⁺) in the step 1. In step 2, the formation of DNA-adenylate complex takes place through transferring the AMP nucleotide on 5'-phosphate group of the nicked DNA strand. This is followed by the formation of junction in phosphate backbone through attack of 3'-OH group on DNA-adenylate complex in step 3. These three steps are shown in Fig. 1.¹ All three steps involve divalent metal ions. In this thesis, we have focused on the first step of the reaction. Using QM/MM based nudged elastic band calculation, we have explored various mechanisms and the minimum free energy path and the associate mechanisms.

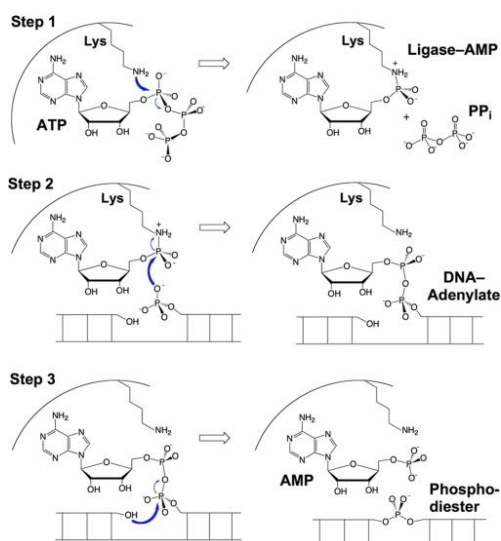


Figure 1. Three-step mechanism of phosphodiester bond formation by DNA-ligase enzyme¹. Reproduced with permission from ref. 1. Copyright 2011 American society for Biochemistry and Molecular Biology.

2. Introduction:

The presence of the nicks in the DNA backbone affects biological processes such as transcription, replication etc. Hence, repairing of the nicks is important for preserving the active maintenance of the genetic elements in the cell. These breaks appear in the DNA strands of the undamaged cells during several processes like replication and recombination. For example, during replication process of human genome approximately 2×10^7 Okazaki fragments are created by discontinuous lagging strand DNA synthesis². In the DNA there are some specific sites where damages can happen easily; these sites are named as DNA lesion. The breaks or nicks in the DNA can be generated either as a consequence of DNA lesion removal by selective repair pathways or directly by a DNA damaging agent. Two hundred thousand single strand breaks are generated per mammalian cell per day by the processing of spontaneous lesions³, whereas exposure to 1 Gy (Gray) of ionizing radiation causes 600-1000 single strand breaks and 16-40 double strand breaks.⁴ DNA ligases are the enzymes which catalyze the joining mechanism of these nicks. These enzymes were discovered by several laboratories in 1967. These enzymes utilize a nucleotide cofactor to catalyze the phosphodiester bond formation. Based on the cofactor, the enzymes are classified into two groups. They are either ATP-dependent or NAD⁺-dependent. All eukaryotic cells contain ATP dependent DNA Ligases while bacteria and viruses contain both ATP and NAD⁺-dependent DNA Ligase enzymes. All DNA Ligases consist of two preserved domains -- an N-terminal nucleotidyl transferase (NTase) domain and a C-terminal oligonucleotide binding (OB) domain. Mammalian and bacterial DNA Ligases have an additional N- and C-terminal domains. These additional domains play an important role in biological processes and in the participation on protein-protein interaction and DNA-protein binding.⁵

2.1. Structural aspects of *Chlorella virus* DNA Ligase

Chlorella virus DNA ligase is the smallest eukaryotic (298 amino acids, weighing 34 kDa) ATP-dependent DNA ligase which consists of only two domains. The enzyme contains the smallest catalytic site among different DNA ligases. N-terminal domain of

the enzyme consists of an active site where the nucleotide cofactor binds. This domain consists of six peptide motifs represented by I, Ia, III, IIIa, IV, and V. (Fig. 3). These motifs show a major role in binding of cofactor through hydrogen bonds to the exocyclic groups of the adenine base and play important roles in one or more steps of the ligation pathway. The C-terminal domain contains motif VI residues which help in the first step of the mechanism. During the first step it orients the phosphates of nucleotide cofactor for an inline attack of the active site lysine⁶. Subsequently, it functions as a DNA binding domain during steps 2 and 3. The OB fold domain is supplely connected to the NTase domain, and it experiences a large rotation between the step 1 and the DNA dependent steps 2 and 3. The motif VI located on the OB fold domain orients towards the active site during the step 1 whereas in steps 2 and 3 the OB domain rotates in such a way that DNA binding residues can interact with the DNA substrate which results the movement of the motif VI residues away from the active site. It also contains a C-terminal latch. The latch represents a disordered shape in the absence of the DNA while it forms a ring shaped structure in the presence of DNA⁷. This latch connects the NTase and OB fold domains and enters the major groove of the DNA. The crystal structures of open and close conformations are shown in Fig. 2.

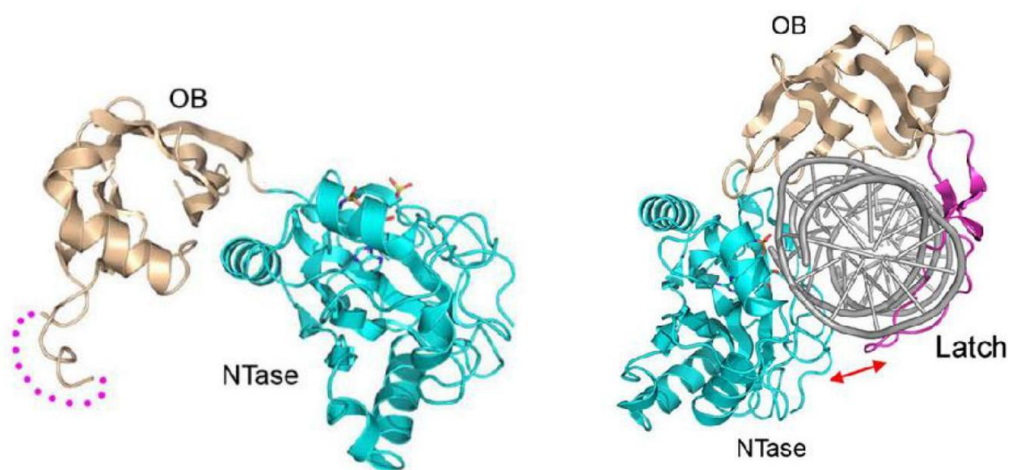


Figure 2. The left side structure represents the open form of the chlorella virus DNA-ligase and the right side structure represents the closed conformation of *chlorella* virus

DNA Ligase¹. Reproduced with permission from ref. 1. Copyright 2011 American society for Biochemistry and Molecular Biology.

It has been observed that the crystal structures of the mammalian and bacterial DNA ligases contain an extra electron density within the active site.⁸This observation leads us to interpret that there is a possibility of the presence of metal ion(s) in the active site. We presumed that these metal ion(s) might coordinate with the active site residues. This extra interaction may lead to an extent of stability during the mechanism. However, there is still a debate about the position and number of metal ions during each step of the reaction because of the inadequate quality of the electron density.

2.2. Mutational studies of *chlorella* virus DNA ligase:

Shuman laboratory has done a rigorous mutational analysis of the active site residues of chlorella virus DNA Ligase. The enzyme differentiates at the DNA binding step between DNA substrates containing a 5' phosphate versus a 5' hydroxyl at the nick. The active site motif I consists of residues TPKIDGIR (residues 25-32). Thr25 is not important for strand joining, whereas conserved residues Lys27, Asp29, Gly30, and Arg32 play significant roles for nick ligation. Lys27 attacks at the alpha phosphate group of the cofactor in step 1. The Lys27 does not show any significant role in step 2 and 3. The mutational analysis reveals that mutating this residue leads to terminate the overall ligation process. Asp29 is important for step 2 but not for step 1 and 3. Mutation of Gly30 does not affect the mechanism, which indicates that the residue does not play any role in the ligation process. Arg32 is important for step 1 and the motif I arginine residue forms a hydrogen bond with the 3'-OH of the ribose sugar of ATP in the T7 DNA ligase co-crystal.⁹ The same hydrogen bond contacts with the 3'-OH of the ribose sugar of the GTP in the chlorella virus mRNA capping enzyme co-crystal⁶ was observed. This observation led us to hypothesize that similar interaction between Arg32 and the ribose sugar is possible for the chlorella virus DNA ligase also during step 1 and step 3.¹⁰

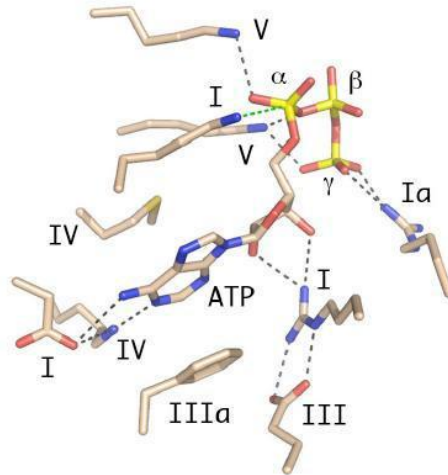


Figure 3. Active site residues of DNA Ligase,¹ reproduced with permission from ref. 1. Copyright 2011 American society for Biochemistry and Molecular Biology.

Motif III consists of two acidic side chains Asp65 and Glu67, which are broadly conserved in other members of the NTase super family. Asp65 forms a salt bridge to Arg32 of motif I residue. Glu67 does not form contact the adenylate in the crystal structure of chlorella virus ligase-AMP intermediate, but it is close to the ribose 2'-O in the structures of T7 ligase with ATP and the capping enzyme-GMP intermediate⁶. Motif IIIa contains conserved aromatic residue Phe98 that stacks on the purine base of the nucleotide.¹¹ Motif IV is preserved among the covalent NTases and contains a (Glu/Asp) Gly dipeptide followed by a triplet of hydrophobic residues. This has been found by the position of a lutetium atom in the ligase-adenylate crystal that the motif IV carboxylate Glu161 might coordinate with a divalent cation bound to the phosphate of AMP.¹¹ Motif V contains residues LLKMKQFKDAEAT (184-196) of which Lys186 and Lys 188 are conserved in all organisms except NAD⁺ dependent DNA Ligases. Lys 186, Lys188, Asp192, and Glu194 are important for step 1. Lys188 is also important for step 3.¹² Motif VI contains residues HEEDR (294-298) which are important for the step 1.¹⁰

Although we have identified the important residues in the active side of the enzyme through mutational studies, but how these residues participate in the reaction is unclear. In the first step of the mechanism it is believed that the active site lysine attacks without transferring a proton, and contains a positive charge on it, which can be higher in

energy. We hypothesized that the proton must be transferred to the nearby residues or to the water medium.

It was shown that all three steps utilize divalent metal ions. But the positions of these ions and their interacting partners during the mechanism are still not clear. Hence, to address these issues we have chosen *chlorella* virus DNA Ligase to study the mechanism using QM/MM method.

We have chosen *chlorella* virus DNA Ligase for two reasons. First one is, it the smallest eukaryotic ATP dependent Ligase having only 298 residues. Second one is crystal structures are available at different steps of mechanism.

3. Methods and Techniques:

3.1. System Selection and description:

We have taken the *chlorella* virus DNA-Ligase enzyme (PDB ID: 1FVI) for the study. The enzyme was captured in the product state of step 1. It has some missing residues at different positions. Also, there is a point mutation in the crystal structure product state to avoid further reactions. The missing residues are 1, 87-90, 206-224, 279-283 and 294-298. The 29th residue was mutated from aspartate to alanine. As we already mentioned that residues 294-298 play an active part in the step 1. The initial structure is shown in fig.4.

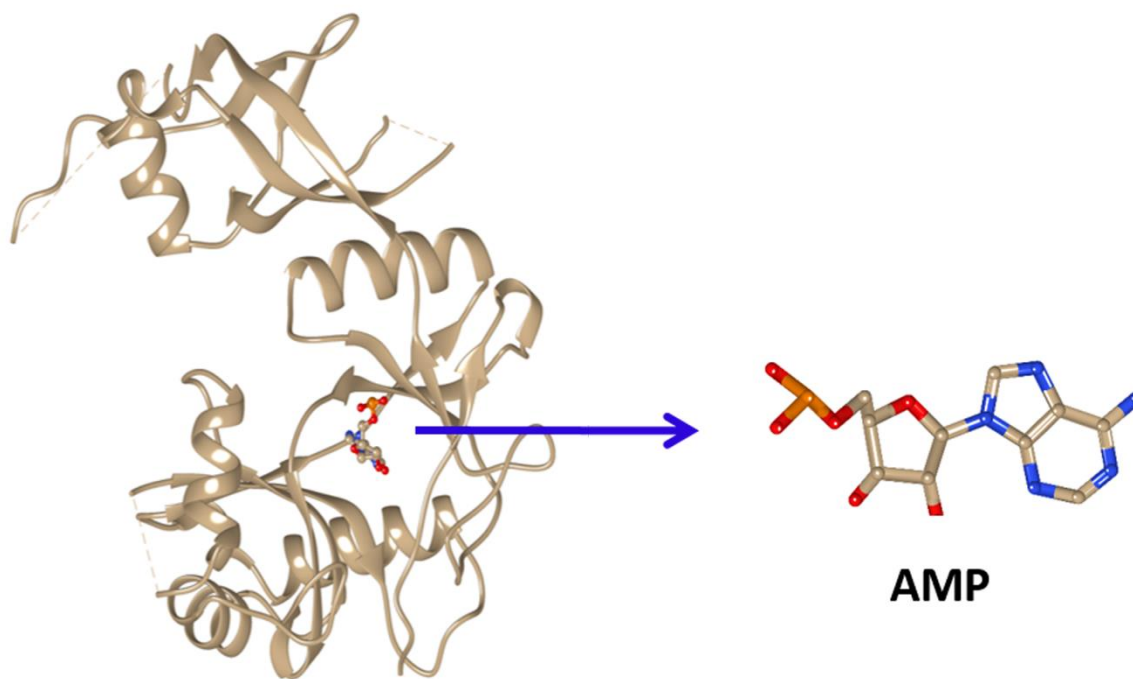


Figure 4: Crystal structure (product state of step 1) of chlorella virus DNA-Ligase (left) and AMP (right) in the active site.

3.2. Loop-modeling:

Hence, to model the missing loops we have used the MODELLER-9.12¹³ software. MODELLER is a program which generates a set of structures of the protein from a given sequence using homology modeling and constraints optimization. The input of the program is the amino acid sequences of the missing loops. The structure is further compared to a database of homologous proteins having similar sequence. The comparison is done to optimize some spatial constraints. These spatial constraints can be residue types, the sequential similarities of two sequences, etc. Finally, through optimizing the constraints we get a set of structures.

Through MODELLER we were able to create five structures out of which we have selected the lowest energy structure for the product state. Through this structure we have filled missing gaps of ligase enzyme by aligning the residues 86 and 91 for the 87-90 gap, 203, 204, 225, 226 and 230 for the 206-224 gap, 278 and 284 for the 279-283

gap, and 292, 293 for the 294-298 gaps. We used modloop server of Fiser loop to model the loop 206-224.¹⁴ We also have replaced the mutated alanine residue in the crystal structure with the aspartate residue. The coordinates of aspartate residue was taken from the crystal structure of Chlorella virus DNA Ligase-adenylate bound to a 5' phosphorylated nick (PDB ID: 2Q2T) by aligning the backbone atoms. The structure of the product obtained after modeling the missing loops is shown in Fig. 5.

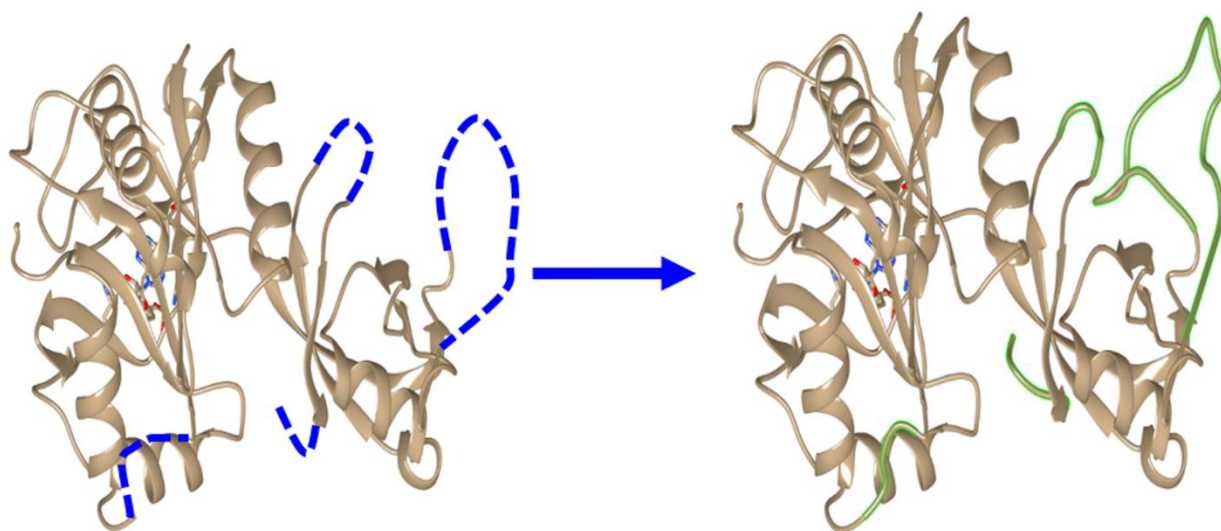


Figure 5: Crystal structure of chlorella virus, missing residues are denoted with the dotted lines (left) and in modeled system with green line (right).

A recent study¹⁵ has shown that magnesium ions affect the acidity of the active site lysine group in a similar kind of system. This variation in the acidity increases the nucleophilicity of the active site lysine group. Hence, we have converted the active site lysine group to the deprotonated form.

Since, no reactant structure was available to study, we have created it from the modeled product structure. As shown in Fig. 1, the reactant state required ATP as the cofactor. Hence, to generate the reactant configuration from the modeled system, we have replaced the AMP moiety of the product structure with ATP as shown in Fig.6. We have taken the ATP coordinates from the crystal structure of ATP-dependent DNA

bacteriophage T7 complex with ATP (PDB ID: 1A0I) which was in the reactant state. We have modified the dihedral angles of the active site Lysine residue and ATP in such a way that the reactant state can be differentiated easily from the product state. We took the modeled reactant structure for further equilibration.

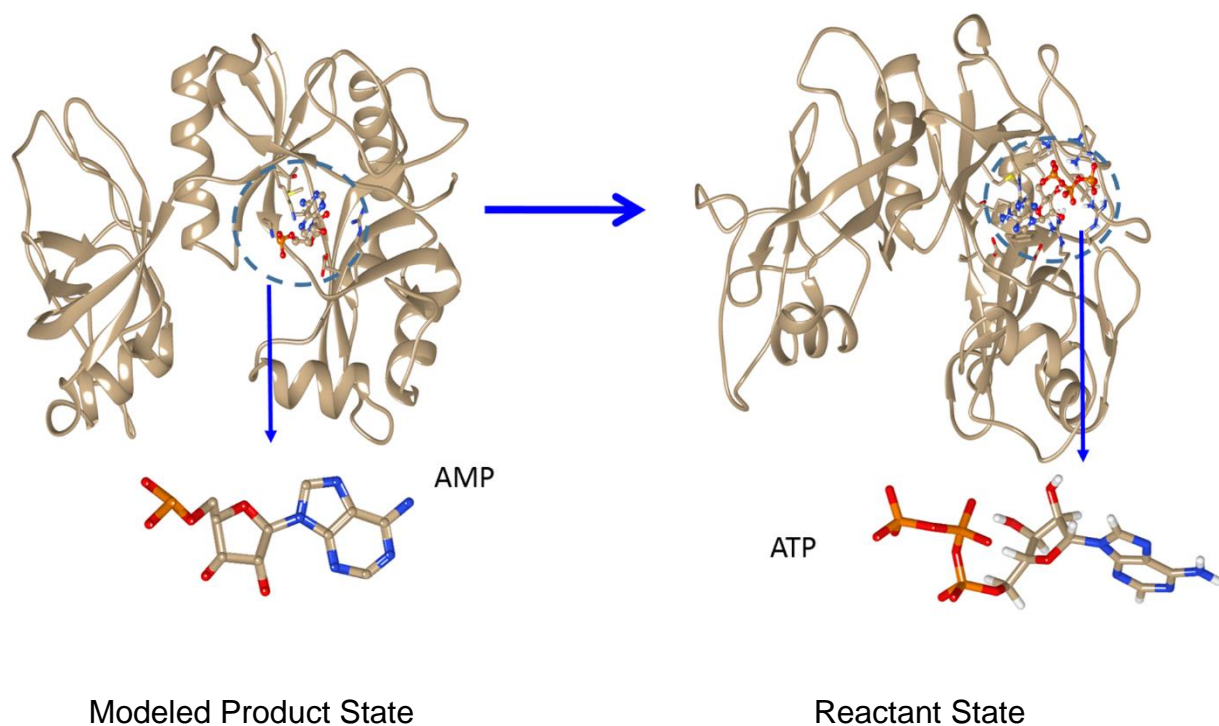


Figure 6: Product state after Loop-modeling (left) and reactant structure (right) modeled from the product state (left).

3.3. Simulation details:

We have generated the parameter file using CHARMM forcefield. We added hydrogen atoms in the reactant structure using HBUILD module of CHARMM¹⁶ software. The system was minimized using 100 steps with steepest descent¹⁷ and 1000 steps with ABNR¹⁸ minimization algorithms. The total charge on the system was -4. We added the physiological concentration of Mg^{+2} and Cl^- (turns out to be 15 Mg^{+2} and 26 Cl^-). The system was solvated in a 81.8 Å cubic box with 16010 water molecules. We used TIP3P¹⁷ water model for solvation. Further, the system was heated to 300 K with restraining the heavy atoms of the enzyme. Thereafter, we have removed the restrains slowly through a series of six simulations; each simulation contained 100 step energy minimization

followed by a 50 ps equilibration. Then the system was equilibrated without any restraints for 12 ns under NPT conditions. From the equilibration we have found that two Mg^{+2} ions are close to the active site. Therefore, we have chosen the structure in such a way that two Mg^{+2} ions were at the minimum distance from the active site.

3.4. Partitioning the system: QM and MM partition:

For further calculations, we have partitioned our system into two regions: (a) QM region (b) MM region. QM region is the center where the reaction takes place while the MM region acts as the environment. We have selected those residues as a part of QM region which have direct interactions with the active site lysine or ATP. We have divided the QM/MM boundary and added link atoms between the non-polarizable bonds of the residues. The QM region consists of residues LYS27, GLU161, LYS186, LYS188, ARG298, ATP, two Mg^{+2} ions and seven water molecules. Total number of atoms in the QM region including link atoms was 83. The QM region alone was used to explore the reaction mechanism in the gas phase. The QM region is shown in Fig. 7.

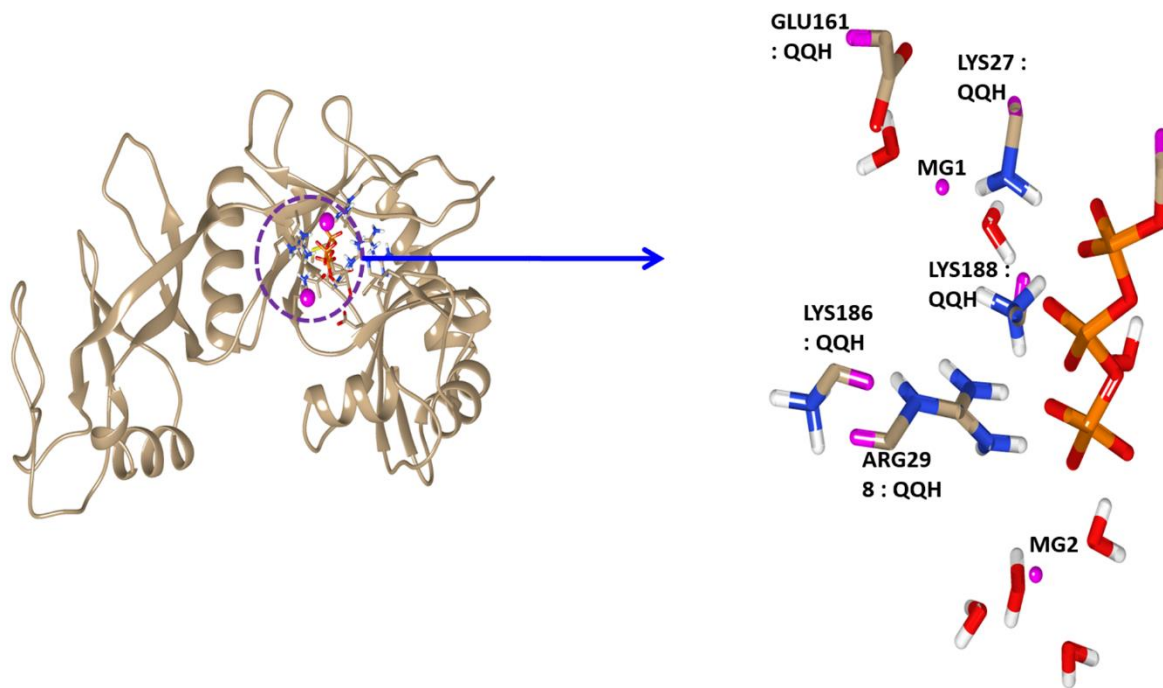


Figure 7: Equilibrated reactant structure (left) with the active site is denoted with dotted circle. Zoomed view of the active site (right).

Subsequently, we have created the gas phase product structure from the reactant one. We then quantum mechanically optimized both the reactant and product structures in the gas phase using CHARMM/GAMESS interface. To find out the reaction mechanism and the minimum free energy path for the reaction between the above reactant and product states, we used Nudged Elastic Band (NEB)¹⁹ approach using CHARMM/GAMESS interface.

3.5. Nudged Elastic Band (NEB) calculations:

We have studied the probable mechanism for the first step of the reaction using Nudged Elastic Band (NEB) method. The method is incorporated in CHARMM software. NEB is a chain-of-state method to find out the minimum energy pathway (MEP) for a transition between two stable states. It starts with the creation of equally spaced replicas or images through a linear interpolation scheme between the two states. The two consecutive images are connected through a spring. These harmonic spring forces between two images ensure that during the optimization no image falls down towards its local minima along the path i.e. the reactant or the product state. Hence, the optimization process is carried out in a force projection scheme in which potential forces act perpendicular while the harmonic spring forces act parallel to the band.

To optimization the replicas we have used Hartree Fock(HF)/6-31G* level of theory. This ensures that the replicas are optimized orthogonal to the path of the reaction providing a MEP. A schematic diagram of force projections on each replica in NEB calculations are shown in the figure 8:

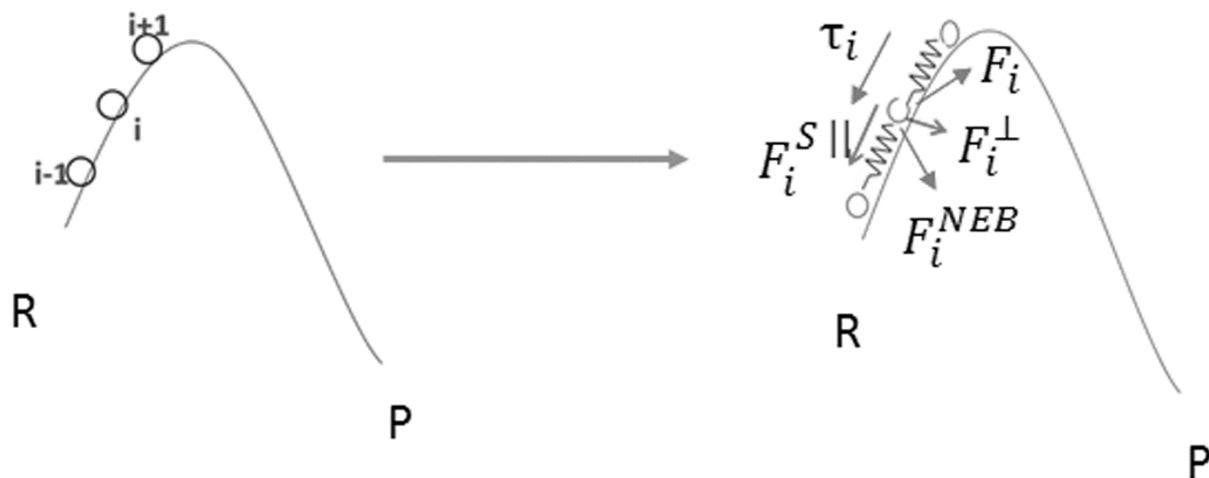


Figure 8: Schematic representation of the NEB method. “R” and “P” represent reactant and product states, respectively. “i” represents the i^{th} replica and the red arrows show the direction of the force on the replica in a NEB calculation. τ_i represents the unit vector along the path, F_i represents the force acting on i^{th} replica. F_i^{\perp} and F_i^{\parallel} show the perpendicular and parallel projections of the force respectively.

The NEB force on image i contains two independent components,

$$F_i^{NEB} = \nabla E(R_i)|_{perp} + F_i^S \cdot \tau_{\parallel} \tau_{\parallel}$$

Here, τ_{\parallel} represents the unit vector along the path. The first term in the above equation is the component of the force due to the potential perpendicular to the band,

$$\nabla E(R_i)|_{perp} = -(\nabla E(R_i) - \nabla E(R_i) \cdot \tau_i)$$

The second term is the spring force parallel to the band,

$$F_i^S \equiv k_{i+1}(R_{i+1} - R_i) - k_i(R_i - R_{i-1})$$

In the final expression, \mathbf{R}_i is the position of the i_{th} image and k is the spring constant.

3.6. QM/MM GAMESS optimization:

The above gas phase optimization of the QM system provided us with the basic chemical reaction. However, it does not have the effect of the protein environment. Therefore, to understand the role of the MM environment, we have performed QM/MM optimization as mentioned below. Further, we have partitioned the system in three concentric spherical regions: (a) QM region (b) Buffer region which consists the residues which are in the radius of 22 Å from the center of the system except the QM region residues and (c) Reservoir region which consists residues which are not in both the QM and buffer regions. Since, we did not use any periodic boundary condition, we had to restrain water molecules at the boundary. For this, we have applied a predefined distance-based potential on water molecules in such a way that water molecules QM region can be freely optimize while the reservoir region will remain fixed during the optimization.

We have created 16 replicas from the equilibrated reactant and product structures. We have replaced the coordinates of the QM region with the coordinates that we obtained from the gas phase NEB optimization.

We have then optimized the replicas having both QM and MM regions (whole protein along with water and ions) through NEB method. The optimization was carried out in such a way that QM region was optimized quantum mechanically while the MM part was optimized classically. Through this optimization we analyzed the effect of the protein environment on the reaction. We also were interested to see if the reaction mechanism itself was different in the presence of the protein environment.

4. Results:

4.1. Structural changes during the equilibration:

From the 12 ns equilibration of the reactant structure of DNA ligase we probed the movement of the ARG298. We observed that ARG298 which was initially far away from the active region (as obtained from the crystal structure of the product state) before the equilibration step has now reached the active site and is coordinated with the gamma phosphate oxygen of ATP. Fig. 9a shows the distance between the ARG298 and the gamma phosphate group of the ATP and its structural change is shown in Fig. 9b. When ARG298 became a part of the active region the distance between the ARG298 and the gamma phosphate group becomes less (see Fig 9a).

As mentioned earlier in the report, there is a debate about the actual position of the Mg^{+2} ions. We found that during the equilibration two of the Mg^{+2} ions came to the active site of the enzyme. The first Mg^{+2} ion was located at the point close to the active site LYS27, α phosphate and GLU161 residue and second Mg^{+2} is located between the β and γ phosphates after the equilibration. Fig. 9c and 9d show the distance between the Mg^{+2} ions and the alpha phosphate group of ATP during equilibration. Corresponding changes in the position of Mg^{+2} are shown in Fig. 9e. Distance become very less when Mg^{+2} coordinated with the active site amino acids (see Fig. 9c, d). Fig. 9f shows the root mean square deviation (RMSD) of the loop regions and the whole protein without loops which clearly indicates that major fluctuation occurred in the loop regions. Whereas there is no significant changes occur in the protein structure shown in Fig. 9f.

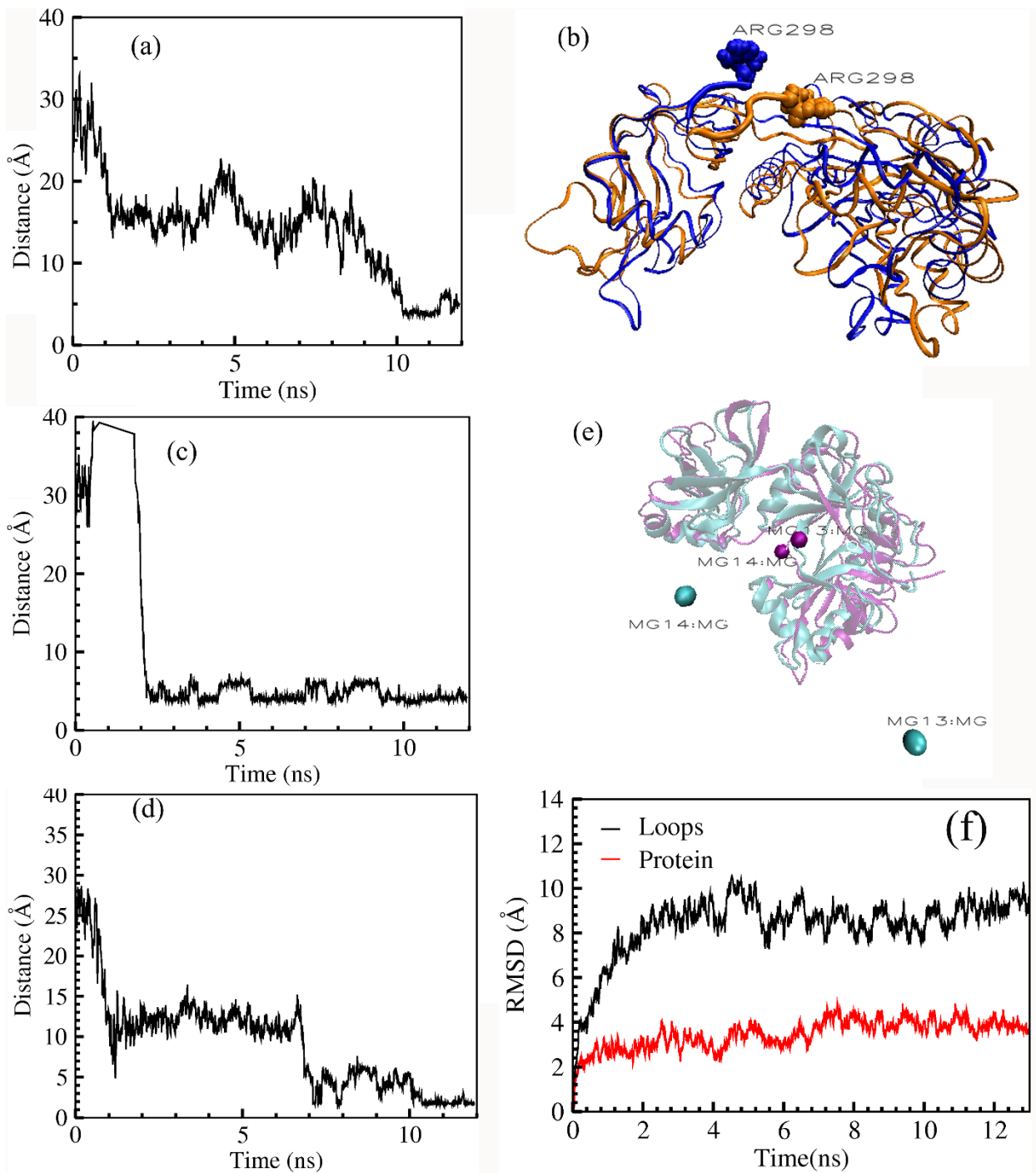


Figure 9: (a) Distance between the ARG298 and the gamma phosphate of ATP. (b) Movement of the ARG298 (blue: modeled structure, orange: after equilibration). Distance between the (c) first Mg²⁺ and (d) second Mg²⁺ from the alpha phosphate of ATP. (e) Representative configuration of Mg²⁺ ions (cyan: modeled structure, purple: after equilibration). (f) Black color represents the RMSD of the loop regions (res 87-90,

res 206-224, res 279-283, res 294-298) and red color represents the RSMD of the whole protein without considering loops.

4.2. The minimum free energy path

Like any other optimization processes, NEB depends heavily on the initial configuration. Therefore, we have constructed different initial pathways (mechanisms) of the reaction and subjected each of them to optimize. Each optimized pathway is the MEP for that particular mechanism. However, the minimum amongst all the MEP should, represent, the appropriate reaction mechanism for a given reaction. We have studied five different mechanisms. We furnish below the mechanism and the energy profile for each of them.

4.2.1. QM Model 1: Proton transfer from water:

Fig. 10 shows the model for the first type of mechanism where the active lysine residue attacks at the alpha phosphate group while the proton is transferred from the water close to the diphosphate group.

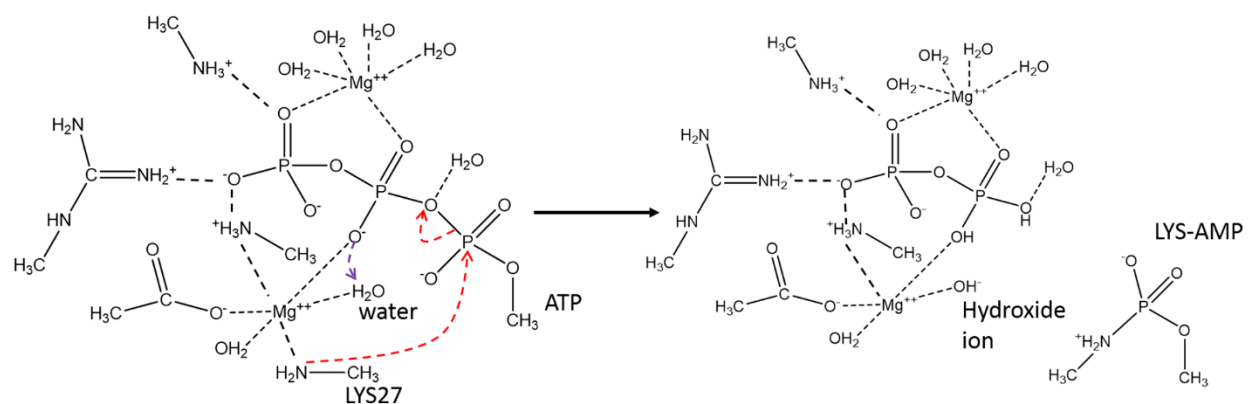
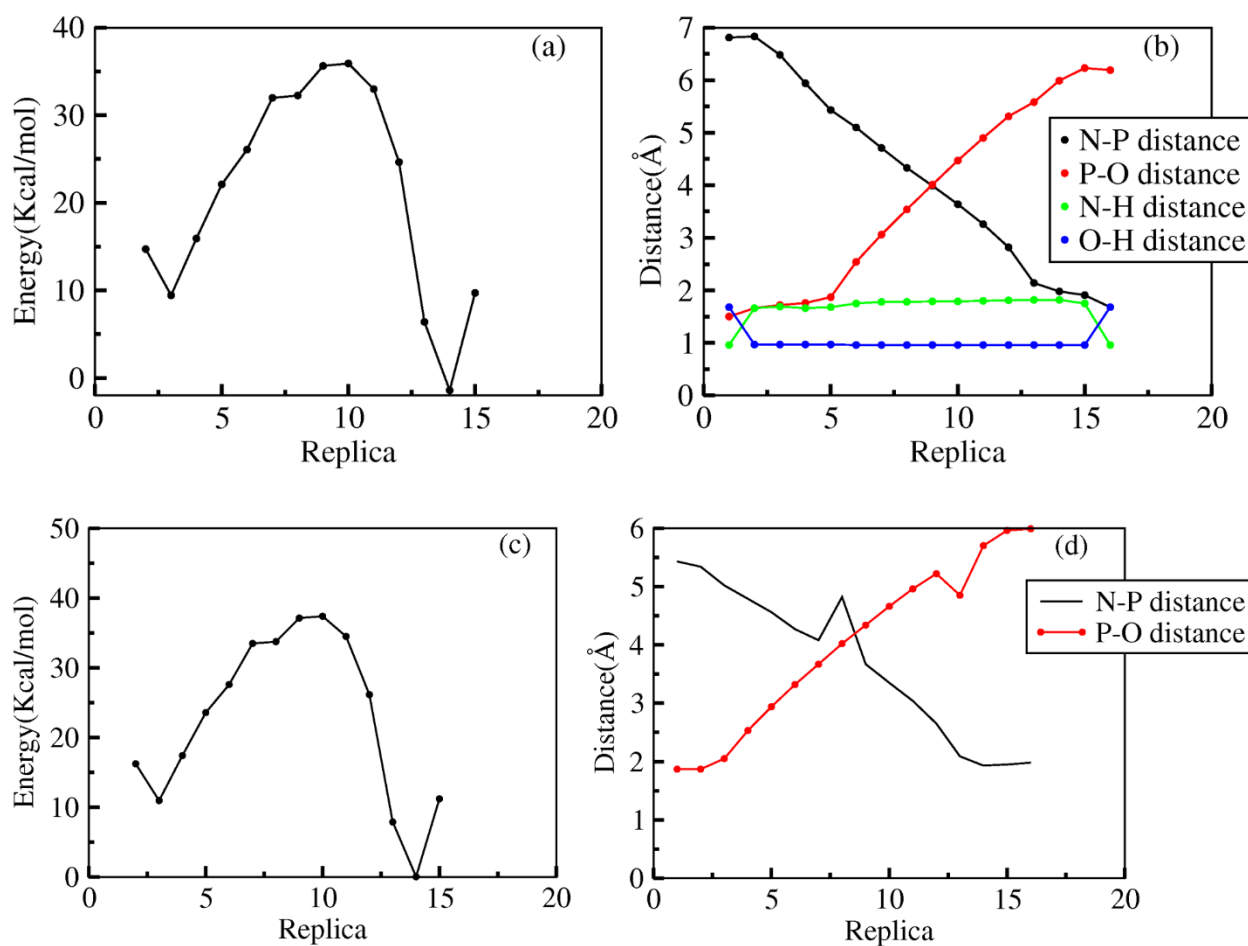


Figure 10: QM1 model: The reactant (left) and product (right) structure of the first mechanism.

We have optimized the QM part taken from the QM/MM model of the reactant state (replica 1). Further, we have created the product structure (replica 16) from the reactant and optimized freely. The intermediate replicas were created by linear interpolation technique from the reactant and the product state. The total number of replicas were 16 including reactant and product states. Product state was found to be more stable than the reactant one. NEB calculations were performed using 50 steps of steepest descent minimization algorithm through CHARMM/GAMESS interface. We have used HF

method and 6-31G* basis set. Spring constant was kept 1000 Kcal/mol/Å² during the optimization. The final energy profile is shown in fig. 11(a). The energy barrier for the mechanism was observed to be 26.3 Kcal/mol. The bond forming distances N-P and O-H and bond breaking distances P-O and O-H distances were also plotted along the profile which is shown in fig. 11(b).



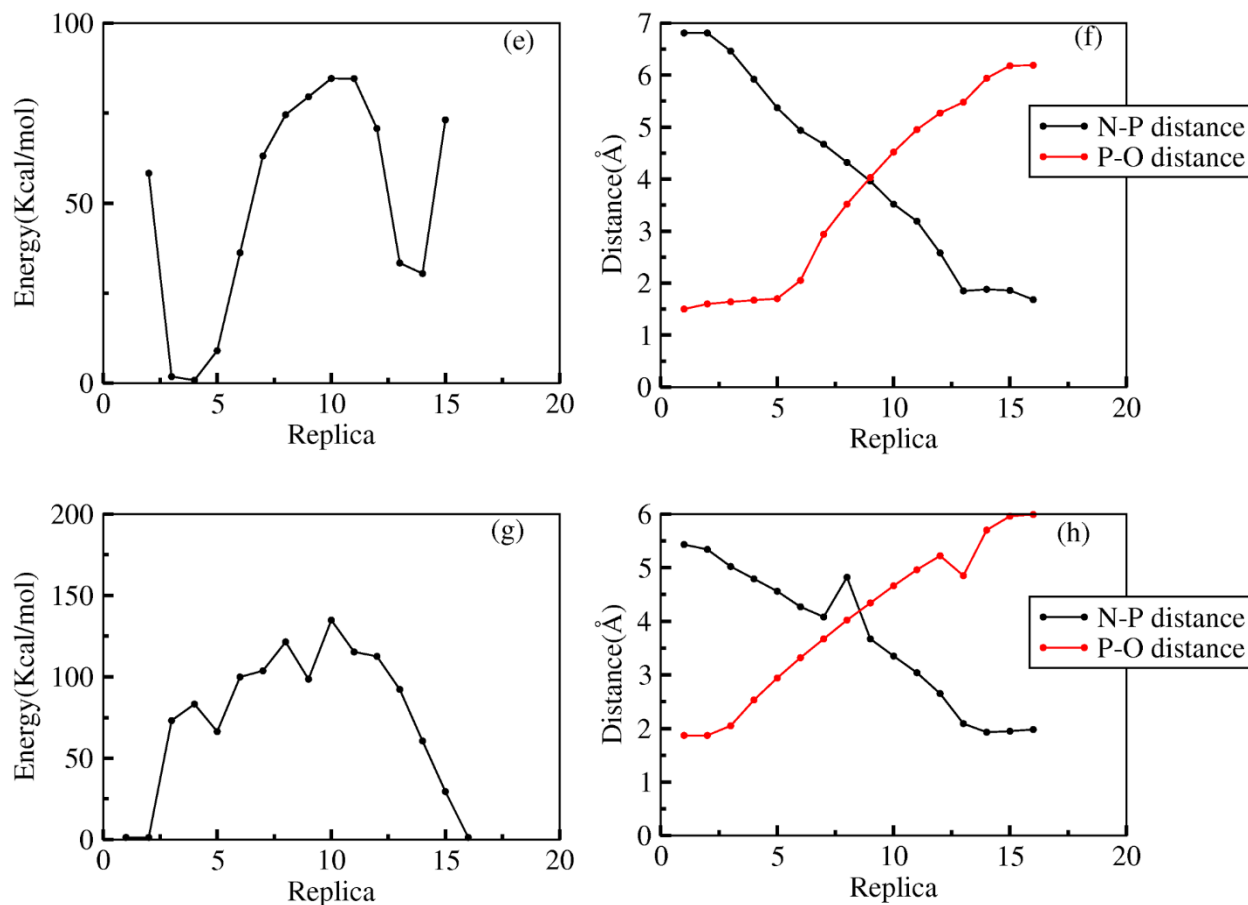


Figure 11: (a) Represents the energy profile for the first optimization. (b) Black color represents the N-P bond distance, red color represents the P-O bond distance, and green color represents the O-H bond distance (proton transfer from water), and blue color represents the O-H bond distance (proton accepted by diphosphate group). (c) Represents the energy profile after interpolating between replica 3 and 12. (d) Red color represents the P-O bond distance and black color represents the N-P bond distance. (e) Represents the energy profile in the absence of magnesium ions. (f) Red color represents the P-O bond distance and black color represents the N-P bond distance. (g) Represents the energy profile in the presence of the environment. (h) Red color represents the P-O bond distance and black color represents the N-P bond distance. All distances are plotted from reactant (replica 1) to product (replica 16) along the profile.

Since proton transfer is a fast process, we observed that the proton transfer happened between replica 1 and replica 2. Hence, replica 3 contains proton attached to the pyrophosphate group which was stabilized that particular replica (Fig. 11 (a)). This

stabilization process made the replica to be more stable as compare to the reactant one. In a similar way the replica 12 was observed to be more stable as compare to the product one (replica 16).

Hence, we recreated the intermediates between these two replicas using the linear interpolation technique. Finally, similar NEB calculations were performed using 50 steps of steepest descent algorithm. The final energy profile is shown in fig. 11(c). The energy barrier for this optimization was observed to be 26.3 Kcal/mol. Since proton was already transfer in the reactant and the product state, the mechanism only involves attack of active lysine group at the alpha phosphate group. N-P and P-O distances were plotted against the profile which are the bond forming and the bond breaking distances respectively. The distance plot is shown in fig. 11(d).

Replica 10 was observed to be the highest energy intermediate whose structure is shown in fig. 12.

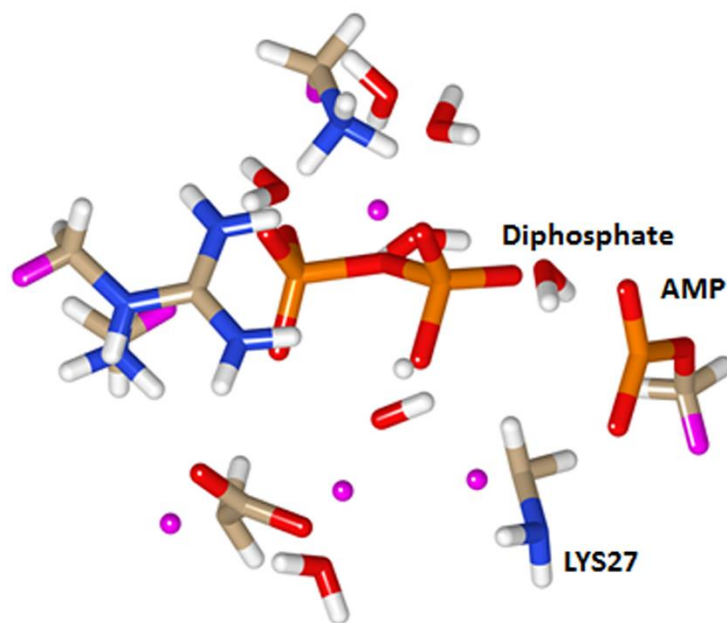


Figure 12: The highest energy intermediate structure, i.e. the probable transition state.

Effect of the magnesium ions:

As we already mentioned that a significant role of the divalent ions is observed in the previous studies.⁸ To verify the role of the magnesium ions, we removed the ions from the reactant state and optimized it in the gas phase. Further, we created the product structure from the optimized reactant structure. The intermediate structures were created using the linear interpolation method. Similar NEB calculations were carried out to optimize the reaction again.

The final energy profile and distances plots are shown in the fig.11 (e) and fig.11 (f), respectively. The energy barrier for the mechanism was 84.2 kcal/mol. This barrier is unusually high and we are currently investigating for the probable cause of its origin.

QM/MM profile:

Further, to see the effect of the environment we have taken the classically optimized reactant and product structures. Replicas were created between these two metastable states. We then replaced the QM region coordinates for all the replicas to the optimized gas phase replicas which we got from the NEB calculations. Further, we performed QM/MM NEB calculations using 50 steps of steepest descent algorithm. The QM part was optimized quantum mechanically while the MM part was optimized using AMBER force field.

The final energy profile and distance plots are shown in the fig. 11(g) and fig. 11(h) respectively. The energy barrier for the mechanism was turned out to be 119.3 Kcal/mol. We are still understanding the reason behind the unusual increase in the barrier height.

4.2.2. QM Model 2: Approaching of the active site lysine group followed by proton transfer:

The scheme is shown in the fig. 13.

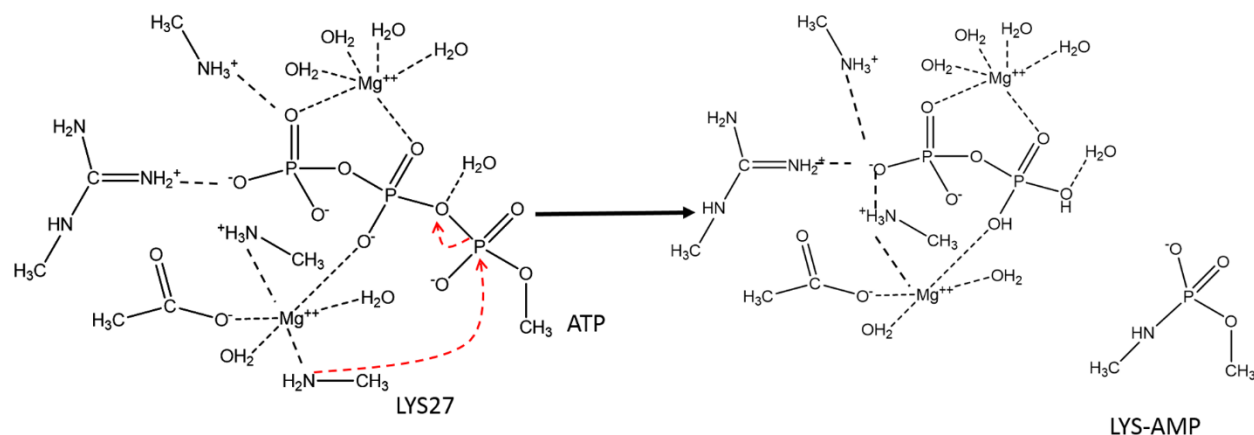


Figure 13: QM Model 2. The reactant (left) and product (right) structure of the second mechanism.

In this mechanism, the reactant state was stable as compared to the product state by 118.7 Kcal/mol. The mechanism proceeds with the approach of the active site lysine group towards the alpha phosphate group without transferring its proton. Once the lysine group was in the close vicinity of the ATP group, the proton was transferred from the lysine to the pyrophosphate group (bond breaking distance N-H distance and bond forming distance O-H) and simultaneously bond breaking occurred between the pyrophosphate and the α phosphate group (P-O distance). At the same time bond formation happened between the active site lysine group and the α phosphate group (N-P distance). The energy barrier was found to be 183.8 Kcal/mol for the profile. The final energy profile and distances plot are shown in the Fig. 14. The replica 11 was observed to be the highest energy containing intermediate whose structure is shown in fig. 15. The N-P and P-O distances are 2.1 Å and 2.25 Å respectively for the replica.

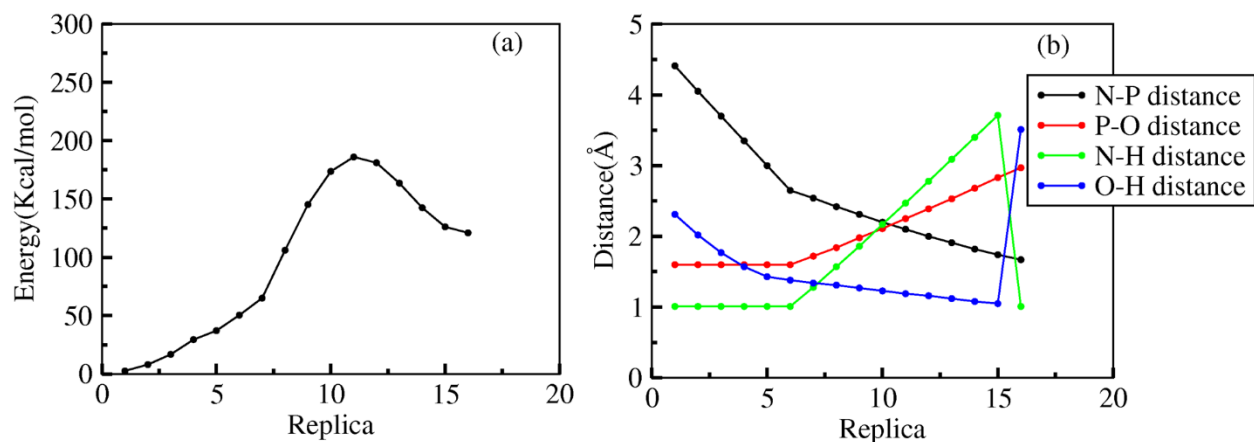


Figure 14: (a) Represents the energy profile for the first optimization. (b) Black color represents the N-P bond distance, red color represents the P-O bond distance, and green color represents the N-H bond distance, and blue color represents the O-H bond distance. All distances are plotted from reactant (replica 1) to product (replica 16) along the profile.

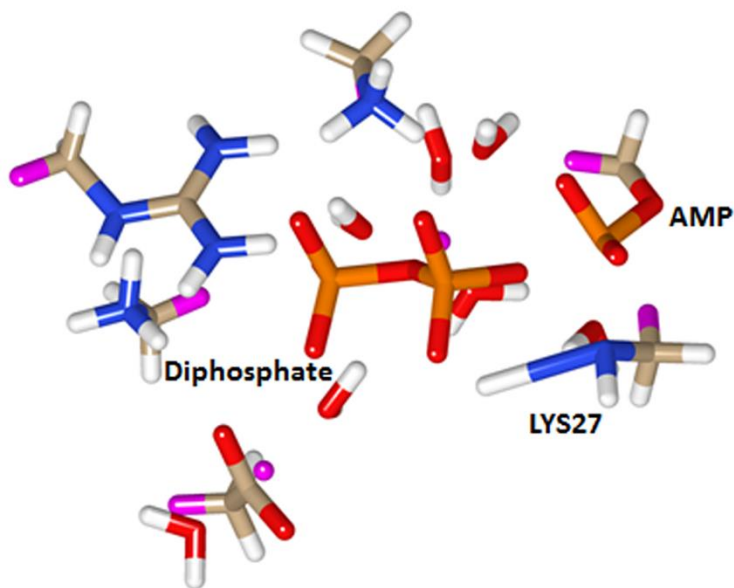


Figure 15: The highest energy intermediate (transition state).

4.2.3. QM Model 3: Proton transfer from the active lysine group:

The scheme for the approach is shown in the Fig. 16.

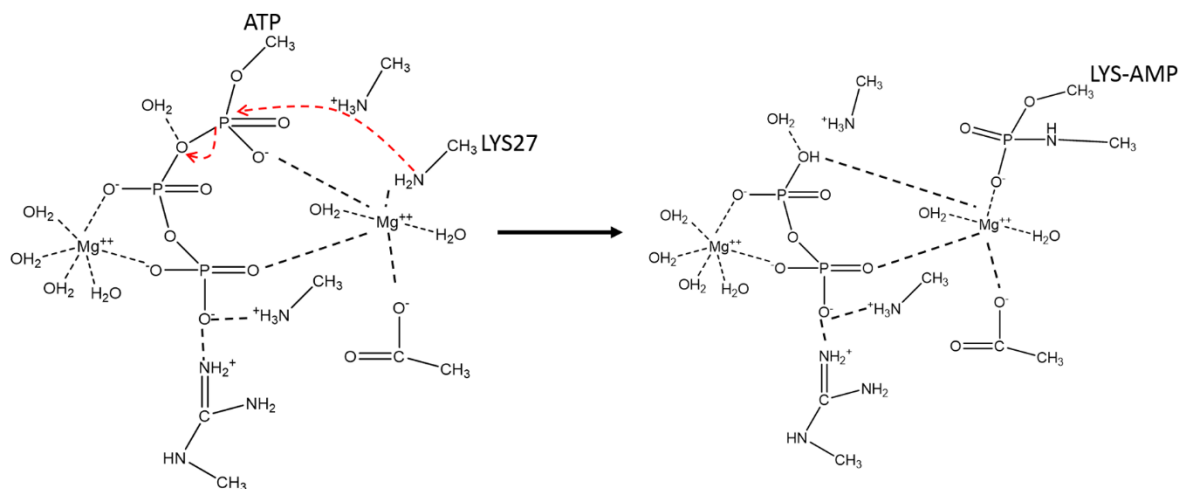


Figure 16: QM Model 3. The reactant (left) and product (right) structure of the third mechanism.

The reactant structure was found to be 33.9 Kcal/mol more stable than the product structure. The mechanism proceeds with the bond breaking between the α phosphate and the pyrophosphate group of the ATP (bond breaking distance P-O). Then proton was transferred initially from the active lysine group to the α phosphate group which was finally transferred to the pyrophosphate group (bond breaking distance N-H and bond forming distance O-H). Meanwhile, the bond formation occurred between the lysine group and the α phosphate of the ATP (bond forming distance N-P). The energy barrier was found to be 77.7 Kcal/mol. The final energy profile and the distances plot are shown in the fig. 17. The replica 9 was observed to be the highest energy state whose structure is shown in the Fig. 18. The N-P and P-O distances are 1.68 Å and 3.14 Å respectively for the replica.

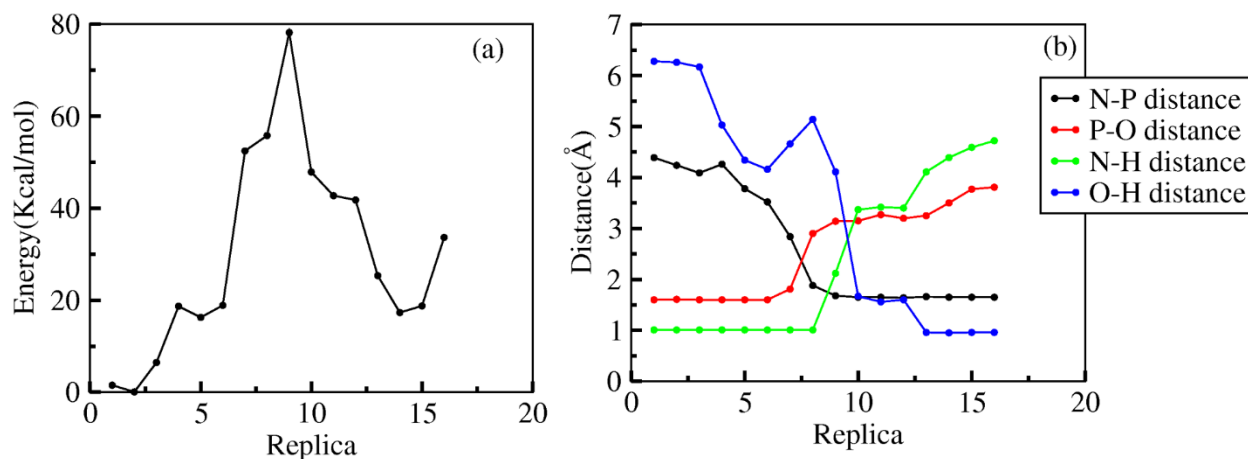


Figure 17: (a) represents the energy plot. (b) Black color represents the N-P bond distance, red color represents the P-O bond distance, green color represents the N-H bond distance, and blue color represents the O-H bond distance. All distances are plotted from reactant (replica 1) to product (replica 16) along the profile.

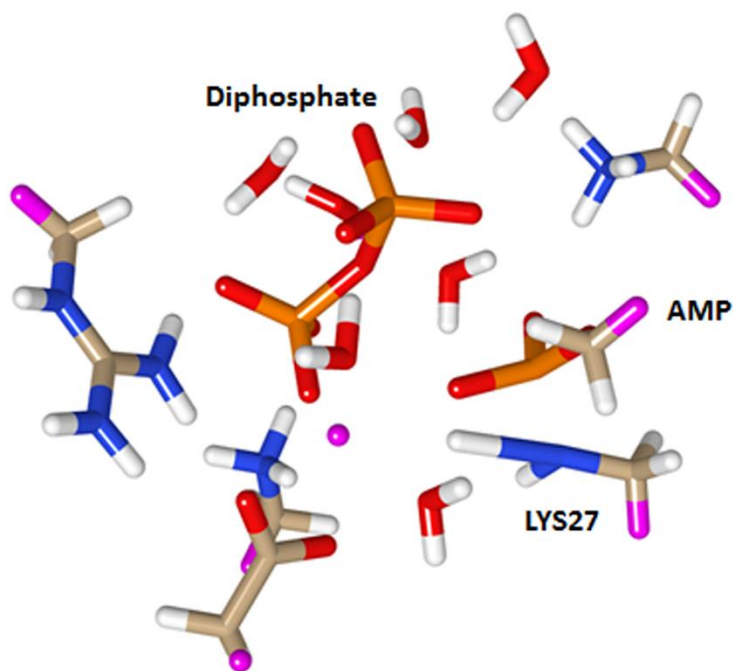


Figure 18: The highest energy replica structure.

4.2.4. QM Model 4: A concerted mechanism

The scheme is shown in the Fig. 19.

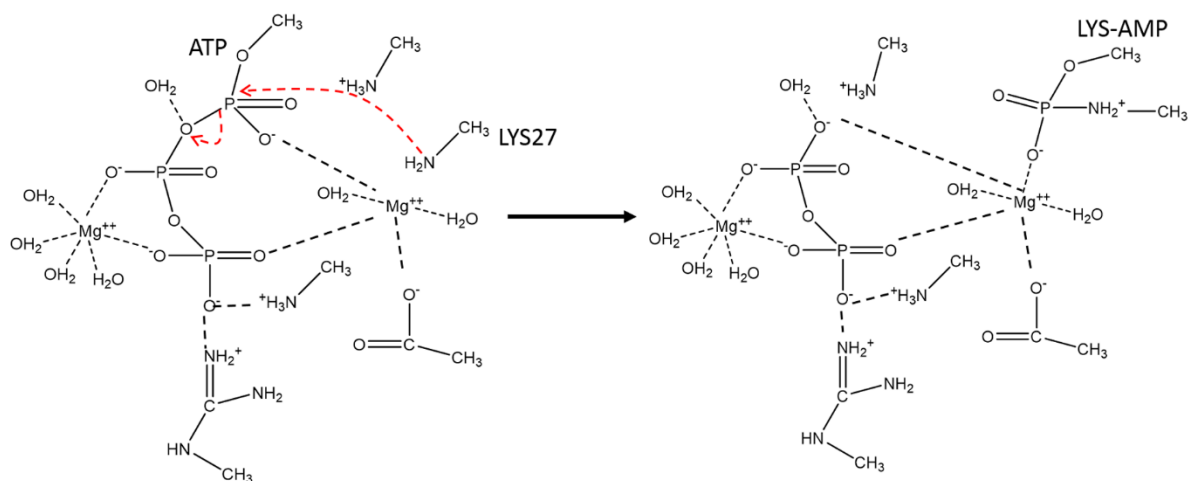


Figure 19: QM Model 4. The reactant (left) and product (right) structure of the fourth mechanism.

The reactant state was 40.1 Kcal/mol more stable as compare to the product state. In this mechanism the active site lysine group is attacking directly to the alpha phosphate group without any proton transfer. Once the lysine group comes in the vicinity of the ATP group (bond forming distance N-P), the pyrophosphate group starts moving away from the alpha phosphate group (bond breaking distance P-O). The energy barrier for the reaction mechanism is 57.84 Kcal/mol. The final energy profile and the distances plot are shown in the Fig.20. The replica 12th was observed to be the highest energy intermediate in the profile which is shown in the Fig. 21. The P-O and N-P distances are 2.21 Å and 2.49 Å respectively for the replica.

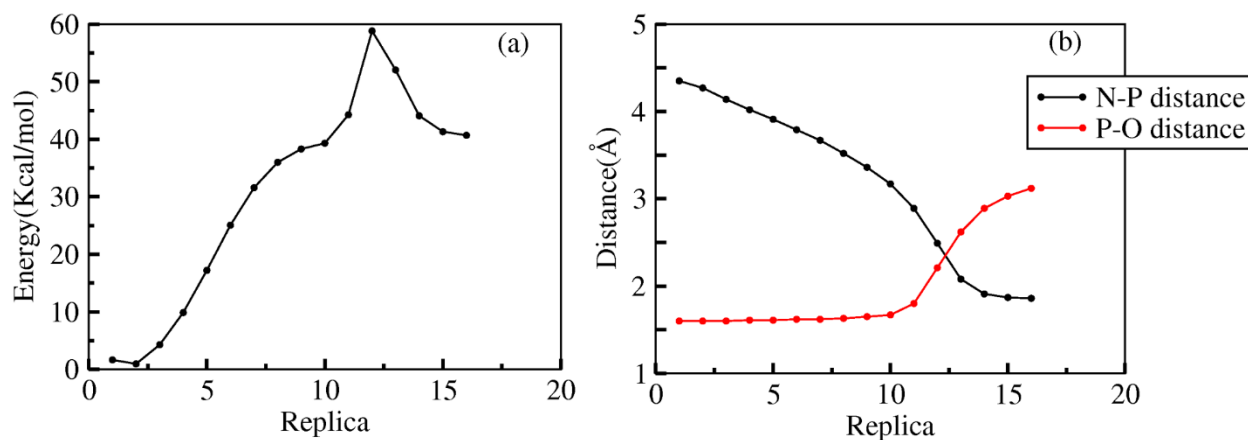


Figure 20: (a) Represents energy profile for the mechanism. (b) Red color represents the P-O bond distance and black color represents the N-P bond distance. All distances are plotted from reactant (replica 1) to product (replica 16) along the profile.

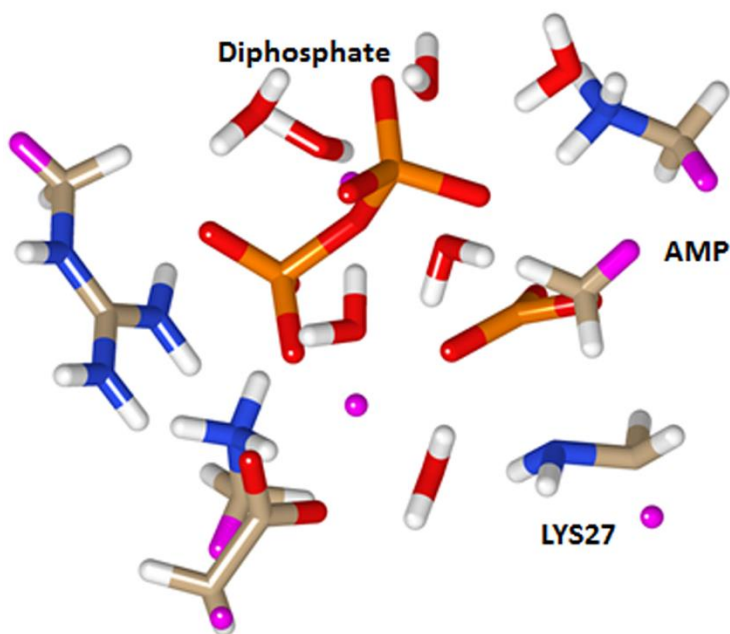


Figure 21: The highest energy intermediate structure.

4.2.5. QM Model 5: A constrained mechanism:

The approach is shown the Fig.22.

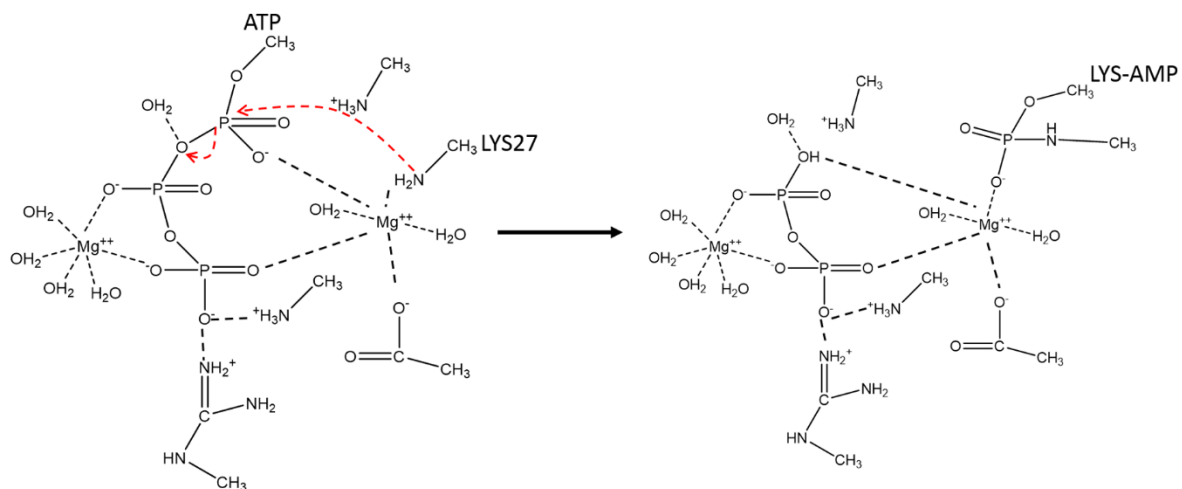


Figure 22: QM Model 5. The reactant (left) and product (right) structure of the fifth mechanism.

In this mechanism, we kept the positions of the transferring proton fix during the optimization in all the replicas. The reactant state was found to be 13.7 Kcal/mol more stable than the product state. The mechanism initiates with the approaching of the active site lysine group towards the alpha phosphate group without any proton transfer. Further, the proton was being transferred initially to the alpha phosphate group and finally to the pyrophosphate group. At the same time bond breaking between the pyrophosphate and the alpha phosphate group of the ATP, and the bond formation between the lysine and the alpha phosphate group occurred. The energy barrier was observed to be 133.8 Kcal/mol. The final energy profile and the distance plot are shown in the Fig. 23. The replica 10 was observed to be the highest energy intermediate in the profile whose structure is shown in the Fig. 24.

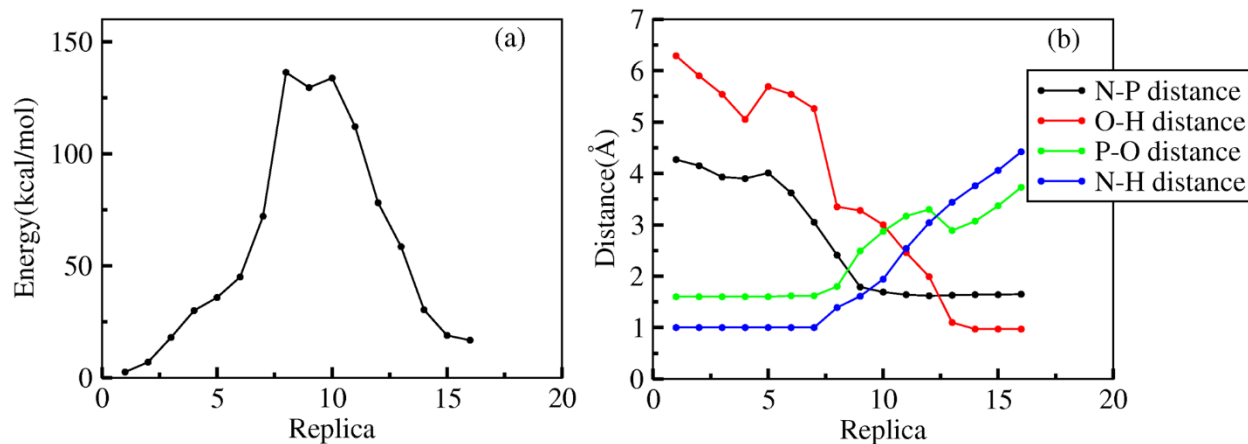


Figure 23: (a) Represents energy plot along the profile.(b)Black color represents the N-P bond distance, red color represents the O-H bond distance. Green color represents the P-O bond distance and blue color represents the N-H bond distance. All distances are plotted from reactant (replica 1) to product (replica 16) along the profile.

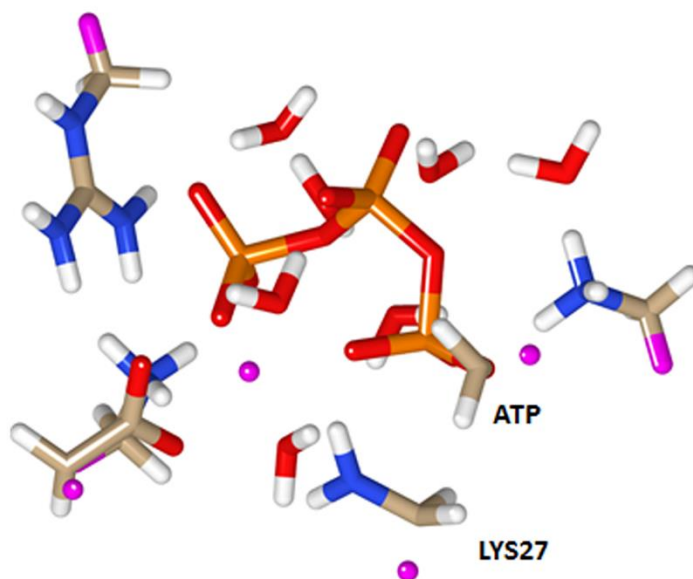


Figure 24: The highest energy structure.

5. Discussion:

We observed that the lowest energy barrier is coming from the 1st mechanism (QM1) while the 2nd mechanism (QM2) provides the highest energy barrier. Since the 1st mechanism (QM1) is driven through proton transfer from the water to the pyrophosphate group followed by the bond formation between active site lysine group to the α phosphate of the ATP whereas, the 2nd mechanism (QM2) proceeds with the approaching of the lysine group to the ATP group followed by the attack of the lysine group at the α phosphate group, we infer that the actual reaction mechanism proceeds with the proton transfer first followed by the bond formation process. However, whether the proton is transferred from water or from lysine is still not properly resolved.

It is believed that lysine attacks the alpha phosphate group without transferring the proton. Hence, to address this we compared the 1st mechanism to the 4th mechanism (QM4). We observe that the 4th mechanism contains an extra activation energy of 31.58 Kcal/mol compare to the 1st mechanism. This comparison clearly shows that during the mechanism a proton must transfer which is proving out assumption.

We have also looked at the effect of magnesium and the environment on the first QM model. The removal of magnesium ions was increasing the energy barrier by 57.95 Kcal/mol. This observation indicates that magnesium ions are playing a significant role through decreasing the energy barrier in the mechanism.

We presumed that the environment is going to decrease the barrier height compared to the gas phase energy barrier. But, we observed that energy barrier increased by 93.07 Kcal/mol. To find out the reason behind the energy profile, we compared the QM energies for the gas phase and for the protein environment. The stabilization energy is defined as the difference of QM energies of the protein environment from the gas phase. Further, we plotted the stabilization energy along the profile which is shown in the fig. 25. It is clear from the plot that environment is stabilizing to the individual replica, but the extent of stabilization varies from one replica to another. It is observed that the maximum stabilization occurs to the reactant and the product structure while the least stabilization observes for the transition state. This was the reason for the increase in the

energy barrier from the gas phase energy profile. We are still working to find out the reason behind the variation in the extent of stabilization.

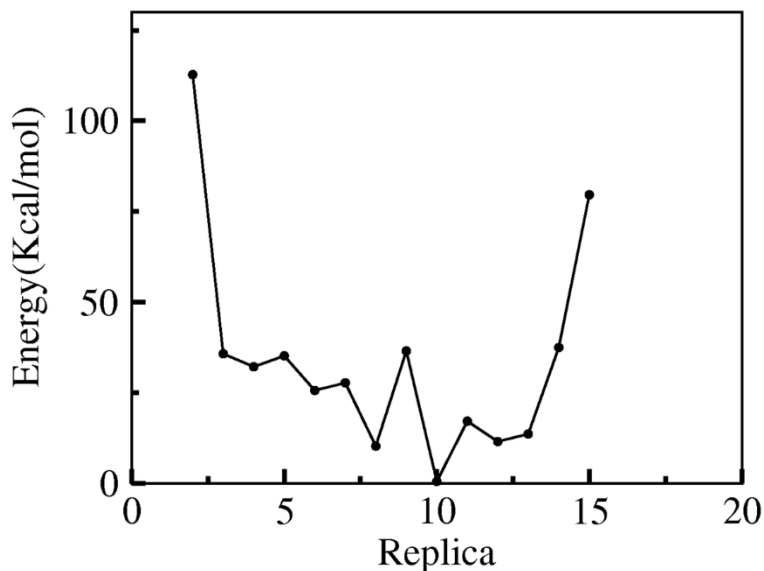


Figure 25: Represents the stabilization energy along the profile.

6. Conclusion:

In this study, we have modeled the chlorella virus DNA ligase enzyme and studied five different types of mechanisms for the first step of the ligation process of DNA. Among all the five steps we observed that mechanism 1 (QM1) provides the lowest energy barrier of 26.3 Kcal/mol. We also tried to look at the effect of magnesium ions and the environment for the same profile. We observed that magnesium ions stabilize the profile by 58.0 Kcal/mol. We also observed the environment stabilizes individual replica in a different manner. In future, we are going to resolve this variation in the stabilization energy from the environment. We kept this profile to be a reference profile.

We also tried four other mechanisms and compared the profile with the reference one. Mechanism 2 (QM2) starts with the approach of the lysine group to the ATP followed by the formation of the bond between these two residues. The energy barrier for the

mechanism is 157.6 Kcal/mol more than the reference barrier. In the 3rd mechanism, the proton transfer happens from the active site lysine group to the pyrophosphate group and the attack of the lysine group at the alpha phosphate of the ATP occurs simultaneously. We observed in this profile also the activation energy is 57.5 Kcal/mol higher than the reference one.

In a similar way we compared the activation energies for the mechanism 4 and 5 to the reference profile which turned out to be 31.6 Kcal/mol and 107.5 Kcal/mol, respectively. Through these comparisons of the activation energies we can conclude that the reference profile (QM1) is the most suitable one amongst all five profiles.

7. References:

1. Shuman.S., *J.Biol.Chem.* 2009, 284, 17365-17369.
2. Yao, N.; Turner, J.; Kelman, Z.; Stukenberg, P. T.; Dean, F.; Shechter, D.; Pan, Z.-Q.; Hurwitz, J.; O'Donnell, M., Clamp loading, unloading and intrinsic stability of the PCNA, β and gp45 sliding clamps of human, *E. coli* and T4 replicases. *Genes Cells* **1996**, 1, 101-113.
3. Barnes, D. E.; Lindahl, T., Repair and genetic consequences of endogenous DNA base damage in mammalian cells. *Annu. Rev. Genet.* **2004**, 38, 445-476.
4. Pursell, Z. F.; Kunkel, T. A., DNA Polymerase ϵ : A Polymerase of Unusual Size (and Complexity). In *Prog. Nucleic Acid Res. Mol. Biol.*, Conn, P. M., Ed. Academic Press: 2008; Vol. Volume 82, pp 101-145.
5. Ellenberger, T.; Tomkinson, A. E., Eukaryotic DNA Ligases: Structural and Functional Insights. *Annu. Rev. Biochem.* **2008**, 77, 313-338.
6. Fabrega, C.; Shen, V.; Shuman, S.; Lima, C. D., Structure of an mRNA Capping Enzyme Bound to the Phosphorylated Carboxy-Terminal Domain of RNA Polymerase II. *Mol. Cell* 11, 1549-1561.
7. Nair, P. A.; Nandakumar, J.; Smith, P.; Odell, M.; Lima, C. D.; Shuman, S., Structural basis for nick recognition by a minimal pluripotent DNA ligase. *Nat Struct Mol Biol* **2007**, 14, 770-778.
8. Zhu, H.; Shuman, S., Bacterial Nonhomologous End Joining Ligases Preferentially Seal Breaks with a 3'-OH Monoribonucleotide. *J. Biol. Chem.* **2008**, 283, 8331-8339.
9. Subramanya, H. S.; Doherty, A. J.; Ashford, S. R.; Wigley, D. B., Crystal Structure of an ATP-Dependent DNA Ligase from Bacteriophage T7. *Cell* **1996**, 85, 607-615.
10. Sriskanda, V.; Shuman, S., Mutational analysis of Chlorella virus DNA ligase: catalytic roles of domain I and motif VI. *Nucleic Acids Res* **1998**, 26, 4618-4625.
11. Odell, M.; Sriskanda, V.; Shuman, S.; Nikolov, D. B., Crystal Structure of Eukaryotic DNA Ligase—Adenylate Illuminates the Mechanism of Nick Sensing and Strand Joining. *Mol. Cell* **2000**, 6, 1183-1193.

12. Sriskanda, V.; Shuman, S., Role of Nucleotidyl Transferase Motif V in Strand Joining by Chlorella Virus DNA Ligase. *J. Biol. Chem.* **2002**, *277*, 9661-9667.
13. Šali, A.; Blundell, T. L., Comparative Protein Modelling by Satisfaction of Spatial Restraints. *J. Mol. Biol.* **1993**, *234*, 779-815.
14. Fiser, A.; Do, R. K. G.; Šali, A., Modeling of loops in protein structures. *Protein Sci.* **2000**, *9*, 1753-1773.
15. Swift, R. V.; Ong, C. D.; Amaro, R. E., Magnesium-Induced Nucleophile Activation in the Guanylyltransferase mRNA Capping Enzyme. *Biochemistry* **2012**, *51*, 10236-10243.
16. Brooks, B. R.; Brooks, C. L.; Mackerell, A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caffisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M., CHARMM: The Biomolecular Simulation Program. *J. Comput. Chem.* **2009**, *30*, 1545-1614.
17. Petrova, S. S.; Solov'ev, A. D., The Origin of the Method of Steepest Descent. *Historia Mathematica* **1997**, *24*, 361-375.
18. Zhang, X.; Bruice, T. C., Reaction mechanism of guanidinoacetate methyltransferase, concerted or step-wise. *Proc. Natl. Acad. Sci.* **2006**, *103*, 16141-16146.
19. Nudged-elastic band method with two climbing images: Finding transition states in complex energy landscapes. *J. Chem. Phys.* **2015**, *142*, 024106.