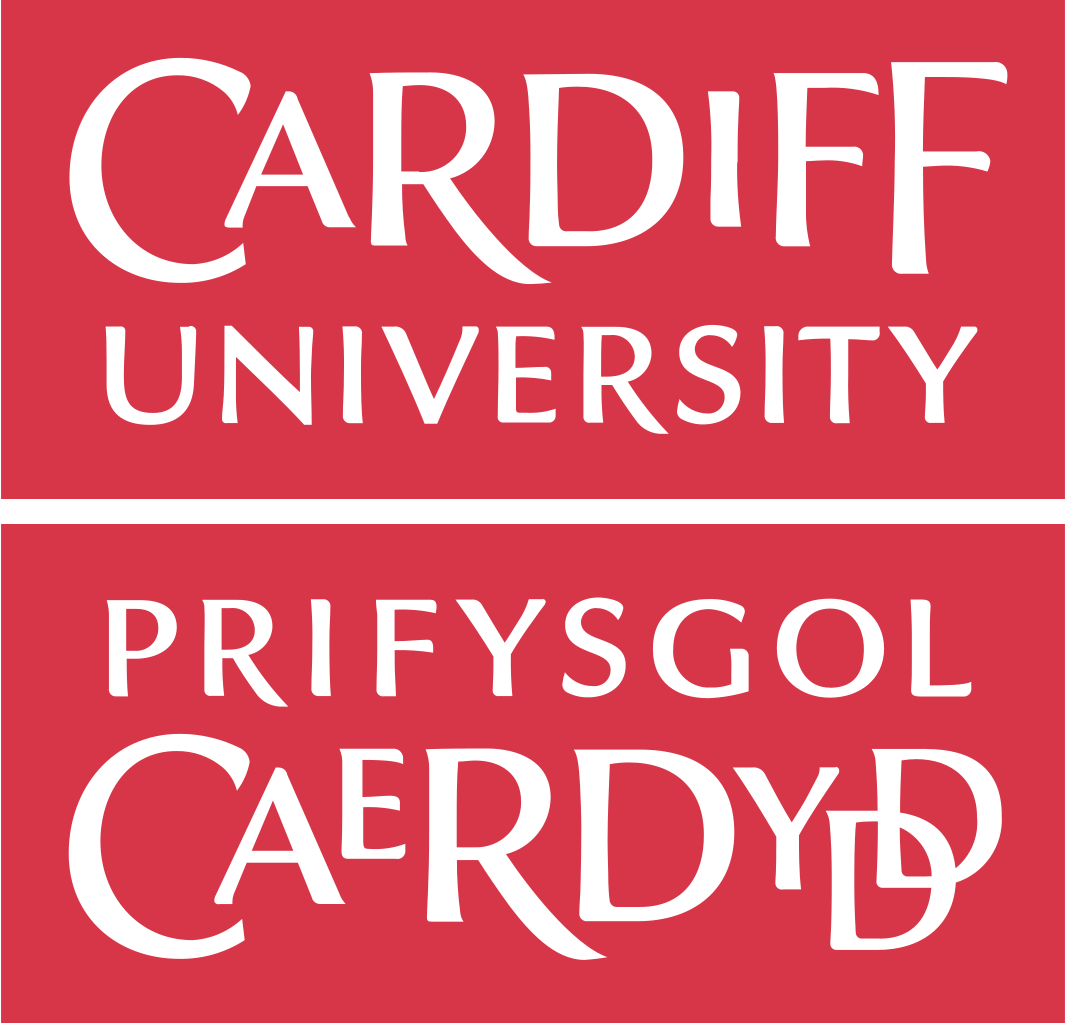
**Investigating ATP ADP & AMP binding and conformational changes in the enzyme Adenylate Kinase using Molecular Dynamics**



Cardiff University [School of Physics and Astronomy](https://www.cardiff.ac.uk/physics-astronomy)

**Author:** Pablo Del Olmo Mier-C21009159

**PXT999**: MSc Dissertation

**Supervisor**: Dr Emyr Macdonald

**Programme:** MSc Data Intensive Physics

**Date of Submission**: 3rd of September 2021

Table of Contents

[1 Introduction 1](#_Toc81158553)

[2 Précis of literature 2](#_Toc81158554)

[2.1 Contextualization 2](#_Toc81158555)

[2.2 Molecular Dynamics 2](#_Toc81158556)

[2.2.1 CHARMM Force Field 2](#_Toc81158557)

[2.2.2 GROMACS 3](#_Toc81158558)

[2.3 Mg2+ Role in the catalytic process 3](#_Toc81158559)

[2.4 Conformational changes in Adenylate Kinase reactions 3](#_Toc81158560)

[2.5 Free Energy Landscape in Adenylate Kinase reactions 4](#_Toc81158561)

[3 Report 5](#_Toc81158562)

[3.1 Aims & Objectives 5](#_Toc81158563)

[3.2 Methodology 5](#_Toc81158564)

[3.2.1 Molecular Dynamics Simulations 5](#_Toc81158565)

[3.2.2 Analysis of the trajectory 7](#_Toc81158566)

[3.3 Results & Discussion 8](#_Toc81158567)

[3.3.1 Study of system configuration 9](#_Toc81158568)

[3.3.2 Ligand free Adenylate Kinase simulation 11](#_Toc81158569)

[3.3.3 Ligand-binding Adenylate Kinase simulations 12](#_Toc81158570)

[3.3.4 Ligand-binding Adenylate Kinase Simulations with Mg2+ ions 14](#_Toc81158571)

[3.3.5 Mapping the transitions and pathways of Adenylate Kinase 16](#_Toc81158572)

[3.4 Conclusions 18](#_Toc81158573)

[3.5 Suggestions for future work 18](#_Toc81158574)

[4 Critique of project 19](#_Toc81158575)

[4.1 Evaluation of the initial plan and the actual plan 19](#_Toc81158576)

[4.2 Scientific merits 20](#_Toc81158577)

[4.3 Reflections on the dissertation process 21](#_Toc81158578)

[References 22](#_Toc81158579)

[Appendix 25](#_Toc81158580)

[A GROMACS, MPI set up and VMD tutorial 25](#_Toc81158581)

[B Compilation of GROMACS abbreviations 27](#_Toc81158582)

[C Initial Files and generated files through the simulation 28](#_Toc81158583)

[D ATP, ADP and AMP atom list used for generating topology files 29](#_Toc81158584)

[E Code & Commands tutorial to facilitate methodology replication 30](#_Toc81158585)

[F Batch files and Python scripts 33](#_Toc81158586)

**ACKNOWLEDGEMENTS**

Thanks a lot, to my supervisor, Dr Emyr MacDonald, for guiding me thought the difficult moments as well as Dr Georgine Mendez for providing extremely useful insights during the whole dissertation. I would also like to thank Dr Richard Lewis and Dr Paul Roche for preparing me for the marvellous world of scientific research. Finally, a big thanks to my parents, María and Jose, because without them I am nothing. The days are long, but the years are short and what I want most in this life is to be with them.

# Introduction

Kinases are enzymes that complete phosphotransferase reactions in a few microseconds; without these enzymes, these reactions would require about 7500 years under normal conditions. (Kerns et al,2015). It is important to fully define the dynamics of the enzyme to determine how the biological mechanism work. (Schroeder GK et al, 2006). Adenylate Kinase, referred to as “AK”, is quite different from many of the protein kinases because is only active in nucleotide substrates and usually catalyses reversible reactions. The most important reaction of this type is the one that maintains the balance of the concentrations of adenosine triphosphate (ATP), which is the main fuel for chemical reactions, and the corresponding diphosphate (ADP) and monophosphate (AMP).

Although multiple investigations have been carried out, there is a gap in our knowledge of the energetic landscapes of kinases and the large-scale conformational changes that describe their catalytic activity. (Li et al, 2015). In this dissertation, we assessed this problem by exploring and understanding the role of AK, a key kinase on cellular energy homeostasis, in regulating and catalysing the interchange of ATP, ADP and AMP molecules. Molecular Dynamics simulations were performed using GROMACS a versatile C++ based package that allowed determining at the molecular scale the midway positions of AK. Necessary computational power was provided by executing the simulations with the ARCCA and Hawk supercomputer cluster. Long time-scale simulations (up to microseconds) were executed for an initial un-bound open AK conformation. In the ligand-bound AK, simulations were executed with and without Mg2+ ions attached on ATP, ADP and AMP for initial open and close conformations.

We found that all domains of AK operate in a clearly chronological manner. In particular, in the un-bound AK case, there are multiple intermediate states separating the open and closed states, which makes quick transitions of AK possible by lowering the energy barrier. In the ligand-bound AK, the large energy barrier favoured the structure, and the enzyme prefers to turn into the closed states in the form of coupled transitions. We also find that Mg2+ ions accelerate the overall reactions in all simulations performed.

Results highlight the fundamental inherent kinetics of AK as well as the complex conformational changes that AK uses to carry out its enzymatic activities. The developed methodology has been compiled in an easy to replicate tutorial that can also be used with other biomolecules and systems of proteins.

In terms of potential significance, the primary application of this study is in the context of biomedicine, and primarily in pharmaceutical development, where molecular dynamics is employed. (Durrant and McCammon, 2011). AK activity can play an important role in cancer cells transformation because the enzyme induces reprogramming of energy metabolism that is required to boost the metabolic pathways necessary for rapid tumour growth and division. (Ji et al, 2017). AK energy calculations to be conducted in this study could enable additional investigations.

This project will be structured in three main sections: Section 2 is the précis of the literature, and it includes the related background knowledge, essential to understand the big picture of the project. Section 3 describes the simulation methodologies that will be implemented throughout the project. It also includes the results obtained and discuss the in vivo action of AK in AMP, ATP, and ADP. Section 4 is the project critique, in which a critical evaluation of the project in terms of planning, reflections and scientific merits is carried out.

Finally, this dissertation also includes an Appendix section. It provides a step-by-step methodology tutorial with full tables of code, software instructions, atoms list, GROMACS commands and VMD images together with further explanations that aims to facilitate other researchers the replication of this project.

# Précis of literature

## Contextualization

Living organisms engage in three primary activities that requires a continuous supply of energy. These three activities include performance of mechanical work such as muscle contractions, transportation of substances in the body and the synthesis of macromolecules (Rosing and Slater,1972). For these processes, the main source of energy is ATP, which is made up of a ribose, an adenine, and a trisphosphate. In standard conditions, a high number of Gibbs free energy () is released on hydrolysis of ATP to ADP and orthophosphate (). Even more Gibbs free energy is released ( when ATP is converted into AMP and pyrophosphate (:

(1)

Where is the standard reaction free Gibbs energy, is the temperature,is the Boltzmann constant and the constant quantities in square brackets are molar concentrations. Phosphate esters () and () are high energy linkages that esters have advantages such as extreme resistant to nucleophilic attack which provides genomic stability, storage of biochemical energy, and long signalling lifetimes (Kamerlin SC et al,2011).

Adenylate kinase (AK) is a critical phosphoryl-transfer enzyme identified in all living cells. Main characteristics of AK include its extended intracellular lifespan and its amino acid sequence entailing 200 residues in terms of length. Primarily, AK aids in the catalysing of magnesium ions (Mg2+) during the process of interconversion between ATP and AMP molecules (Bommer, Van Schaftingen and Veiga-da-Cunha, 2020):

(2)­­

Various studies have been performed employing Molecular Dynamics simulations (MD) in the evaluation of properties of AK and further study of energy landscapes. (Kerns et al, 2015). MD simulations have allowed for development of a better understanding of P-transfer in AK catalytic process (Lassila JK et al, 2011). However, despite all the studies and information collected and analysed regarding the P-transfer enzymes, there is still a gap in literature on the comprehension of kinase catalysed phosphorylation (Li et al, 2015). This work, therefore, proposes for comprehension of AK’s operation dynamics. Moreover, the comprehension will entail an explanation of the conformational transition mechanism based on time specifically microseconds and explicit MD (LT-MD) simulations. The purpose of this work and detailed explanation of the AK transition pathways, and free energy landscapes is to further the knowledge on the impact of conformational changes of the enzyme’s biological functioning.

## Molecular Dynamics

MD simulates the movement of molecules over time using Newton's second law. Each particle with mass and location (according to the arbitrary origin) in an isolated system experiences a force of , which is described by:

(3)

The important feature of MD simulations is that they give details on the system’s main characteristics, such as the system's stability, conformational exploration, and approximating of molecular forces (Karplus, 2005). However, MD simulations are complicated because they need considering every molecule and atom in the process and are heavily dependent on both processing power and memory.

### CHARMM Force Field

MD simulations are mostly depending on how the interaction forces potential (V) is shaped. For this work we will be using CHARMM (Chemistry at Harvard Macromolecular Mechanics) force field because it is a well-stablished package that considers all the atoms of the system, including the hydrogen ones, and is periodically updated to learn from past simulations errors include errors. In the field each atom has a kinetic energy and a potential energy *V* described in terms of the positions of its neighbours within a force-field.

The workflow of the algorithm that used CHARMM is as follows: First; the positions and velocities of each atom are indexed. Second, the potential energy V of each atom is calculated. Third; the instantaneous force on the atom is calculated as , Forth: accelerations, velocities and positions can be updated from the previous step. Finally, the process is repeated for a succession of small timesteps (which need to be shorter than the atomic vibration period).

### GROMACS

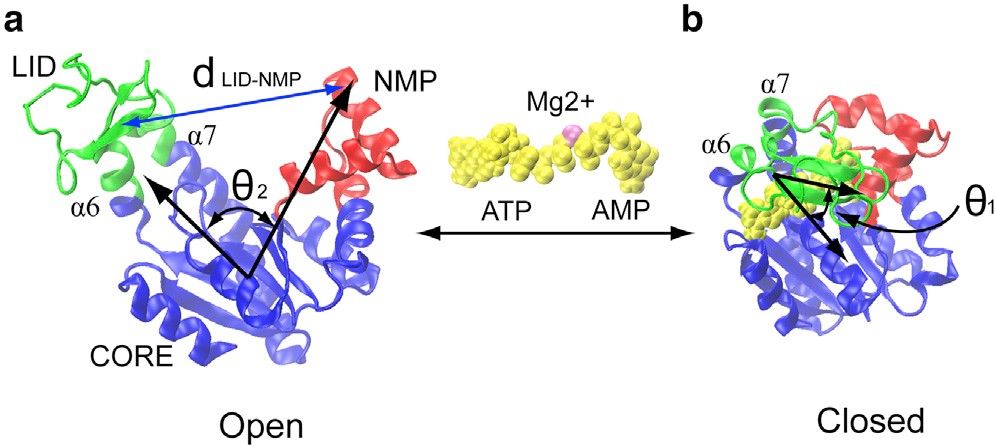
The GROMACS (GROningen MAchine for Chemical Simulations) program is one of the most used systems for molecular dynamics simulations. GROMACS is a flexible software that is meant to run MD simulations on macromolecules such as proteins, lipids, and nucleic acids, amongst other types of molecules .We have selected this program for the dissertation because is generally five times faster than existing MD software and it also allow scripts to be executed in conjunction with supercomputer clusters such the one it will be used in this project in order to save computational power and accelerate the simulations. Another reasons to select this software is because there plenty of documentation available related with similar works (Astuti, A. & Mutiara, 2009).For example, meta dynamics simulations using different Force fields such as [MARTINI](https://www.sciencedirect.com/topics/biochemistry-genetics-and-molecular-biology/martini) has given insights about the motion of different proteins (Bekker et al, 1993).

## Mg2+ Role in the catalytic process

Mg2+ in the catalytic process, enables the quick reaction between the molecules, hence reducing the time taken for the reactions to be complete (Tan, Y. W et al, 2009). Additionally, Mg2+ having been associated with different mechanisms has resulted in unsolved disagreements concerning the dynamics strategies in catalysis. (Tan YW, Hanson JA, Yang H,2009). Moreover, these debates have also taken over, focusing on the changes in the enzymatic reactions (Bhabha G et al, 2013) since the presence of the Mg2+ in the chemical reaction is critical for high energy production. A quantitative description of the process of energy production is necessary for the reconciliation of the dissimilar process including the function of Mg2+ ions in the enzyme catalysis.

## Conformational changes in Adenylate Kinase reactions

The AK arrangement has been broadly studied in the scientific community over the years (Wang et al ,2020). AK's structure, as depicted in Figure (1), is composed primarily of three major components: the NMP, the CORE, and the LID.AK is considered to remain in an open conformation when there are no substrates available, and to undergo large conformational changes when AMP ADP or ATP are bound to the NMP and LID domains of this enzyme (Endicott JA et al, 2012).Although the core domain is rather inflexible throughout significant conformational transitions, structural movements are mostly observed in the LID and NMP domains during these transitions (Aviram et al, 2018).

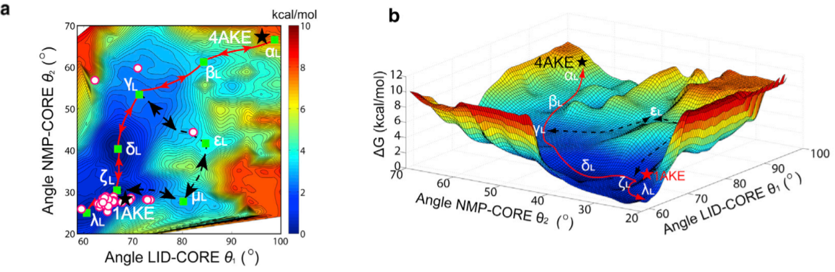


**Figure (1):** Conformational structure in Adenylate Kinase: (*a*) the open state without ligands and (*b*) the closed state with ATP and ADP ligands (yellow spheres). The enzyme consists of three well-defined domains: the rigid domain CORE (*blue*), the ATP binding domain LID (*green*) and the AMP binding domain NMP (*red*). The angles are formed between the centre of mass of the CORE residues and the centre of mass Of LID and NMP residues respectively. *dLID-NMP is* distance by the centre of mass between domains LID and NMP. Source: Li et al,2015

An experiment conducted by Hanson J and Yang H in 2009 showed that the balanced state of AK is best shown in the closed conformation. This finding applies even in the not bound state but following a solution form NMR experiment by Aden and Wolf-Watz ,2007, the attachment of ATP triggers an active equilibrium of both the closed and open conformations. The experiment also showed that the open to closed transitions take place on the NMP and LID domains with the Core domain remaining rigid. Another useful study was done by (H. Willdman et al, 2007) who demonstrated that large-scale domain movements in unbound AK are not random. These past experiments have shown that the substrate-protein exchanges follow a procedure differentiated from the common orthodox theories such as induced-fit theories. These experiments also support and show that conformational selection occurs regardless of ligands. (Koshland, Jr, D.E et al,1994).

## Free Energy Landscape in Adenylate Kinase reactions

It has been possible to determine the free energy landscapes of a few proteins based on their properties and free energy landscapes because it is difficult to relate the dynamic properties and the landscape of free energy activities (Pontiggia et al, 2008). In an Adenylate kinase-mediated reaction, energy transfer is made of numerous steps that each require the release of energy from ATP. Figure (2) quantifies numerous different kinetic states of AK along with the reaction routes as part of energy landscape studied using BE-META dynamics simulations:



**Figure (2).** The energy landscape of Adenylate Kinase with ATP and ADP ligands. a) A two-dimensional contour map. (b) A three-dimensional contour map. Both maps show the angles LID-CORE and NMP-CORE and the free energy, with the scale bar in kcal/mol units. The states that undergo AK are shorted as follow: αL (open conformation), βL, ζL, γL, μL, εL and λL (closed conformation). The red lines indicated the minimum free energy paths. The dashed black lines are the possible alternative pathways. Circular symbols represent additional conformational states. 4AKE and 1AKE correspond to the open and close AK structure respectively. Source: (Li et al,2015).

Plots above reveal that the ligand-bound open state has no apparent minimum, indicating that the ligand does not bind effectively to a completely open conformation of AK. However, there is a definite bottom toward the ligand-bound closed state (Wolf-Watz et al,2010). In addition, several research have proven that the unligated enzyme energy landscape supports conformational population support (Sullivan, S. M., and T. Holyoak. 2008). However, the Adenylate Kinase in a fully closed state cannot be able to capture the ligands because the ligands cannot be able to bind to their active sites. According to (Hammes et al, 2009) conformational selection and the operations on the induced fit are dependent on how concentrated the ligand and protein are. In contrast, high ligand concentration dominated by induced fit will result when the ligand binding is in little stochastic frequency.

# Report

## Aims & Objectives

The **main aim** of the dissertation is therefore to understand and quantified the conformational changes that are crucial to the role of AK in regulating and catalysing the interchange of ATP, ADP, and AMP. This involves setting up a Molecular Dynamics model using GROMACS and running several simulations in conjunction with ARCAA and Hawk supercomputer cluster for different key scenarios to obtain clarifying insights. The achievement of this aim has involved the following six objectives:

1. Develop from scratch a full Molecular Dynamics simulation model of AK catalysing the interchange of ATP, ADP and AMP using GROMACS and provide a detailed tutorial on the methodology to facilitate replication of the research.
2. Obtain readable x-ray diffraction topology file structures for the protein as well as for the nucleotides ADP, AMP, and ATP. Construct different initial conformations states.
3. Simulate the trajectory of open AK structure without any nucleotide for a total time of 4 ns
4. Simulate the trajectory of close AK structure ligated in both binding sites with a) ATP-AMP, b) ADP-ADP, c) AMP-ATP with and without Mg2+ions for a total time of 4 ns
5. Simulate the trajectory of open AK structure (with the CORE aligned) ligated in both binding sites with a) ATP-AMP, b) ADP-ADP, c) AMP-ATP with and without Mg2+ ions for a total time of 4 ns
6. Monitor key variables for the different conformations and obtain the dynamics trajectories showing the transitional pathway from the unbound to the bound conformations.

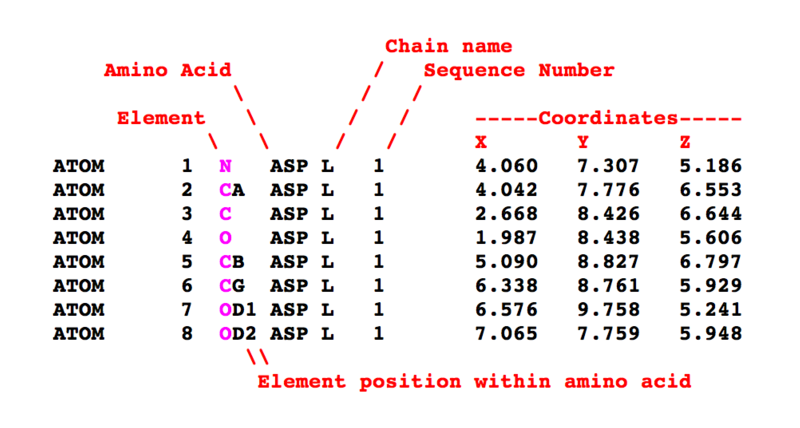
The four fundamental questions of this research are as follows:

1. What are the conformational changes that undergoes AK binding process?
2. How many factors play an important role during the transitional pathway on AK for the overall rate acceleration?
3. How the activation process begins?
4. What is the effect of bringing Mg2+ to the catalysis process?

## Methodology

### Molecular Dynamics Simulations

AK topology files contains position and connectivity information of all the atoms of the system. Topology files has been created from scratch (Appendix D) using as reference X-RAY AK crystallography structures obtained from bacteria E. coli. Reference topology file (extension .pdb) selected is pd4b.ake for an initial open AK conformation and pd1.bake for an initial close AK conformation since these files have been broadly used in previous studies (Berry et al,1994). Topology files has been downloaded from protein base bank (https://www.rcsb.org/.) This project has performed a total of thirteen (13) MD simulations corresponding to different initial AK conformations: one ligand free AK, six ligand combinations bound to AK binding sites with Mg2+ and the same six ligands’ combinations bound to AK but without Mg2+. structures. The ligands bound to AK binding sites NMP and LID are the molecules, ATP, ADP, and AMP. Combinations used were ATP(LID)-ATP(NMP), ADP(LID)-AMP(NMP), AMP(LID)-ADP(NMP). They were attached by overwritten the initial close and open conformations from AK topology files. New atoms coordinates were written using a simple text editor and an Atom map for ATP, ADP, and AMP molecules (Appendix D). Initial topology files contain all the necessary information to perform the simulations: it includes all the molecules, atoms names, charges, connections, coordinates, and masses and are readable by GROMACS.



**Figure (3):** Preview of a PDB file of close AK. The Protein Data Bank (pdb) file format is a text-based file format that describes the three-dimensional structures of molecules in the Protein Data Bank. It shows the different elements that it contains when read as a plain text. We examined the topology files for any incomplete internal sequences and any amino acid residues with missing atoms before running the simulations. The results were satisfactory and free of errors; therefore, no further actions were taken.

All 13 MD simulations were performed in GROMACS and using the same parameter settings (Appendix C). The methodology below summarises the process followed in all the simulations:

1) Deletion of existent water molecules and generation of a new GROMASCS readable topology. This preliminary step is essential for two main reasons. First, GROMACS cannot work with the initial pdb file alone and need to translate the information into a new topology file (extension. top). Second, GROMACS automatically add hydrogens to the molecules when generating the topology so a pdb file without water needs to be used as an input.

2) Unit cell definition and addition of solvent. CHARM 36-protein was the force field used and (TIP3P) was the water model because they are the most stable and accurate available within the GROMAS library. We have centred the protein in a box with cube form and with the appropriate size the reason of creating a box is that periodic boundary conditions are meet. Periodic boundary conditions are necessary because we must adhere to the bare minimum of picture convention. That is, a protein should never be exposed to its periodic image; otherwise, the forces computed would be inaccurate. The process-entailed solvating the system in a cubic 50 × 50 × 50 A˚ TIP3P water box using 5.000 water molecules.

3) Electric charge neutralization of the system. Non-intrusive Na sodium ions were utilized for neutralizing the system box. For electrostatic interactions, the particle mesh Ewald technique was employed. The method further required that the intermittent boundary conditions were attached to the simulation cell.

4) Energy Minimization. This phase tries to produce a structure that is acceptable in terms of shape and solvent orientation. The energy of the system was minimized by the application of 50000 steps as well as the gradual heating of the system to 300K in 40ps. Minimization step size were set to 0.01, a maximum force 1000.0 KJ/ mol.

5) Equilibration of Ions and solvent around the protein. The equilibration phase is required because the solvent system must be brought to the temperature we desire to imitate, and it is during this stage that the solvent and ions are properly oriented around the protein. There are two phases in equilibration: Phase 1: An NVT ensemble is used for NVT equilibration (constant Number of particles, Volume, and Temperature). The length of time required for such an operation varies depending on the contents of the system, but in NVT, the system's temperature should achieve an appropriate value. The second step is NPT equilibration, which is carried out in an NPT ensemble with constant particle numbers, pressure, and temperature. The "isothermal-isobaric" ensemble is often known as the "experimental" ensemble since it most closely mimics experimental settings. For both equilibrations, the restraint force was reduced from 5.39 to 0 kcal. At 300K the production simulations were unrestricted and a relaxation time of 0.002 fs, a 100 ps as total number of steps and pressure of 1 bar using the Perriello-Rahman approach. Using a time step of 2.0fs in the production simulation the LINCS algorithm was used for constrained hydrogen atoms. A set of house scripts were used in converting the equilibrated structures to GROMCAS.

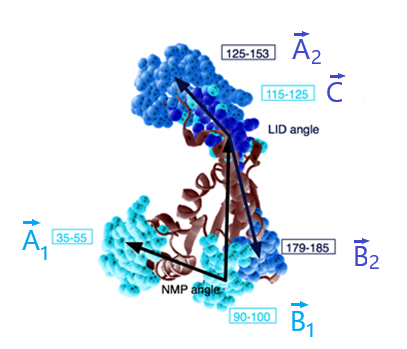
6) Molecular trajectory simulation. This is the last step of MD simulations and is based on data collection of the protein for each time frame. Protein trajectories were pushed to 4 ns. The Visual Molecular Dynamics (VMD) program was used for visualizing the protein trajectories and key conformational states.

### Analysis of the trajectory

Once the simulations have been carried out, the result is a file containing, among others, the information of the spatial coordinates of each atom of the system as a function of time. Several variables can be derived from the position of the atoms that allow the results to be analysed. In this work, five variables were used for monitoring the conformational transitions and analysing the trajectories:

* + - * + The angle between the centre of mass of the carbon atoms and NMP residues. This is usually defined as the NMP-CORE angle or NMP angle (Tikhonova et al,2013).
        + The angle between the centre of mass of the carbon atoms and LID residues. This is usually defined as the LIC-CORE angle or LID angle angle.
        + Distance between the centre of mass NMP and LID domains .

For this first three variables simple vector relations have been used from the three-dimensional coordinates of certain centre of mass of residues. The following image visualizes the AK domains together with the defined vectors:



**Figure (4):** Adenylate Kinase domains indicating , represent the three-dimensional spatial coordinates of the centre of mass of the NMP residues 35-55 and 90-100 respectively. , represent the three-dimensional spatial coordinates of the centre of mass of the LID residues 125-153 and 179-185 respectively. represent the 3-dimensional spatial coordinates of the centre of mass of the CORE residues (115-125). Image was created using VMD.

From the image above the following vectors can be defined:

(5)

Using basic vector analysis, we can obtain the angles and using the following relation:

(6)

On the other hand, the distance between the centre of mass NMP and LID domains, is simply obtain by taken the magnitude of the vector that is obtained by taken together the centre of mass of the NMP residues from , and the LID residues from , as shown below:

(7)

Thus, the distance is given by:

(8)

According with previous studies. (Beckstein et al, 2009 and Müller et al,1996) for the topology files selected in this work, pd4b.ake and pd1b.ake, the critical angles and distances that correspond to the open (superscript “o”) and close (superscript “c”) conformation are:

In this work we will analyze the possible combinations to distinguish between close- semi-open- open transitions.

* + - * + Fourth variable is the root-mean-square deviation (RMSD) of the carbon atoms in terms of the closed and open configuration. RMSD is a useful metric to assess the protein structure stability. The lower RMSD value the better target structure

(9)

* Last variable is the radius of Gyration . A protein's compactness is determined by its gyration radius. If a protein is folded correctly, the value of will most likely remain stable. will vary over time as a protein unfolds. It is given by:

­

(10)

where and are the mass and the position of each individual atoms and the system respectively. On the other hand, and are the total mass and the centre of mass of the protein as a whole system.

## Results & Discussion

The following sections present the results obtained by strictly following the methodological process described in section 3.2. This section is mainly focused on fulfilling the objective (6) of the dissertation and answering the established research questions. All sections from this section (3.3.1-3.3.5) are related to the research questions (1) and (2). To answer them, the aim is to find the conformational changes of AK and their main important factors by studying important variables during the catalytic process. To answer question (3) we will analyse how the activation process starts in AK once the ligands have attached to the binding sites of the protein in sections 3.3.3 and 3.3.4. Finally, section 3.3.4 is focused on answering question (4), related to the effect of Mg2+ ions on the catalytic process.

### Study of system configuration

Molecular dynamics simulations are time-consuming and resource intensive. In this project, despite using the ARCAA and Hawk supercomputer clusters to run all the code, the total time to perform the 13 simulations was 420 hours (17 and a half days). During the six step described in the methodology, several parameters had to be defined. A poor choice of parameters can ruin an entire simulation as protein+ ligand molecular systems are very sensitive (Muller et al,1996). Therefore, a series of small simulations were performed to test the parameter set described above. To save time, parameters were chosen from previous studies in AK where their effectiveness had been proven for certain simulations (Zheng et al,2018). Analysis was performed only for NPT, NVT and energy minimisation balancing as these are the fundamental steps where the results can change the most depending on the parameters chosen. The open AK structure without ligands was chosen because, unlike the other cases, this topological structure did not require any additional modification. As soon as it was downloaded from the protein database it was tested during the first stages of the project. Once the results were validated, the same configuration and parameters were used for the remaining twelve simulations. Figure (5) shows the main graphs analysed to check if parameters were situ able so NPT, NVT and energy minimization stages worked well, and behaviour is as expected.

First and second graph starting from the top shows the variation of pressure and density during the NPT equilibration phrase. Pressure values fluctuates in a stable manner thought 100-ps NPT equilibration phase. Running average of data shows that the average value of the 235 ± 6 bar for the pressure and 1033 ± 4 kg/m3 for the density. Outcome is acceptable since reference values (Childers, M. C., & Daggett, V,2018) for pressure and density fall within the uncertainty range. In the case of pressure, fluctuation is more significant. As a result, both density and pressure levels are highly steady over time, suggesting that the system is well balanced.

Third graph shows the fluctuation of the system temperature. The temperature quickly converges to the target value of 300 K and remains stable between 295 K and 305 K over the NVT equilibration. As in the previous cases, running average of temperature also indicates a stability around an average value of 299 ±2 k. Punctual temperature picks don’t represent a problem since position restraints are not 100% perfect during the equilibration (Childers, M. C., & Daggett, V,2018)).

Finally, we looked at two important factors to evaluate whether energy minimization process was successful or not: potential energy and maximum force. Analysing the first factor, we observed that potential energy minimization rapidly converges to -5000 KJ /mol. Average values for a simple protein in water ranges between 105 and 106 although this value depends on the number of molecules and hydrogen bonds (Budhayash Gautam ,2020). For the second factor we look at, we observed that simulation stopped when the force reached > -5000.0 kJ/mol/ (Budhayash Gautam ,2020), indicating that the system arrives at a reasonable potential energy that is below the maximum forced allowed. Thus, we conclude that the systems stable enough to advance further steps in the simulation process.

Graphical user interface

Description automatically generated

**Figure (5):** Initial Calibration simulation. Four graphs above show NPT NVT and Energy minimization phases for open Adenylate Kinase simulation. Firs three graphs plot the average system density, pressure, and temperature over a time of 100 ps during NPT and NVT equilibration phases. Last graph shows how the system stabilizes the potential energy during the energy minimization for each time steps, corresponding to a total of 100ps.

### Ligand free Adenylate Kinase simulation

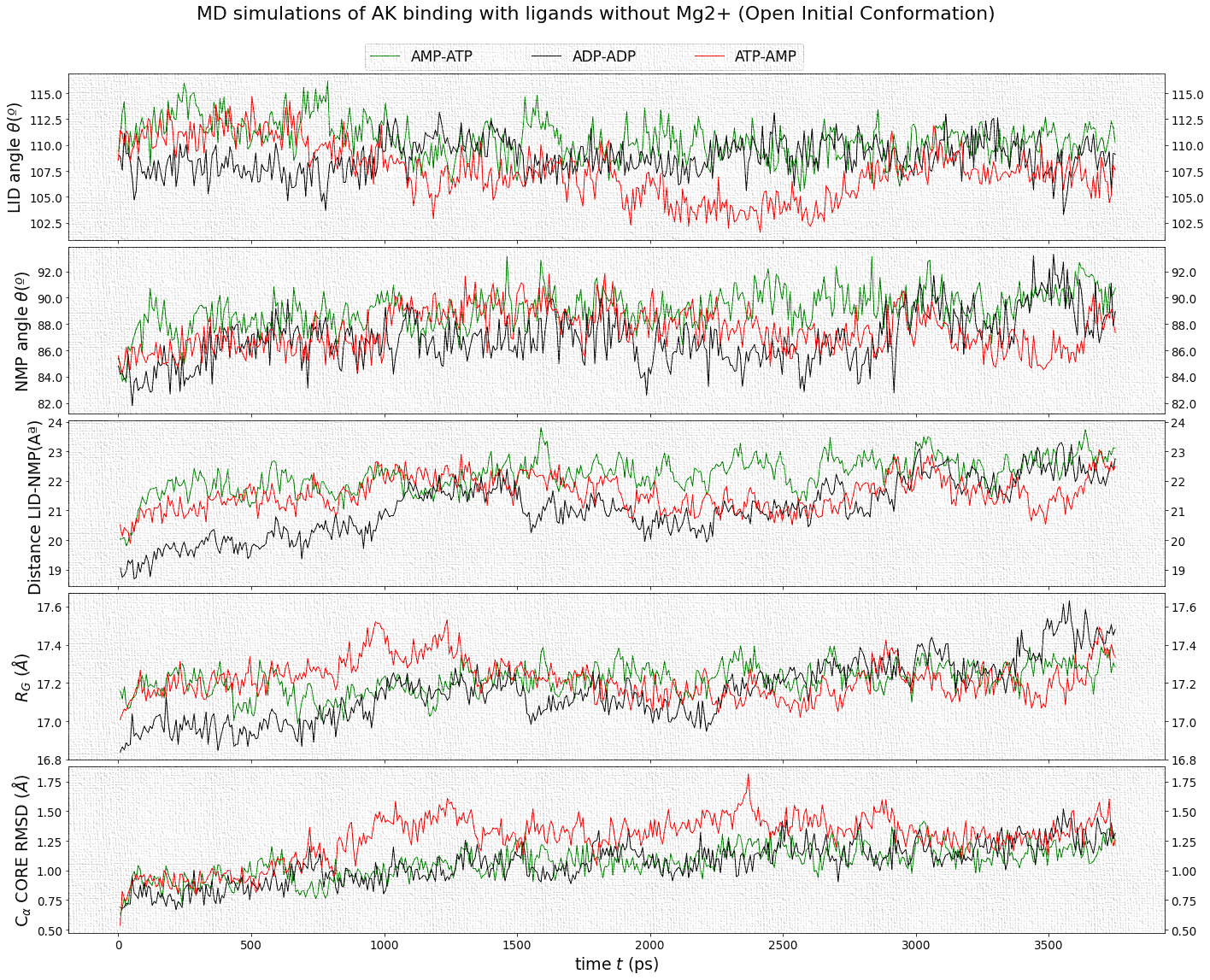
The study results show that the conformational changes and pathways in the ligand-free AK starting in the unbound open states evidence distinctive pathways. Figure (6) shows the of all the variables monitored during the simulation. AK preferred to shift its conformation from open form to closed form using the following distinct pathways. In this case, in the short time of 4 ns AK conformations remain very close to their equilibrium point. However, the final structure ends up being more closed than at the beginning of the simulation since the distance between the centre of mass NMP and LID domains, at initial time t=0 ns is 37 Å and at final time t=4 ns is 33.5 Å had decreased. LID and NMP angles don’t change significantly .However, there are some local-non permanent changes to consider: First, NMP domain semi-opens first at t=1.2 ns, then the LID domain moves to a semi-closed state (t=1.2 ns) and the NMP domain semi-opens at t=0.3 ns and semi-closed at t=1.2 ns, and finally the LID domain opens again (t=2.5 ns) This observation made from an open conformation to semi-close conformation to close to end in a semi open conformation is coherent and similar to the transitions obtained in previous studies.(Song HD, Zhu F,2013).Three time series show the RMSD levels off to ~1.2 Å, indicating that the structure is very stable. (Kufareva, I., & Abagyan, R, 2012). Finally, Gyrotron radius time series also indicates that the protein structure loses compactness through the simulation. However, it ranges between normal values of 28-22 Å and the mayor loos of compactness occurs at the same moment in time in which LID domain close, approximately (t=1.2 ns).



**Figure (6):** Results on a 4ns MD simulation trajectory for AK without ligands attached to binding sites LID and NMP. Initial conformation is an open structure (PDB: 4AKE). From top to bottom, variables used to monitor trajectories, described on methodology section 3.2.2 are: The angle between the cantesrs of mass of the carbon atoms and NMP and LID residues, the distance between the centre of mass NMP and LID domains, RMSD of the carbon atoms and the radius of Gyrotron . Red Arrows indicate local non-permanent fluctuations.

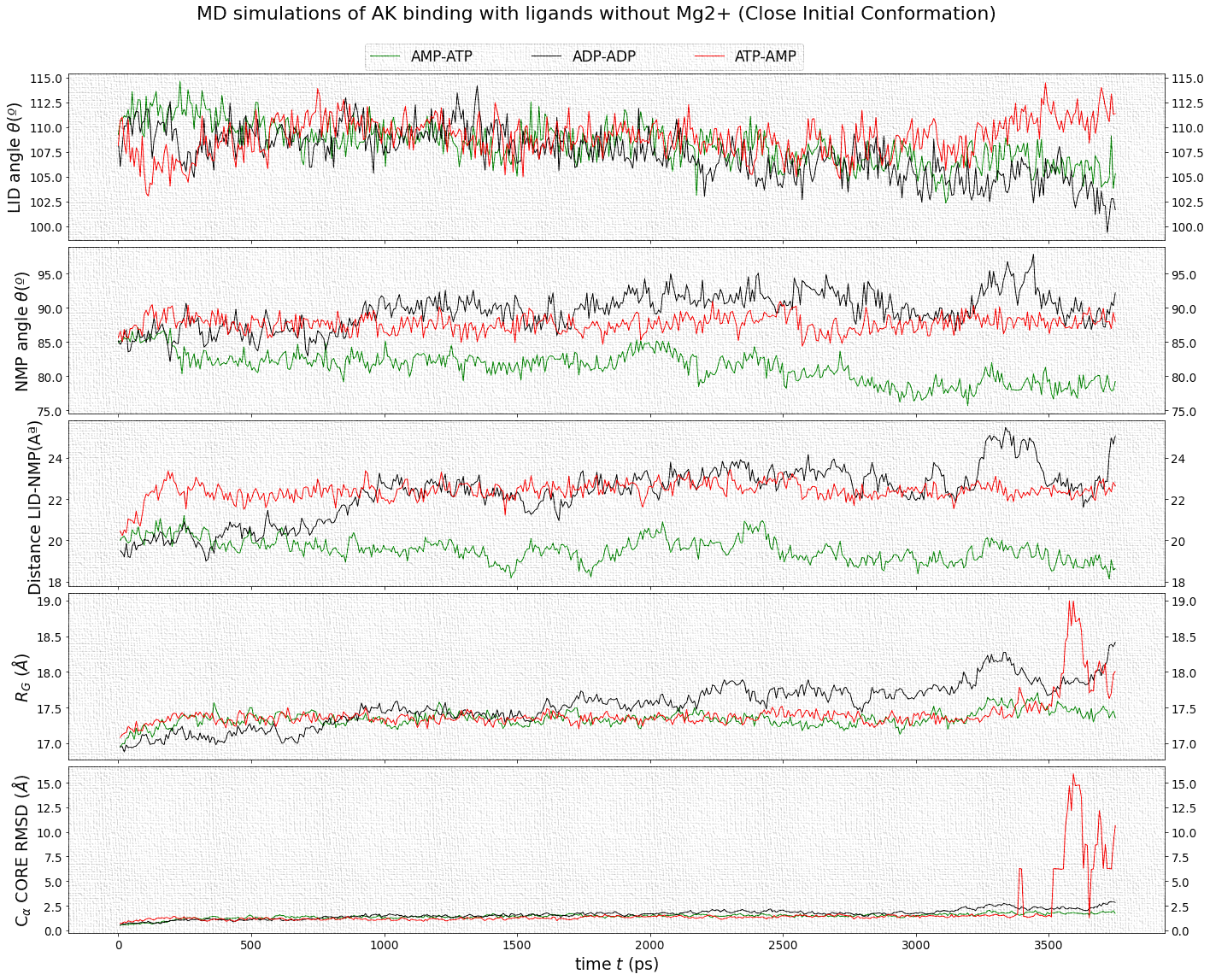
### Ligand-binding Adenylate Kinase simulations

Second set of six simulations started from both open (three simulations) and close (three simulations) initial conformations with ligands attached to the binding sites LID and NMP. The ligands bound to AK binding sites NMP and LID are the molecules, ATP, ADP, and AMP. Combinations used were ATP(LID)-ATP(NMP), ADP(LID)-AMP(NMP), AMP(LID)-ADP(NMP). Previous studies indicate that these combinations are the only allowed (Jana, B, et al ,2011). The changes we found are likely to be mainly fluctuations which can easily be reversed rather than significant conformational changes.. LID domain semi-close at t=0.75 ns (ADP-ADP) and t=1.1 ns (ATP-AMP) and semi- opens again at t=2.4 ns (ADP-ADP) and t=2.7 ns (ATP-AMP). On the other hand, NMP domain only performs local fluctuation instead of a significative movement in the ADP-ADP ligand situation. Observations indicates that AK prefers to range between intermediate semi-conformations without undergoing fully conformational Thus, results concludes that AK trends to be very stables at the open conformation when ligands are attached. Figure 7 show the of all the variables monitored during the simulation for the open initial conformation case:



**Figure (7):** Results on a 4 ns MD simulation trajectory for AK with ligands attached to the binding sites LID and NMP. Initial conformation is an open structure (PDB: 4AKE). From top to bottom, variables used to monitor trajectories, described on methodology section 3.2.2 are: The angle between the centre of mass of the carbon atoms and NMP and LID domains residues, the distance between the centre of mass NMP and LID domains, RMSD of the carbon atoms and the radius of Gyrotron . Black arrows indicate AK indicate non-permanent local changes for ADP-ADP and red arrows indicate non-permanent local changes for ATP-AMP.

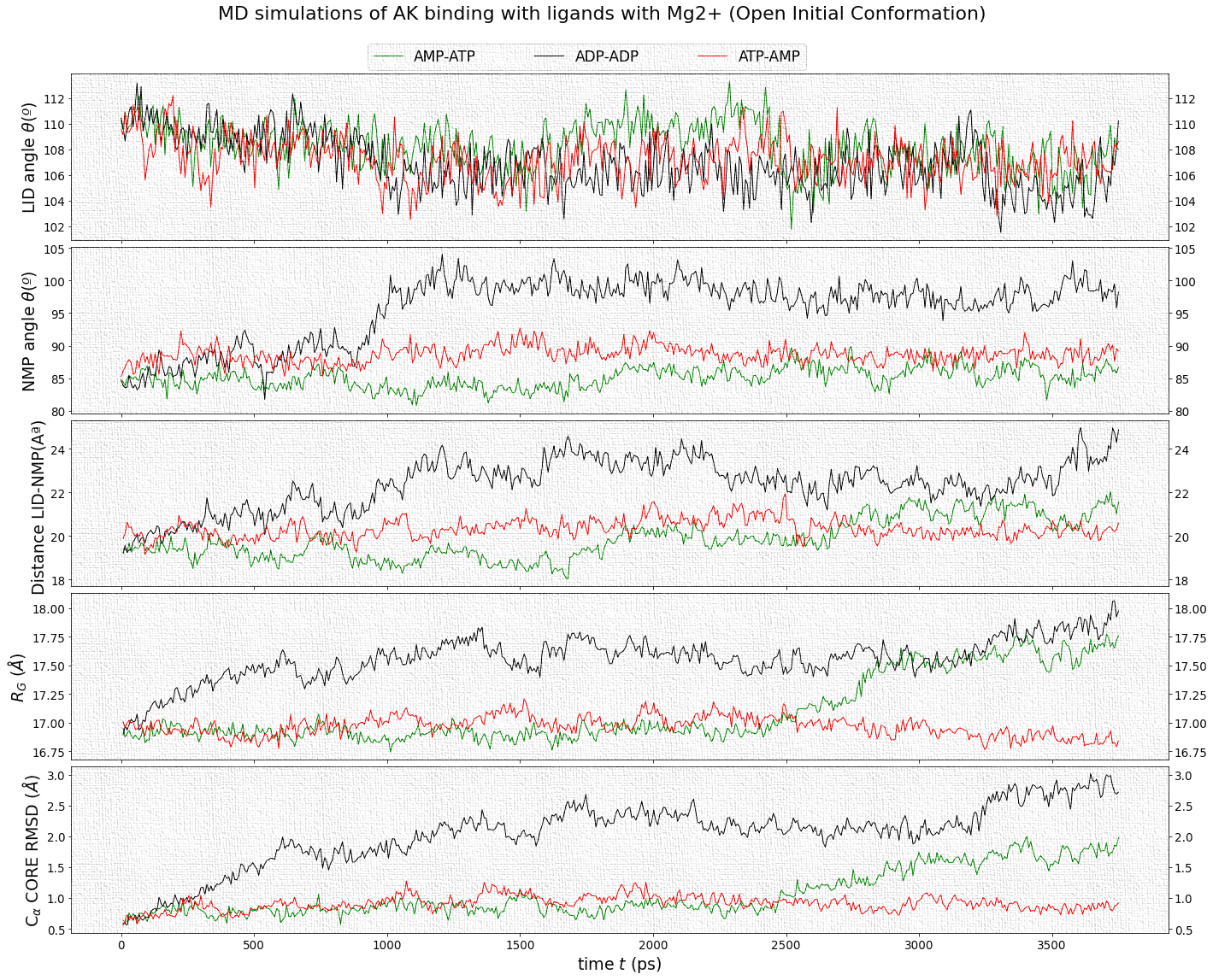
On the other, AK starting from a close conformation did not show full conformation changes but rather intermediate fluctuation between states. Initial close conformation starts very close to equilibrium and hence we wouldn’t expect major changes. As in the previous case, this shows that molecular dynamics approach works and gives sensible results These intermediate states are labelled semi-open and semi-closed structures which explains the motions by the NMP domain and the LID domain. In the case of ATP-AMP LID domain opens at t= 0.5 ns, closes at 1.7 ns, and opens again at t=3.3 ns where a very considerable increase of 10x in its RMSD is observed, which indicates together with the radius of gyration a clear transformation of the closed structure towards a more open one. In the case of ADP-ADP, only opening of the LID and NMP domain are observed around t= 1.1 ns and t= 3.3 ns respectively, which leads AK to a semi-closed and semi-open state. For the AMP-ATP case there is not present any significant fluctuations in both close and open initial configurations. Figure 8 shows all the variable monitored through the simulation for the close initial conformation case:



**Figure (8):** Results on a 4 ns MD simulation trajectory for AK with ligands attached to the binding sites LID and NMP. Initial conformation was obtained from a close structure topology file (PDB: 1AKE). From top to bottom, variables used to monitor trajectories, described on methodology section 3.2.2 are: The angle between the centres of mass of the carbon atoms and NMP and LID domains residues, the distance between the centre of mass NMP and LID domains, RMSD of the carbon atoms and the radius of Gyrotron . Black arrows indicate AK indicate non-permanent local changes for ADP-ADP and red arrows indicate non-permanent local changes for ATP-AMP.

### Ligand-binding Adenylate Kinase Simulations with Mg2+ ions

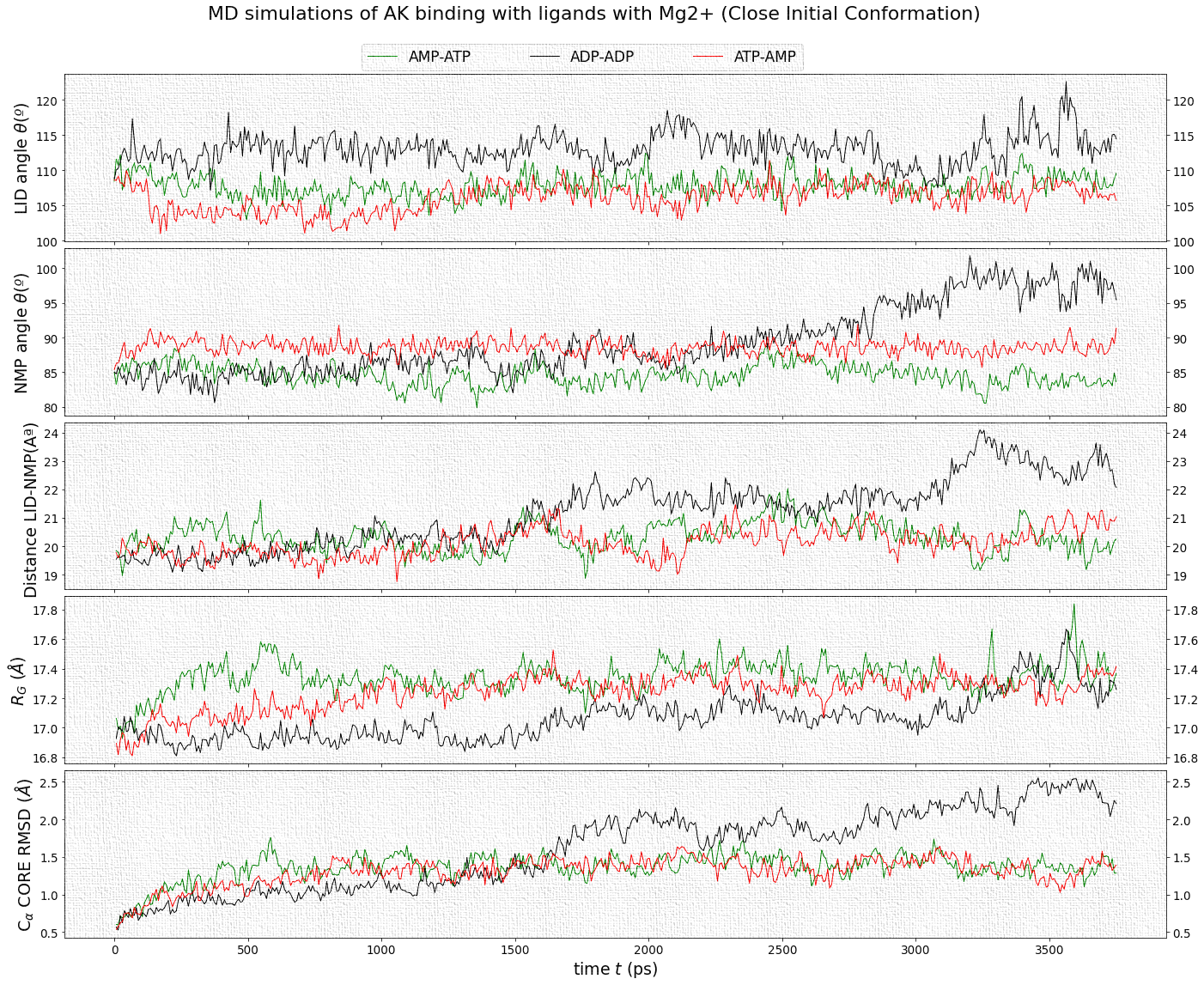
Final set of six simulations started from both open (three simulations) and close (three simulations) initial conformations with ligands attached to the binding sites LID and NMP. The ligands bound to AK binding sites NMP and LID are the molecules, ATP, ADP, and AMP. Combinations used were ATP(LID)-ATP(NMP), ADP(LID)-AMP(NMP), AMP(LID)-ADP(NMP). Results in both close and open initial conformations indicates that Mg2+ accelerates NMP opening only in the ADP-ADP ligand combination, which directly affect the number of intermediate steps that AK undergoes during the catalytic reaction. The effect is much more pronounced in the case of the open initial conformation (Figure 9), although compared to the reaction without Mg2+ ions in the previous section (3.3.4), strong changes are also seen in the case of the closed initial conformation (Figure 10). ATP-AMP ligand combinations only shows small non-permanent local- fluctuations between semi-open and semi-close states in the open initial conformation configuration. Thus, we conclude that configuration is very close to its respective equilibrium. For the AMP-ATP case there is not present any significant fluctuations in both close and open initial configurations.



**Figure (9):** Results on a 4 ns MD simulation trajectory for AK with ligands attached to the binding sites LID and NMP. Initial conformation was obtained from an open structure topology file (PDB: 4AKE). Mg 2+ ions are attached to all the ligands. From top to bottom, variables used to monitor trajectories, described on methodology section 3.2.2 are: The angle between the centre of mass of the carbon atoms and NMP and LID domains residues, the distance between the centre of mass NMP and LID domains, RMSD of the carbon atoms and the radius of Gyrotron . Black arrows indicate AK conformational changes for ADP-ADP and red arrows indicate non-permanent local changes for ATP-AMP-

As can be seen in Figure (9), specifically in the case of ADP-ADP ligands combination, the closure of the NMP domain and the LID domain are very marked for time t=0.5 ns and t=1 ns respectively. In this case, since the critical angle values are exceeded, it is considered a conformational change and not a fluctuation. Then LID and NMP domain open again at t= 3.4 ns and t= 1.1 ns respectively. In the case of AMP-ATP ligands combination, the changes are just-locally and non-permanent. Finally, the behaviour for the ATP-AMP ligands combination case is is very similar to the AMP-ATP ligands combination case and no significant change in the NMP and LID angle are observed. It can therefore be concluded that for the ATP-AMP AK goes from an open to a semi-closed state through the semi-closure of the LID domain at time t=0.5 ns, semi-opens at t=1.1 ns and finally semi-closes at t=2.5 ns.

On the other hand, in Figure (10) we can observe a large variation of conformational changes for the initial closed configuration. In this case, it is once again the ADP-ADP ligand that most intensely acquires a open configuration at t=2.8 ns with the opening of the NMP domain. LID domain on the other hand opens at t=2 ns, closes at t=2.8 ns and opens again at t= 3.4 ns. In this case, since the critical angle values are exceeded, it is considered a conformational change and not a fluctuation. In the case of the ATP-AMP ligands, no significant changes are observed, indicating that the system oscillates around equilibrium. Finally, for the AMP-ATP ligands while the NMP and LID domain remains stable.



**Figure (10):** Results on a 4ns MD simulation trajectory for AK with ligands attached to the binding sites LID and NMP. Initial conformation was obtained from a close structure topology file (PDB: 1ake). Mg2+ ions are attached to all the ligands. From top to bottom, variables used to monitor trajectories, described on methodology section 3.2.2 are: The angle between the centre of mass of the carbon atoms and NMP and LID domains residues, the distance between the centre of mass NMP and LID domains, RMSD of the carbon atoms and the radius of Gyrotron . Black arrows indicate AK conformational changes for ADP-ADP.

### Mapping the transitions and pathways of Adenylate Kinase

As we can see from table (1), transitions in ADP-ADP case are the only ones that undergoes conformational changes, varying the folding angle and the distance between LID and NMP domains significantly (with peaks of ±2 º for the angles and ±3 A for the distance). For the rest of ligand combinations transitions correspond to local fluctuations around equilibrium, so they cannot be categorized as conformational changes However, several fluctuations between intermediate states were observed on the ligand free simulations comparing it to the ligand-bound case. Thus, the main implication extracted from the simulation results is that AK is activated more by ADP-ADP than by other ligands and NMP opening is more predominant transition.

Another conclusion is drawn regarding the role of the metal cofactor, and it is that Mg+2 ions aids in accelerating NMP and LID closure and opening. The investigation of the metal cofactor leads to studying how Mg+2 aids in lowering the activation barriers of the primary steps in a catalytic cycle. Additionally, the process correctly positions the active-site players for effective P-transfer. The suggestion made in this case is that substitution of divalent metals for Mg+2 would activate NMP-opening just like P-transfer aids in charging the density of Mg+2. One possible hypothesis is that Mg2+ prevents H20 molecules trapped in the water solvent boxed to get exchange with bulk water (Guardia et al, 1999). Since there are more water molecules with Mg2+ in the cavity AK can performed easily conformational transitions. Other of the perceptions formed is that the Mg+2 prevents the high electrostatic interactions between substrates and the active site thus facilitating NMP-opening.

|  |  |  |  |
| --- | --- | --- | --- |
| **Starting Structure** | **Ligands** | **Conformational Transitions** | |
| **LID domain** | **NMP domain** |
| Open | Ligand-free  ATP-AMP  ATP-Mg2+-AMP  ADP-ADP  ADP-Mg2+-ADP  AMP-ATP  AMP-Mg2+-ATP | semi-close (t= 1.2 ns) ↔ semi-open (t=2.5 ns)  none  semi-close (t=0.5 ns) ↔ semi-open (t=1.1 ns) ↔semi-close (t=2.5 ns)  close (t=0.75 ns) ↔ open (t=2.7 ns)  close (t=0.5 ns) ↔ open (t=1 ns)  none  none | semi-open (t= 0.5 ns) ↔ semi- close (t=1.1 ns) ↔ closed (t=2 ns)  none  none  close (t= 2 ns) ↔ open (t= 3 ns)  none  none  none |
| Close | ATP-AMP  ATP-Mg2+-AMP  ADP-ADP  ADP-Mg2+-ADP  AMP-ATP  AMP-Mg2+-ATP | open (t=0.5 ns) ↔semi- close (t=1.7 ns) ↔ semi-open (t=3.3 ns)  none  semi-open (t=2.7 ns)  semi-open (t=2 ns)- ↔close (t=2.8 ns)  none  none | none  none  semi-open (t=2.7 ns)  semi-open (t=2.8 ns)  none  none |

**Table (1):** Summary of conformational transitions and transitional pathways observed by MD simulations.

It is useful to be able to observe graphically the conformation of the enzyme as time progresses in a clear and distinguishable way between its CORE, LID and NMP dominions. Following image created with VMD summaries all the conformational changes captured through the transitional pathways:

|  |  |  |
| --- | --- | --- |
| **Initial Open Conformation** | | |
| Ligand free | |  |
| ATP-Mg2+  -AMP | |  |
| ADP-ADP | |  |
| **Initial Close Conformation** | | |
| ATP-AMP |  | |
| ADP-Mg2+  -ADP |  | |

**Figure (11):** VMD snapshots of AK at time t=0 ns, t=2 ns and t= 4 ns. GREEN, BLUE, and RED colours represent LID, CORE and NMP domain respectively. Only most relevant cases from Table (1) were selected and represented.

## Conclusions

At the end of the project, if we analyse the results in relation to the proposed objectives, we may come to the following conclusions: The first objective of the project, which consisted of documenting the entire methodological process in a simple tutorial to facilitate the replication of the simulations, has been fulfilled; all the information can be found in the Appendix of the project. The second objective has also been completed satisfactorily, from the close and open conformation of AK acquire from the E. coli bacteria (Appendix D) it has been possible to create all the proposed topological files to carry out the simulations. Objectives 3, 4 and 5 aimed to implement molecular dynamics simulations in AK for different scenarios. A total of 13 simulations of 4000 ps for ligand free AK and ligand binding AK to ATP, ADP, and AMP, starting from an open and close conformation and including the effects of Mg2+ ions, were carried out. Finally, objective 6 sought to monitor the variables of interest using the trajectories of the atoms as a function of time. This is perhaps the most important objective of the dissertation as it allows us to answer all the research questions from a quantitative point of view.

Extensive research on AKs and Kinases has revealed that enzyme catalysis is reduced to a steady state that is a collection of numerous reaction steps. This way of thinking has aided in the misunderstanding of mechanical concepts (Wang, Z., & Cole, P. A, 2014) because conformational movements are overlooked or hidden in many instances. Thus, to check consistency of the results, previous studies within a different set of parameters and time frame were considered, especially those performed on AK from (Li et al,2015) and (Kerns et al, 2015). Our MD simulations results (Table 1) show that AK remains in a kinetic stability between close-semi-open and open–semi-closed states without any ligands. Attachment of ADP, AMP or ATP ligands alter the close and open conformation populations and cause AK to fluctuates towards an open/semi-open or semi-close state. We also found that there is a clear effect of Mg2+ of ADP-ADP leading to an opening of the NMP. However, it is not known exactly how Mg2+accelerates the conformational changes and what the inner workings are. This interesting result opens the door for future research to answer how Mg+2 aids in accelerating the activation process of the primary steps in a catalytic cycle.

## Suggestions for future work

Determining the conformational change of protein kinase AK is extremely difficult since protein conformational changes normally spread a large variety of timeframes, ranging from nanosecond to microseconds and milliseconds as well as size scales. Future work will require perform simulations within a more extensive range of time and size scales. This will allow to further investigate the phosphoryl-transfer process and could open new avenues for studying crystal structure. Using the methodology described in the discussion, the investigation was able to comprehensively establish at an atomistic level the transition pathways of AK. Apart from the determination of transition pathways the study investigates the different AK forms in the absence and existence of ligands. Following the results, the discussion reveals the functional domains in AK and more so in chronological operation. The results show that the binding of AK with a ligand leads to the AK shifting to a closed form expounding on the insights related to the subtleties of AK. The complex conformational conversions of AK aid in its role of regulating enzymatic functions. The suggestion following the end of the study is a hybrid of conformational selection of ligand biding to AK.

Other aspect that was not consider in this project for establishing rational relationships is the quantitative effect of quantum mechanics (QM). Molecular dynamics is based entirely on Newton’s laws of motion so cannot be used for QM. A possible technique is to use MD for an outer region and use QM calculations for a very small inner region(Shen, Lin & Yang,2018).However, the technique is quite computationally intensive and limits to small systems for short timescales. Due to the resources of the project, it was not possible to request such a high amount of computational power. Thus, questions concerning the transition routes and the selection process for conformation / population transfer in AK remain unanswered. As a result, employing alternative approaches to gain a quantitative knowledge of mechanical variations of AK will be extremely helpful in future studies.

Finally, it is important to note that enzymes can only be improved by continuous reduction of the unwanted and unproductive fluctuations shown in the active sites. Further research is required to obtain a deeper knowledge of the entire landscape of energy in the catalysis of enzymes. The impressive rate of acceleration achieved in Biology by biological catalysts remains an interesting subject. X-rays show structures of the active complexes, kinases’ computational studies, and kinetic characterizations that are detailed but there is more to be studied regarding kinases and their function. In the case of hydrogen-transfer enzymes, more investigation has been carried out regarding conformational sampling, pre-organization, and its impact on active sites. Hopefully, the comprehensive body of work presented on AK catalytic framework will provide a foundation that will improve the understanding of the catalytic power of kinases.

# Critique of project

## Evaluation of the initial plan and the actual plan

This dissertation has been realised during a 12-week period, beginning on the 11/06/21 and ending on 03/09/21.Workflow has been breakdown into 4 distinct phases: First, a theory refresh phase in which the author has learned GROMACS and all the relevant literature. Second, a pre-simulation phase in which all the necessary files and code were created from scratch. Third, a simulation phase, in which all the scripts were executed in conjunction with ARCAA, and Hawk supercomputer cluster and the results files were extracted. Finally, a fourth phase that was used for analysing and discussing the results and writing the dissertation in parallel.

A comparison between the original research proposal plan and the final execution plan of the project can be seen in the following Gantt charts:

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **ORIGINAL PLAN** | | | | | | | | | | | | | |
| **TASK** | **Week** | | | | | | | | | | | | |
| **0** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** | **11** | **12** |
| **11/6** | **18/6** | **25/6** | **02/7** | **09/7** | **16/7** | **23/7** | **30/7** | **06/8** | **13/8** | **20/8** | **27/8** | **03/9** |
| Q&A Supervisor meetings |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Literature Review |  |  | ---- |  |  |  |  |  |  |  |  |  |  |
| Learning GROMACS |  |  | transparent background target icon |  |  |  |  |  |  |  |  |  |  |
| Compiling necessary files |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Setting up model & Writing code |  |  |  |  |  | transparent background target icon |  |  |  |  |  |  |  |
| Running Simulations |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Analysing Transional Paths |  |  |  |  |  |  |  |  |  | transparent background target icon |  |  |  |
| Writing Up first formal draft |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Implementing Feedback |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Final Draft Proofreading |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Submitting Dissertation |  |  |  |  |  |  |  |  |  |  |  |  | transparent background target icon |
| **FINAL PLAN** | | | | | | | | | | | | | |
| Q&A Supervisor meetings |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Literature Review |  |  | ---- |  |  |  |  |  |  |  |  |  |  |
| Learning GROMACS |  |  |  | transparent background target icon |  |  |  |  |  |  |  |  |  |
| Compiling necessary files |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Setting up model & Writing code |  |  |  |  | transparent background target icon |  |  |  |  |  |  |  |  |
| Running Simulations |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Analysing Transional Paths |  |  |  |  |  |  |  |  |  |  | transparent background target icon |  |  |
| Writing Up first formal draft |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Implementing Feedback |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Final Draft Proofreading |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Submitting Dissertation |  |  |  |  |  |  |  |  |  |  |  |  | transparent background target icon |

**Figure (12):** A Gantt chart comparing the original plan in the top and the actual plan in the bottom. Project plan is composed by 4 Phases represented with different colours. For the original plan, darker colours represent scheduled work weeks, and lighter colours represent time left over if contingencies arise.

As can be seen in Figure (12), both real and planed schedules were very similar and all and almost all the proposed contingency blocks accomplish the time demands of the project. Furthermore, the main 4 phases remain the same since, as discussed before, the objectives requirements were identically as the ones from the proposal. However, the implemented plan underwent modifications in several ways, detailed below:

* In the original plan,3 weeks were allocated for obtaining the enzyme and ligands topology files, however it takes more time to obtain because there need to be created. Thanks to Georgina Mendez, ATP, ADP, and AMP molecules could be constructed from previous .tpr files using a full nucleotide atom list and marking the necessary changes. These modifications were essential to generate all the ligand-attached topology files and run close Ak simulations.
* Supervisor meetings were planned to be on a weekly basis during the whole project. However, weekly 30 mins meetings were performed only during the last stage of the project, until week 4. Thereafter, 1-hour meetings were performed on a bi-weekly basis as it was found to be more effective to space them out and collect more information to save the time of the supervisor and the project author.
* Literature review and learning GROMACS section took more time than expected, taking an extra week off the contingency plan. This is since most of the literature had to be re-collected and rewritten in smaller, more understandable paragraphs, and had to be expanded further on the energy landscape as proposed in the proposal feedback. On the other hand, unexpected bugs appeared in the GROMACS scripts and required more time to resolve.
* During Phase 3, generating Plots during Energy minimization and visualizing transitional pathways required to write extra bits of code in Python and Stata. Original plan assumed that the author's knowledge of the software would not delay the analysis phase of the simulations. However, all contingencies block from were finally taken and it could have been foreseen and solved in advance.

In conclusion, although there have been some minor modifications, the final development of the project compared to the original plan has been very similar. This result was expected for two main reasons. First, the inherent nature of this project by strongly relying on the analysis of the results of the simulations allowed for the construction of a general and simplified proposal to facilitate its implementation. Second, the plan had plenty of allocated time for contingencies, and especially for the most exigent phase: the execution and the obtention of quantitative results.

Objectives and research questions presented in in section 3.1 changed significantly to those set out in the original plan. Based on feedback from the supervisor, the performance objectives were further defined for the ligand binding and ligand free simulations and the Mg2+ factor was considered. Once objective 3 and 4 were complete the results showed little or no conformational change, and a stable evolution of the conformational protein to its centre of gravity. Thus, it was decided to add objective 5 together with the execution of a new set of simulations to capture more conformational changes. The research and information gathering objectives were transformed into objective 1: Tutorial for the replication of the methodology. Finally, the objective of structure investigation using x-ray diffraction was removed as it was more effective to obtain these files directly from the protein database. The final number of 6 objectives was increased by 2 with respect to the objectives set out in the proposal (eight objectives). Regarding research questions the first 4 questions of the proposal were kept, and four extra questions were removed as the project evolved since they involved energy analysis, which could not be carried out due to time limitations.

## Scientific merits

The most interesting merits achieved in this project are summarised below:

* The results conclude that AK undergoes multiple transitional pathways from an initial open conformation and less transitions from an initial close conformation. This result implies that short timescales simulations are effective to snapshot protein mechanism and supports previous studies such as (Ping et al,2013), (Sean L et al,2014) and (Zhu F, 2013). Due to the large number of results indicating multiple transitions, this research supports the need for further investigation with more computational power and time resources considering quantum mechanics effects.
* Previous studies have provided guides to perform molecular dynamics with AK. However, most of them don’t include all the necessary steps and can be confusing to replicate the results. This project has presented a tutorial, exposing all the infrastructure and programs used. The appendix of this project includes a well-structured guide with all the GROMACS commands, together with explanations regarding the files generated as well as the abbreviations that GROMACS used when creating new files. We have also included all the pythons’ scripts necessary for analysing the energies and transitional pathways of AK. Methodology of this project can be extrapolated for different biomolecular systems, and we hope that for future AK-specific projects, this paper can be used as a reliable guide by the scientific community.
* The most comparable simulation to this work is carried out by (Li et al,2015), where BE-META dynamics simulations are considered. However, Li assumes large conformational times (up to 100 ns) Furthermore ligand-binding initial states are only analysed from an initial open AK conformation, and in this work, we also analysed an initial close structure conformation. Throughout the project, it was evident that short simulation time of 4 ns could be considered inadequate for observing large conformational transitions. Results of this project show that simulate AK conformational changes are difficult to catch. The implications of this are that there the demand for new, reliable methods which measures Ak transitions within an even larger frame.
* This project fits within the previous work done by Dr Georgine Mendez and Dr Emyr Macdonald within ARCA and Hawk collaboration to better understand the structural mechanics of Ak during catalysis. Although like those applied in previous papers, all techniques, graphs, and code were originally created, what gives this work scientific merits and makes it indispensable for future work.
* Further studies to biomolecular systems and various enzymes may replicate the computational model established in this study. The methods presented in this work are reliable, and the results are consistent. Overall, the rational design of ligands aiming at the shift of the population as well as the flexibility of protein is a crucial practice for the development of biomolecular systems such as enzymes. Thus, rather than a new contribution in this field of itself, a mayor scientific merit of this project lies in the work’s potential applicability to others.

## Reflections on the dissertation process

Below is a thorough list of suggestions that could be useful to a student working on a similar project in the future. This is a list of things I've learned about myself and my work that I wish I'd known at the start of the project:

* It would have been very useful to allocate more extra weeks in the project planning schedule for learning a completely new software. Learning how to code in a new program is a great challenge that may   
  take more time than learning any other subject of the dissertation. It is better to plan the contingencies in advance on this issue. For my project however, and after a lot of effort, I have been able to obtain a good working knowledge of both GROMACS and VMD, and now I feel capable of using them in to explore other proteins and different structures, which will be very useful both for my future PhD tesis.I feel grateful of having taken a rare opportunity to explore and develop a completely new coding skill, And if you had to go back to the beginning and choose whether to carry out a project that involves a challenge like this or not, I would definitely choose this project again.
* Supervisor meetings are essential for monitoring the progress of the project. I took the initiative and set most of the meetings in agreement with my supervisor, even asking for extra meetings in case he had a lot of news to discuss. It was very important to keep a bi-weekly frequency since it allowed me to rectify in case the results were moving in the wrong direction. During all meetings I generated a record in a notebook with questions, notes, and other considerations for future meetings. The main mistake I made is that I wasted a lot of time, especially in the first meetings, with technical questions about programming in GROMACS and VMD.
* Through all the project I created a project diary that helped me to compile all the essential information that I was doing every week. I was constantly updated this document and it was very useful to remind me what problems and difficulties I found, and I had to ask. As a non-native English speaker one of my difficulties was not being able to communicate doubts in the most effective way. However, constantly writing this diary together with the dissertation helps me tremendously to communicative better to my supervisor. Furthermore, well written and separated documentation in my diary of both results and scripts helped to avoid a lot of confusion and to find faster the required files for the simulations.
* One of the main requirements that determine the quality of the project is the knowledge of the fundamental theoretical background. I studied all the articles proposed by my supervisor. However, a greater understanding of bioinformatics and general biology through my own research would have helped. It is vital to acquire the foundations and backgrounds at the very beginning of the project and constantly read and compile relevant literature.
* I strongly believe that a dissertation is much more than a piece of work: it is a personal development path, and it should be carried out without anxiety and with a positive mindset. This project has given me the opportunity to learn about molecular dynamics, a field of great interest where experts and researchers from different areas such as physics, chemistry and biology collaborate with each other. I always tried to remind me that the dissertation process is not only about the end results but also about the skills acquired, the critical thinking, and the overall scientific research experience.

# References

[1] Kerns, S.J., Agafonov, R.V., Cho, Y.J., Pontiggia, F., Otten, R., Pachov, D.V., Kutter, S., Phung, L.A., Murphy, P.N., Thai, V. and Alber, T., 2015. The energy landscape of adenylate kinase during catalysis. Nature structural & molecular biology, 22(2), p.124.

[2] Schroeder GK, Lad C, Wyman P, Williams NH, Wolfenden R, 2006. The time required for water attack at the phosphorus atom of simple phosphodiesters and of DNA. Proceedings of the National Academy of Sciences of the United States of America.; 103:4052–4055.

[3] Li D, Liu MS, Ji B. 2015 Mapping the Dynamics Landscape of Conformational Transitions in Enzyme: The Adenylate Kinase Case. Biophys J. Aug 4;109(3):647-60.

[4] Durrant, J.D., McCammon,2011. J.A. Molecular dynamics simulations and drug discovery. BMC Biol 9, 713.

[5] Ji, Y., Yang, C., Tang, Z., Yang, Y., Tian, Y., Yao, H., Zhu, X., Zhang, Z., Ji, J. and Zheng, X., 2017. Adenylate kinase hCINAP determines self-renewal of colorectal cancer stem cells by facilitating LDHA phosphorylation. Nature communications, 8(1), pp.1-16.

[6] Rosing, J., and Slater, E.C. The value of G^o of the hydrolysis of ATP. 1972. Biochim Biophys Acta. 267(2):275-90. Table VIII pp. 287

[7]. Kamerlin SC, Sharma PK, Prasad RB, Warshel A, 2013.Why nature really chose phosphate. Q Rev Biophys pp 46:1–132.

[8] Bommer, G.T., Van Schaftingen, E. and Veiga-da-Cunha, M., 2020. Metabolite repair enzymes control metabolic damage in glycolysis. Trends in biochemical sciences, 45(3), pp.228-243

[9] Lassila JK, Zalatan JG, Herschlag D,2011. Biological phosphoryl-transfer reactions: understanding mechanism and catalysis. Annu Rev Biochem.; 80:669–702.

[10]. Thatcher GRJ, Kluger R, 1989.Mechanism and Catalysis of Nucleophilic-Substitution in Phosphate Esters. Advances in Physical Organic Chemistry; 25:99–265.

[11] Karplus and Maragakis, P,2005. Large amplitude conformational change in proteins explored with a plastic network model: adenylate kinase. J. Mol. Biol. 352:807–822.

[12] Astuti, A. & Mutiara, Achmad, 2009. Performance Analysis on Molecular Dynamics Simulation of Protein Using GROMACS.

[13] Bekker et al,1993. Gromacs: A parallel computer for molecular dynamics simulations. Physics Computing. 92. 252-256.

[14] Tan, Y. W., Hanson, J. A., & Yang, H. 2009. Direct Mg(2+) binding activates adenylate kinase from Escherichia coli. The Journal of biological chemistry, 284(5), 3306–3313.

[15] Bhabha, G., Ekiert, D. C., Jennewein, M., Zmasek, C. M., Tuttle, L. M., Kroon, G., Dyson, H. J., Godzik, A., Wilson, I. A., & Wright, P. E. 2013. Divergent evolution of protein conformational dynamics in dihydrofolate reductase. Nature structural & molecular biology, 20(11), 1243–1249.

[16] Wang, J., Peng, C., Yu, Y., Chen, Z., Xu, Z., Cai, T., Shao, Q., Shi, J. and Zhu, W., 2020. Exploring conformational change of adenylate kinase by replica exchange molecular dynamic simulation. Biophysical journal, 118(5), pp.1009-1018

[17] Endicott JA, Noble ME, Johnson LN. The structural basis for control of eukaryotic protein kinases,2007. Annu Rev Biochem. pp.81:587-613.

[18] Aviram, H.Y., Pirchi, M., Mazal, H., Barak, Y., Riven, I. and Haran, G., 2018. Direct observation of ultrafast large-scale dynamics of an enzyme under turnover conditions. Proceedings of the National Academy of Sciences, 115(13), pp.3243-3248.

[19] Hanson J and Yang H. , . 2009. Direct Mg(2+) binding activates adenylate kinase from Escherichia coli. J Biol ChemJan 30;284(5):3306-3313

[20] Adén, J., & Wolf-Watz, M. 2007. NMR identification of transient complexes critical to adenylate kinase catalysis. Journal of the American Chemical Society, 129(45), 14003–14012.

[21] Henzler-Wildman, K.A., Thai, V., Lei, M., Ott, M., Wolf-Watz, M., Fenn, T., Pozharski, E., Wilson, M.A., Petsko, G.A., Karplus, M. and Hübner, C.G., 2007. Intrinsic motions along an enzymatic reaction trajectory. Nature, 450(7171), pp.838-844.

[22] Danielson, M. A., Biemann, H. P., Koshland, D. E., Jr, & Falke, J. J, 1994. Attractant- and disulfide-induced conformational changes in the ligand binding domain of the chemotaxis aspartate receptor: a 19F NMR study. Biochemistry, 33(20), 6100–6109.

[23] Pontiggia, Francesco & Zen, Andrea & Micheletti, Cristian, 2008. Small- and Large-Scale Conformational Changes of Adenylate Kinase: A Molecular Dynamics Study of the Subdomain Motion and Mechanics. Biophysical journal. 95. 5901-12.

[24] Wolf-Watz, M., Olsson, U. , 2010 Overlap between folding and functional energy landscapes for adenylate kinase conformational change. Nat Commun 1, pp 111.

[25] Sullivan, S.M., & Holyoak, T, 2008. Enzymes with lid-gated active sites must operate by an induced fit mechanism instead of conformational selection. *Proceedings of the National Academy of Sciences, 105*, 13829 - 13834.

[26] Hammes, G. G., Chang, Y. C., & Oas, T. G., 2009. Conformational selection or induced fit: a flux description of reaction mechanism. Proceedings of the National Academy of Sciences of the United States of America, 106(33), 13737–13741.

[27] Berry, M. B., Meador, B., Bilderback, T., Liang, P., Glaser, M., & Phillips, G. N., Jr., 1994. The closed conformation of a highly flexible protein: the structure of E. coli adenylate kinase with bound AMP and AMPPNP. Proteins, 19(3), 183–198.

[28] Beckstein, Oliver & Denning, Elizabeth & Perilla, Juan & Woolf, Thomas. ,2009. Zipping and Unzipping of Adenylate Kinase: Atomistic Insights into the Ensemble of Open ↔ Closed Transitions. Journal of molecular biology,pp 394. 160-76.

[29] Muller, C. W., G. J. Schlauderer, J. Reinstein, and G. E. Schulz. ,1996. Adenylate kinase motions during catalysis: An energetic counterweight balancing substrate binding. Structure 4:147-156.

[30] Zheng, Y. and Cui, Q., 2018. Multiple pathways and time scales for conformational transitions in apo-adenylate kinase. Journal of chemical theory and computation, 14(3), pp.1716-1726.

[31] Childers, M. C., & Daggett, V. ,2018. Validating Molecular Dynamics Simulations against Experimental Observables considering Underlying Conformational Ensembles. The journal of physical chemistry. B, 122(26), 6673–6689.

[32] Budhayash Gautam , 2020. Energy Minimization, Homology Molecular Modeling - Perspectives and Applications, Rafael Trindade Maia, Rômulo Maciel de Moraes Filho and Magnólia Campos, IntechOpen,

[33] Song HD, Zhu F ,2013. Conformational Dynamics of a Ligand-Free Adenylate Kinase. PLoS ONE 8(7): e68023.

[34] Kufareva, I., & Abagyan, R.,2012. Methods of protein structure comparison. Methods in molecular biology (Clifton, N.J.), 857, 231–257.

[35] Jana, B., Adkar, B. V., Biswas, R., & Bagchi, B.,2011. Dynamic coupling between the LID and NMP domain motions in the catalytic conversion of ATP and AMP to ADP by adenylate kinase. The Journal of chemical physics, 134(3), 035101.

[36] Guardia, Elvira & Sesé, Gemma & Padró, J. & Kalko, Susana.,1999. Molecular Dynamics Simulation of Mg2+ and Ca2+ Ions in Water. Journal of Solution Chemistry. 28.

[37] Wang, Z., & Cole, P. A., 2014. Catalytic mechanisms and regulation of protein kinases. Methods in enzymology, 548, 1–21.

[38] Ping, Jie et al,2013. “Molecular dynamics studies on the conformational transitions of adenylate kinase: a computational evidence for the conformational selection mechanism.” BioMed research international.

[39] Sean L. Seyler & Oliver Beckstein, 2014.Sampling large conformational transitions: adenylate kinase as a testing ground, Molecular Simulation, 40:10-11, 855-877

[40] Zhu F, 2013. Conformational Dynamics of a Ligand-Free Adenylate Kinase. PLoS ONE 8(7): e68023.

[41] Harrison GJ, Willis RJ, Headrick JP.,1998. Extracellular adenosine levels and cellular energy metabolism in ischemically preconditioned rat heart. Cardiovasc. Res; pp 40:74–87.

[42] T. M. Picknett, S. Brenner,2001. “X-Ray Crystallography”, Encyclopedia of Genetics, 2154.

Formoso, E., Limongelli, V. & Parrinello, M.,2015. Energetics and Structural Characterization of the large-scale Functional Motion of Adenylate Kinase. Sci Rep 5, 8425.

[43] Ercheng Wang, Huiyong Sun, Junmei Wang, Zhe Wang, Hui Liu, John Z. H. Zhang\*, and TingjunHou,2005. End-Point Binding Free Energy Calculation with MM/PBSA and MM/GBSA: Strategies and Applications in Drug Design Chem. Rev,47,1589−1614.

[44] Borhani, David & Shaw, David. ,2011. The future of molecular dynamic simulation in drug discovery. Journal of computer-aided molecular design. 26. 15-26.

[45] Ionescu, M.I,2019, Adenylate Kinase: A Ubiquitous Enzyme Correlated with Medical Conditions. Protein J 38, 120–133.

[46] Shen, Lin & Yang,2018. Molecular Dynamics Simulations with Quantum Mechanics / Molecular Mechanics and Adaptive Neural Networks. Journal of Chemical Theory and Computation. 14.

[47] Palsson, B,2011. Systems Biology: Simulation of Dynamic Network States. Cambridge: Cambridge University Press.

[48] Greives, N., and Zhou, H.X., 2014. Both protein dynamics and ligand concentration can shift the binding mechanism between conformational selection and induced fit. Proceedings of the National Academy of Sciences, 111(28), pp.10197-10202.

[49] Liu, X., Shi, D., Zhou, S., Liu, H., Liu, H. and Yao, X., 2018. Molecular dynamics simulations and novel drug discovery. Expert opinion on drug discovery, 13(1), pp.23-37.

[50] Dzeja, P., & Terzic, A.,2009. Adenylate kinase and AMP signaling networks: metabolic monitoring, signal communication and body energy sensing. International journal of molecular sciences, 10(4), 1729–1772.

[51] Whitford PC, Miyashita O, Levy Y, Onuchic JN.,2007. Conformational transitions of adenylate kinase: switching by cracking. J Mol Biol. Mar 9;366(5):1661-71.

[52] Krishnamurthy H, Lou H, Kimple A, Vieille C, Cukier RI,2005. Associative mechanism for phosphoryl transfer: a molecular dynamics simulation of Escherichia coli adenylate kinase complexed with its substrates. Proteins. pp. 1;58(1):88-100.

[53] Wang Y, Wei DQ, Wang JF.,2010. Molecular dynamics studies on T1 lipase: insight into a double-flap mechanism. J Chem Inf Model. pp 24;50(5):875-8.

[54] Bahar, I., Chennubhotla, C. and Tobi, D., 2007. Intrinsic dynamics of enzymes in the unbound state and relation to allosteric regulation. Current opinion in structural biology, 17(6), pp.633-640.

[55] Arora, K., and Brooks, C.L., 2007. Large-scale allosteric conformational transitions of adenylate kinase appear to involve a population-shift mechanism. Proceedings of the National Academy of Sciences, 104(47), pp.18496-18501.

[56] Hosseini SM, Asgari V, Hajian M, Nasr-Esfahani MH,2015. Cytoplasmic, rather than nuclear-DNA, insufficiencies as the major cause of poor competence of vitrified oocytes. Reprod Biomed Online.

[57] Ionescu MI,2019. Adenylate Kinase: A Ubiquitous Enzyme Correlated with Medical Conditions. Protein J. Apr;38(2):120-133.

[58] Tikhonova, I. G., Selvam, B., Ivetac, A., Wereszczynski, J., & McCammon, J. A. (2013). Simulations of biased agonists in the β(2) adrenergic receptor with accelerated molecular dynamics. Biochemistry, 52(33).

# Appendix

## A GROMACS, MPI set up and VMD tutorial

GROMACS is mostly supplied as source code, although it may also be installed as a Linux RPM package. On the Project's homepage ([www.gromacs.org](http://www.gromacs.org)) can found all the necessary information for the installation of the application and additional on-line resources.

The connection to the supercomputer cluster can be made using SSH, HTTP and FTP where data encryption is ensured by adding the SSL/TLS layer and from Windows or Linux, using the steps described below:

Using a Linux Machine

When connecting from a machine using Linux, a command console will be opened and the following line will be executed:

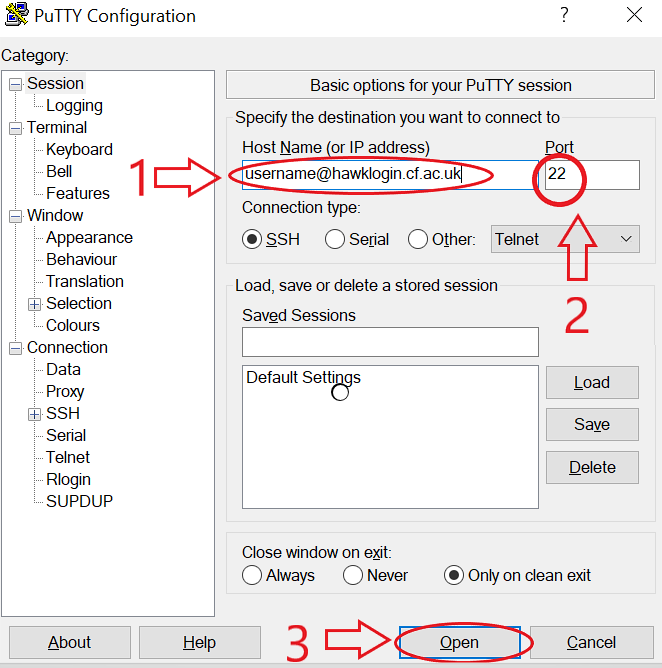
$ ssh <username>@hawklogin.cf.ac.uk -p 22

Once the password belonging to the user has been entered a work session will be established.

Using a Windows Machine

When connecting from a machine with a Windows operating system, Putty tool can be used and is available on the web site: http://www.chiark.greenend.org.uk/~sgtatham/putty

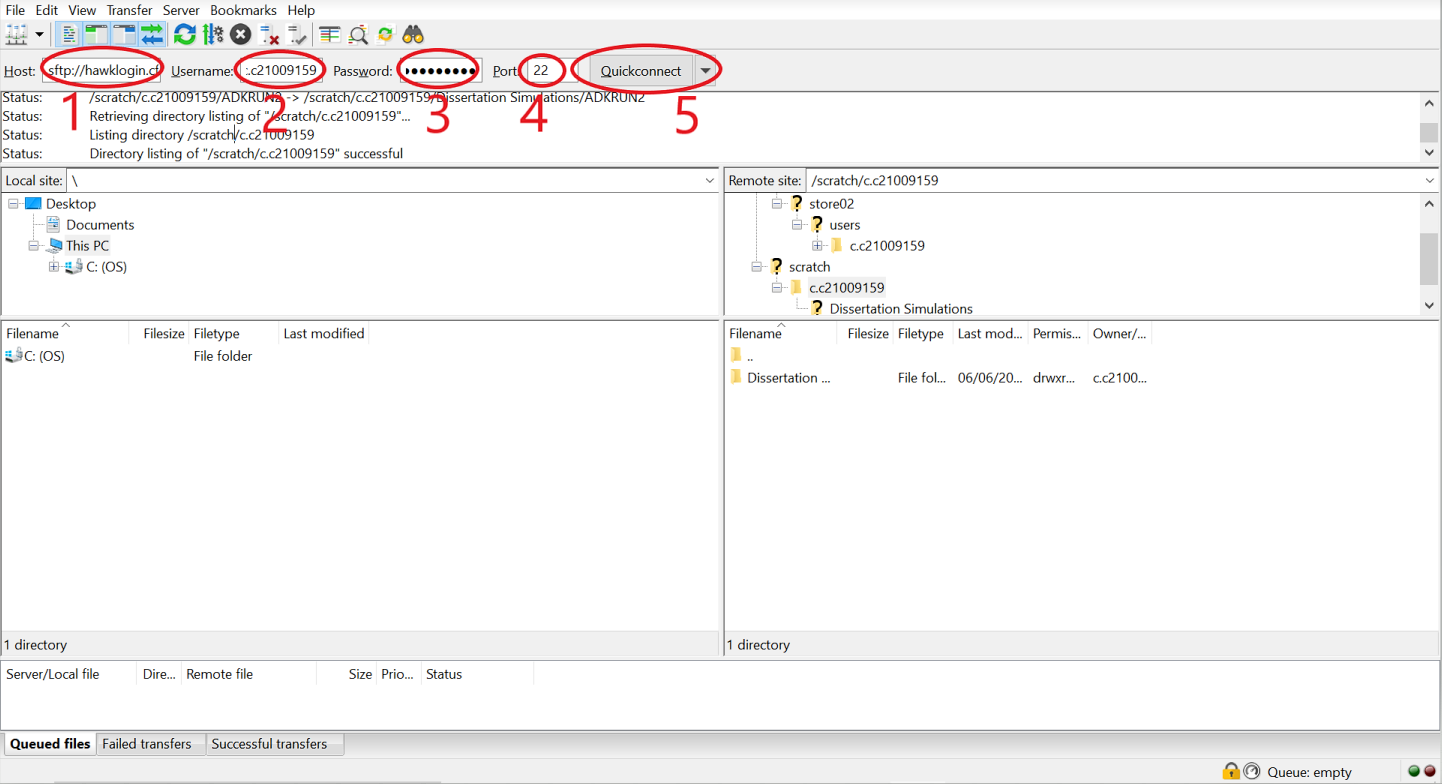
For connecting to ARCAA and Hawk it is necessary to use: [username@hawklogin.cf.ac.uk](mailto:username@hawklogin.cf.ac.uk) as a Host Name and then select port 22 (and fill in the fields as shown in figure 1).For any other supercomputer clusters the settings will be different. Once the user fields and password have been entered, the clustered partition will be in use.



**Figure (12):** Instructions for using Putty tool to get connected to MySCW supercomputer cluster.

The file transfer must always be done from the computer that is connected to the cluster, as the communications from the computer that is connected to the cluster, as communications from the server to remote computers are blocked for security reasons. To access via FTP, we have use Filezilla which is available on the web site: https://filezilla-project.org/

First enter the domain name login username@hawklogin.cf.ac.uk, followed by the username and password and pressing the enter key will send the required files to the cluster files to the cluster. From Figure (2) we can see the necessary steps, left side of the file explorer correspond to the local computer files and the right side with the MySCW. Once the connection has been stabilised is recommend creating a new directory side on MySCW (right side) that allow to run the simulations.

**Figure (13):** Instructions for using FileZilla tool to transfer files with MySCW supercomputer cluster.

Protein Visualization: VMD software

VMD is developed for the modelling, visualization, and analysis of biological systems such as proteins, nucleic acids, and molecules, and is meant to be utilized for protein visualization throughout the project. It may be used to browse the Protein Data Bank's (PDB) database of structures (PDB). It also provides a number of ways to represent and color a molecule. (Theoretical and Computational Biophysics Computational Biophysics Group, 2006). It is a open software available to download in: https://www.ks.uiuc.edu/Development/Download/download.cgi?PackageName=VMD.



**Figure (14):** Sample of the three windows that allow the use of VMD.

Figure (11) from section 3.3.5, showing different graphically AK conformations has been created using VMD. NewCartoon display options have been used to drawn ATP, ADP, and AMP molecules. We have also selected different colours to better distinguish the structure of the enzyme: BLUE for the CORE domain, GREEN for the LID domain and RED for the NMP domain.

## B Compilation of GROMACS abbreviations

Following table describe GROMACS files extensions and abbreviations to provide a better understanding in the MD simulation process:

|  |  |  |
| --- | --- | --- |
| **Abbreviation** | **Full file name** | **Description** |
| .mdp | Molecular Dynamics Parameters file | File that contains all information about the Molecular Dynamics simulation itself e.g. time-step, number of steps, temperature, pressure etc |
| .pdb | Protein Data Bank file | File with textual format describing the three-dimensional structures of molecules |
| .top | Molecular Topology file | File that contains a complete topology description of all the interactions in the molecule |
| .gro | Molecular Structure GROMACS file | File that contains a structure file as a pdb file but it can also hold velocities and it is readable with GROMACS. |
| .sh | Shell Script file | File that contains code for various scripts that can be used in the Unix Bash command interpreter. |
| .tpr | Portable Binary Run Input file | File that contains the starting structure of your simulation, the molecular topology and all the simulation parameters |
| .trr | Portable Binary Trayectory file | File that contains all the coordinates, velocities, forces and energies printed according to an MDP file and is used for simulating a trajectories in GROMACS. |
| .edr | Portable Binary Energy file | File that contains the different types of energies of the molecule using the xdr protocol |
| .log | Log file | File that contains of all events or actions affecting a particular process |
| .cpt | Portable Checkpoint file | File that  contains the complete state of the, including extended thermostat variables, random number states and NMR time averaged data |
| .xvg | Simple xmgrace file | File that contains simple multi-column data files that adds several methods to access the data for a quickly plot it. |
| .itp | Include Topology file | File that contains the topology corresponding to a single molecule type |
| .ndx | Index file | File that contains a definable set of atoms |
| .xtc | Compressed trayectoy file | File that contains the trajectories of the molecular dynamics GROMACS simulation |

**Table (2):** Taxonomy of files used for the MD simulations. Source: ([www.gromacs.org](http://www.gromacs.org))

## C Initial Files and generated files through the simulation

Following table compiles and describes all initial files and generated files through the MD simulation process:

|  |  |
| --- | --- |
| **INITIAL FILES** | |
| **Name** | **File Description** |
| ions.mdp | Parameter set up used for adding ions in step ­­­3 |
| md.mdp | Parameter up used for the MD simulation in step ­­­7 |
| em.mdp | Parameter set up used for the energy minimization used in step ­­­4 |
| npt.mdp | Parameter set up used for the NPT equilibration used in step ­­­6 |
| nvt.mdp | Parameter set up used for the NVT equilibration used in step ­­­6 |
| 4ake.pdb | Structure of Unligated AK |
| 4ake\_ADP-ADP\_Mg | Structure of ADP-DP ligated to AK with Mg2+ ions |
| 4ake\_ATP-AMP\_Mg | Structure of ATP-AMP ligated to AK with Mg2+ ions |
| 4ake\_AMP-ADP\_Mg | Structure of AMP-ADP ligated to AK with Mg2+ ions |
| 4ake\_ADP-ADP\_no\_Mg | Structure of ADP-ADP ligated to AK with Mg2+ ions |
| 4ake\_ATP-AMP\_no\_Mg | Structure of ATP-AMP ligated to AK with Mg2+ ions |
| 4ake\_AMP-ADP\_no\_Mg | Structure of AMP-ADP ligated to AK with Mg2+ ions |
| run.sh | Batch File required to start the whole MD simulation process. |
| charmm36-feb2021.ff | Charm Force Field files |
| **FILES GENERATED** | |
| **Name** | **File Description** |
| XXXX\_clean.pdb | Protein structure cleaned from water molecules |
| posre.itp | File that contains Information used to restrain the positions of heavy atoms |
| topol.top | Topology of the protein |
| conf.gro | GROMOS file that contains all the atoms defined within the force field |
| #conf.gro.1# | GROMOS file that contains the protein counterdeed in a cubic box (-c), |
| #topol.top.1# | Topology of the protein modified by taking solvent into account |
| solv.gro | GROMOS file with the configuration of the solvent (-cs). |
| ions.tpr | File used for adding ions into the protein |
| ions.gro | GROMOS file with the description of the system with the ions added. |
| mdout.mdp | Parameter file containing information por the MD simulation. |
| em.tpr | Necessary GROMACS readable file to run energy minimization |
| em.log | ASCII-text log file of the EM process |
| em.edr | Binary energy file from the EM process |
| em.gro | Energy-minimized structure |
| em.trr | Binary full-precision trajectory |
| nvt.tpr | Necessary GROMACS readable file to run NVT equilibration |
| nvt.log | ASCII-text log file of the NVT equilibration process |
| nvt.cpt | Necessary GROMACS readable file to run NVT equilibration |
| nvt.gro | Binary energy file of the NVT equilibration process |
| nvt.edr | Energy-minimized structure of the NVT equilibration process |
| nvt.trr | Binary full-precision trajectory from the NVT equilibration process |
| npt.tpr | Necessary GROMACS readable file to run NPT equilibration |
| npt.log | ASCII-text log file of the NPT equilibration process |
| npt.edr | Binary energy file of the NPT equilibration process |
| npt.trr | Energy-minimized structure of the NPT equilibration process |
| npt.gro | Binary full-precision trajectory from the NPT equilibration process |
| npt.cpt | Necessary GROMACS readable file to run NPT equilibration |
| md.tpr | Necessary GROMACS readable file to run trajectory simulation |
| md.log | ASCII-text log file of the trajectory |
| md.edr | Trajectory file containing positions of heavy atoms |
| md.xtc | Final trajectory file |
| md.cpt | Necessary GROMACS readable file to run trajectory simulation |

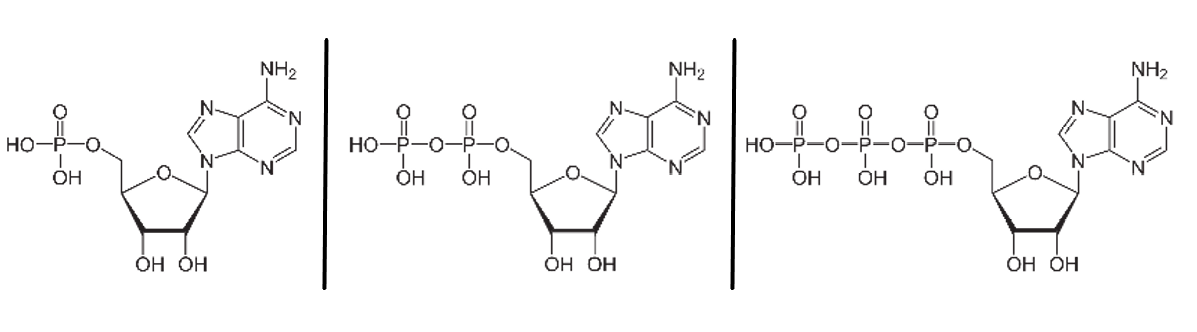
**Table (3):** Files used and generated through the simulation. Initial files Google Drive Source: https://drive.google.com/file/d/1pd5YP8HxJa7LH6flVvXv3hlyI5rfowyn/view?usp=sharing

## D ATP, ADP and AMP atom list used for generating topology files

AK topology file has been obtained using X-RAY crystallography technique from bacteria E. coli, pd4b.ake for open AK conformation and p1d.ake for close AK conformation. (Berry et al,1994). For attaching ligands ATP, ADP, and AMP molecules into the close conformation we have manually include the atom coordinates list using a text editor. This procedure has been done 3 times without for the simulation without attaching Mg2+ ions and another three times attaching Mg2+ (Special Thanks to Dr Georgina Mendez):

|  |  |  |
| --- | --- | --- |
| [ ATP ]  [ atoms ]   1. C4' CN7 0.160 0 2. H4' HN7 0.090 1 3. O4' ON6B -0.500 2 4. C1' CN7B 0.160 3 5. H1' HN7 0.090 4 6. C5 CN5 0.280 5 7. N7 NN4 -0.710 6 8. C8 CN4 0.340 7 9. H8 HN3 0.120 8 10. N9 NN2 -0.050 9 11. N1 NN3A -0.740 10 12. C2 CN4 0.500 11 13. H2 HN3 0.130 12 14. N3 NN3A -0.750 13 15. C4 CN5 0.430 14 16. C6 CN2 0.460 15 17. N6 NN1 -0.770 16 18. H61 HN1 0.380 17 19. H62 HN1 0.380 18 20. C2' CN7B 0.140 19 21. H2'' HN7 0.090 20 22. O2' ON5 -0.660 21 23. H2' HN5 0.430 22 24. C3' CN7 0.140 23 25. H3' HN7 0.090 24 26. O3' ON5 -0.660 25 27. H3T HN5 0.430 26 28. C5' CN8B -0.080 27 29. H5' HN8 0.090 28 30. H5'' HN8 0.090 29 31. O5' ON2 -0.620 30 | [ ADP ]  [ atoms ]  C4' CN7 0.160 0  H4' HN7 0.090 1  O4' ON6B -0.500 2  C1' CN7B 0.160 3  H1' HN7 0.090 4  C5 CN5 0.280 5  N7 NN4 -0.710 6  C8 CN4 0.340 7  H8 HN3 0.120 8  N9 NN2 -0.050 9  N1 NN3A -0.740 10  C2 CN4 0.500 11  H2 HN3 0.130 12  N3 NN3A -0.750 13  C4 CN5 0.430 14  C6 CN2 0.460 15  N6 NN1 -0.770 16  H61 HN1 0.380 17  H62 HN1 0.380 18  C2' CN7B 0.140 19  H2'' HN7 0.090 20  O2' ON5 -0.660 21  H2' HN5 0.430 22  C3' CN7 0.140 23  H3' HN7 0.090 24  O3' ON5 -0.660 25  H3T HN5 0.430 26  C5' CN8B -0.080 27  H5' HN8 0.090 28  H5'' HN8 0.090 29  O5' ON2 -0.620 30 | [ AMP ]  [ atoms ]  C4' CN7 0.160 0  H4' HN7 0.090 1  O4' ON6B -0.500 2  C1' CN7B 0.160 3  H1' HN7 0.090 4  C5 CN5 0.280 5  N7 NN4 -0.710 6  C8 CN4 0.340 7  H8 HN3 0.120 8  N9 NN2 -0.050 9  N1 NN3A -0.740 10  C2 CN4 0.500 11  H2 HN3 0.130 12  N3 NN3A -0.750 13  C4 CN5 0.430 14  C6 CN2 0.460 15  N6 NN1 -0.770 16  H61 HN1 0.380 17  H62 HN1 0.380 18  C2' CN7B 0.140 19  H2'' HN7 0.090 20  O2' ON5 -0.660 21  H2' HN5 0.430 22  C3' CN7 0.140 23  H3' HN7 0.090 24  O3' ON5 -0.660 25  H3T HN5 0.430 26  C5' CN8B -0.180 27  H5' HN8 0.090 28  H5'' HN8 0.090 29  O5' ON2 -0.400 30 |

**Table (4):** Atom list for ATP, ADP and AMP molecules including charge and atom description (created from scratch using the file merged.rtp).



**Figure (15):** Ligands chemical diagrams: a) AMP b) ADP c) ATP.

## E Code & Commands tutorial to facilitate methodology replication

Following table compiles and explain all the necessary commands to perform methodological process explained in section 3.2 commands. It divides the process in seven steps and provides a description for each command aim, together with input and output files generated:

|  |  |
| --- | --- |
| STEP 1: Topology Generation | |
| **Command:** | **grep -v HOH XXXX.pdb > XXXX\_clean.pdb** |
| In files: | XXXX.pdb |
| Out files | XXXX\_clean.pdb |
| Description: | Strip waters molecules from the initial pdb file. This function creates a new pdb file with the water molecules removed. XXXX should be replaced for the corresponding pdb file name. |
| **Command:** | **gmx\_mpi pdb2gmx -f XXXX.pdb** |
| In files: | XXXX\_clean.pdb |
| Out files | conf.gro, topol.top, posre.itp |
| Description: | This function post process the protein and generates the 3 most important files that are necessary for GROMACS to work. The topology for the molecule, A position restraint file and a A post-processed structure file .We have selected Select the Force Field: 1 (AMBER03 protein) Select the Water Model: 1 (TIP3P) |
| STEP 2: Defining the unit cell & adding solvent | |
| **Command:** | **gmx\_mpi editconf -f conf.gro -o conf.gro -c -d 1.0 -bt cubic** |
| In files: | conf.gro |
| Out files | #conf.gro.1# |
| Description: | creates new file in which the protein is centered within the protein in the box (-c), and placed at least 1.0 nm from the box edge (-d 1.0). The box type is defined as a cube (-bt cubic). |
| **Command:** | **gmx\_mpi solvate -cp conf.gro -cs spc216.gro -o solv.gro -p topol.top** |
| In files: | #conf.gro.1# |
| Out files | #topol.top.1# solv.gro |
| Description: | Solvates the protein into water and writes a new topology to reflect the changes that have been made. TIP3P water has been used as the main water model |
| STEP 3: ADDING ions to neutralize electric charge of the protein | |
| **Command:** | **gmx\_mpi grompp -f ions.mdp -c solv.gro -p topol.top -o ions.tpr** |
| In files: | ions.mdp, solv.gro, #topol.top.1# |
| Out files | ions.tpr mdout. |
| Description: | Now we have an atomic-level description of our system in the binary file ions.tpr |
| **Command:** | **gmx\_mpi genion -s ions.tpr -o ions.gro -p topol.top -np 4** |
| In files: | Ions.tpr, #topol.top.1# |
| Out files | ions.gro #topol.top.2# |
| Description: | Adds only the ions necessary to neutralize the net charge on the protein generating a new topology file.We have selected group 13 "SOL" for embedding ions. |
| STEP 4: Enery Minimization | |
| **Command:** | **gmx\_mpi grompp -f em.mdp -c ions.gro -p topol.top -o em.tpr o este o** |
| In files: | ions.groem.mdp, #topol.top.2# |
| Out files | em.tpr |
| Description: | Initial relaxtions and setting up by creating a set of parameters in em-tpr file that are necessary for MD relaxation |
| **Command:** | **gmx\_mpi mdrun -deffnm em -v** |
| In files: | em.tpr |
| Out files | em.log, em.edr, em.trr, em.gro |
| Description: | It carries on the energy minimization. |
| STEP 5 EQUILIBRATION | |
| **Command:** | **gmx\_mpi grompp -f nvt.mdp -c em.gro -r em.gro -p topol.top -o nvt.tpr** |
| In files: | em.gro , nvt.mdp, #topol.top.2# |
| Out files | nvt.tpr |
| Description: | Creates a tpr file to get an atomoic description need for NVT equilibration |
| **Command:** | **gmx\_mpi mdrun -deffnm nvt -v** |
| In files: | nvt.tpr |
| Out files | nvt.log, nvt.edr, nvt.trr, nvt.gro nvt.cpt |
| Description: | Executes NVT equilibration |
| **Command:** | **gmx\_mpi grompp -f npt.mdp -c nvt.gro -r nvt.gro -p topol.top -o npt.tpr** |
| In files: | nvt.gro , npt.mdp , #topol.top.2# |
| Out files | npt.tpr |
| Description: | Creates a tpr file to get an atomoic description need for NPT equilibration |
| **Command:** | **gmx\_mpi mdrun -deffnm npt -v** |
| In files: | npt.tpr |
| Out files | npt.log , npt.edr, npt.trr, npt.gro, npt.cpt |
| Description: | Executes NPT equilibration |
| STEP 6: MOLECULAR DYNAMICS | |
| **Command:** | **gmx\_mpi grompp -f md.mdp -c npt.gro -r npt.gro -p topol.top -o md.tpr** |
| In files: | npt.gro , #topol.top.2# , md.mdp |
| Out files | md.tpr |
| Description: | Creates a tpr file to get an atomoic description needed for MD production simulation |
| **Command:** | **gmx\_mpi mdrun -deffnm md** |
| In files: | md.tpr |
| Out files | md.log, md.edr, md.xtc, md\_prev.cpt , md.cpt |
| Description: | Executes molecular dynamics simulation |
| STEP 7: RESULTS ANALYSYS | |
| **Command:** | **gmx\_mpi energy -f em.edr -o potential.xvg** |
| In files: | em.edr |
| Out files | potential.xvg­­­ |
| Description: | Aarage Potential Energy in KJ.mol^-1 for each time step during energy minimization.We have choose potential (10) and (0) |
| **Command:** | **gmx\_mpi energy -f nvt.edr -o temperature.xvg** |
| In files: | nvt.edr |
| Out files | temperature.xvg |
| Description: | Aarage Tempreature in K for each time step during NVT equilibration. We have chosen potential (16) and (0) |
| **Command:** | **gmx\_mpi energy -f npt.edr -o pressure.xvg** |
| In files: | npt.edr |
| Out files | pressure.xvg |
| Description: | Aarage pressure in bar for each time step during NPT equilibration. We have chosen potential (18) and (0) |
| **Command:** | **gmx\_mpi energy -f npt.edr -o density.xvg** |
| In files: | npt.edr |
| Out files | density.xvg |
| Description: | Aarage density in kg\*m^3 for each time step during NPT equilibration. We have chosen potential (24) and (0) |
| **Command:** | **gmx\_mpi trjconv -s md.tpr -f md.xtc -o mdnoPBC.xtc -pbc mol -center** |
| In files: | md.tpr, mt.xtc |
| Out files | mdnoPBC.xtc |
| Description: | post-processing command to strip out coordinates, correct for periodicity Correct the trajectory for posterior analysys. We have Selected 1 ("Protein") as the group to be centered and 0 ("System") for output. |
| **Command:** | **gmx\_mpi rms -s md.tpr -f mdnoPBC.xtc -o rmsd.xvg -tu ns** |
| In files: | mdnoPBC.xtc, md.tpr |
| Out files | rmsd.xvg |
| Description: | Performs RMSD calculation. We have choose (4) Backbone for both the least-squares fit and the group |
| **Command:** | **gmx\_mpi rms -s em.tpr -f mdnoPBC.xtc -o rmsd\_xtal.xvg -tu ns** |
| In files: | em.tpr, mdnoPBC.xtc |
| Out files | rmsd\_xtal.xvg |
| Description: | Performs RMSD relative to the crystal structure calculation. We have choose (4) Backbone for both the least-squares fit and the group |
| **Command:** | **gmx\_mpi gyrate -s md.tpr -f mdnoPBC.xtc -o gyrate.xvg** |
| In files: | md.tpr , mdnoPBC.xtc |
| Out files | gyrate.xvg |
| Description: | Calculates The radius of gyration We have Choosen group 1 (Protein) for analysis. |

**Table (5):** Overview of the GROMACS process used in the dissertation. Input/Output files with the commands and the description for each piece of GROMACS code.

## F Batch files and Python scripts

Batch file can be used to execute a set of commands in an automatically manner, instead of manually execute each command from table (5). Batch file should be placed in the working directory and can be executed in the GROMACS command terminal by typing the following command:

-sh run /batch file

To generate batch file following code should be copy and pasted into a text editor and save the file with. batch extension:

|  |
| --- |
| #!/bin/bash  #SBATCH -o 4ake.o.%J # Job output file  #SBATCH -e 4ake.e.%J # Job error file  #SBATCH -J 4ake  #SBATCH --ntasks=80  #SBATCH --ntasks-per-node=40  #SBATCH -p compute  #SBATCH --time=72:00:00  #SBATCH --account=scw1702  # #Usage  # sbatch script # submit job  # squeue # job status  module purge  # Load the gromacs module - gromacs/4.6.1-mpi\_d - double precision  module load gromacs  # following variable is default on Hawk  export OMP\_NUM\_THREADS=1  top\_dir=/app/chemistry/gromacs/4.6.1/sb/intel-13.0/intel-4.1/examples  # Settings & directory locations for Lab\_Worksheet\_Gromacs  export my\_dir=$SLURM\_SUBMIT\_DIR  MYLOGS=$my\_dir  NNODES=$SLURM\_NNODES  NCPUS=$SLURM\_NTASKS  PPN=$SLURM\_NTASKS\_PER\_NODE  ###################################  # WORKDIR #  ###################################  # run the simulation on /scratch - change this directory as appropriate  export work\_dir=/scratch/c.spxjem/AK/Run2  # ... and change into this directory  cd $work\_dir  # Run the MD simulation  start="$(date +%s)"  #srun gmx\_mpi mdrun  stop="$(date +%s)"  finish=$(( $stop-$start ))  echo Gromacs/4.6.1-mpi\_d ${case\_name} $SLURM\_JOBID Job-Time $finish seconds  echo Gromacs/4.6.1-mpi\_d ${case\_name} End Time is `date` |

**Table (6):** GROMACS Dissertation Code in batch file format.

To analysis results batch file must been executed and all output files (APPENDIX C) generated. Using this files graphs from section 3.3.1-3.3.4 can be created using the following Python script:

|  |
| --- |
| from MDAnalysis.analysis.rms import rmsd  u = mda.Universe("npt.gro", "md.xtc")  print(u)  u = mda.Universe("npt.gro", "md.xtc")  print(u)  from numpy.linalg import norm  def theta\_NMP(u):      """Calculate the NMP-CORE angle for E. coli AK in degrees"""      C = u.select\_atoms("resid 115-125 and (backbone or name CB)").center\_of\_geometry()      B = u.select\_atoms("resid 90-100 and (backbone or name CB)").center\_of\_geometry()      A = u.select\_atoms("resid 35-55 and (backbone or name CB)").center\_of\_geometry()      BA = A - B      BC = C - B      theta = np.arccos(np.dot(BA, BC)/(norm(BA)\*norm(BC)))      return np.rad2deg(theta)  def theta\_LID(u):      """Calculate the LID-CORE angle for E. coli AK in degrees"""      C = u.select\_atoms("resid 179-185 and (backbone or name CB)").center\_of\_geometry()      B = u.select\_atoms("resid 115-125 and (backbone or name CB)").center\_of\_geometry()      A = u.select\_atoms("resid 125-153 and (backbone or name CB)").center\_of\_geometry()      BA = A - B      BC = C - B      theta = np.arccos(np.dot(BA, BC)/(norm(BA)\*norm(BC)))      return np.rad2deg(theta)  def distance(u):      """Calculate the LID-NMP domain  for E. coli AK  in amstrongs"""      #C = u.select\_atoms("resid 179-185 and (backbone or name CB)").center\_of\_geometry()      #B = u.select\_atoms("resid 115-125 and (backbone or name CB)").center\_of\_geometry()      #A = u.select\_atoms("resid 125-153 and (backbone or name CB)").center\_of\_geometry()      #AC=C-A      #distance= norm(AC)      cm = dict((name, dom.center\_of\_mass()) for name,dom in domains.items())      distance=(norm(cm['NMP'] - cm['LID']))      return distance      from matplotlib.ticker import StrMethodFormatter  time, NMP,distance, LID = data.T  # plotting  fig, axs = plt.subplots(5, sharex=True, sharey=False,figsize=(23,18))  fig.subplots\_adjust(wspace=0, hspace=0.04)  fig.suptitle("MD simulations of AK without ligands (Open Initial Conformation)", fontsize=22,y=0.92)  axs[0].plot(time, LID, 'r-', lw=1, color='green',label=r"$\theta\_{\mathrm{LID}}$")  axs[0].set\_ylabel(r"LID angle $\theta(º)$",fontsize=19)  #axs[0].legend(loc="upper right",fontsize=17)  axs[0].yaxis.set\_ticks\_position('both')  axs[0].tick\_params(labeltop=False, labelright=True)  axs[0].yaxis.set\_major\_formatter(StrMethodFormatter('{x:,.1f}'))  axs[1].plot(time, NMP, 'b-', lw=1, color='blue' ,label=r"$\theta\_{\mathrm{ NMP}}$")  axs[1].set\_ylabel(r"NMP angle $\theta(º) $",fontsize=19)  #axs[1].legend(loc="upper right",fontsize=17)  axs[1].yaxis.set\_ticks\_position('both')  axs[1].tick\_params(labeltop=False, labelright=True)  axs[1].yaxis.set\_major\_formatter(StrMethodFormatter('{x:,.1f}'))  axs[2].plot(time[1:], distance[1:], 'k-', lw=1, label='Distance  LID-NMP')  axs[2].set\_ylabel(r"Distance (Aª) ",fontsize=19)  #axs[2].legend(loc="upper right",fontsize=17)  axs[2].yaxis.set\_ticks\_position('both')  axs[2].tick\_params(labeltop=False, labelright=True)  axs[2].yaxis.set\_major\_formatter(StrMethodFormatter('{x:,.1f}'))  axs[3].plot((Rgyr[:,0])[1:], (Rgyr[:,1])[1:], 'k-', lw=1, label=r"radius of gyration $R\_G$ ($\AA$)")  axs[3].set\_ylabel(r"$R\_G$ ($\AA$)",fontsize=19)  #axs[3].legend(loc="upper right",fontsize=17)  axs[3].yaxis.set\_ticks\_position('both')  axs[3].tick\_params(labeltop=False, labelright=True)  axs[3].yaxis.set\_major\_formatter(StrMethodFormatter('{x:,.1f}'))  for iframe, ts in enumerate(u.trajectory):      for name, g in domains.items():          results[name][iframe, :] = (u.trajectory.time,                                          rmsd(g.positions, xref0[name],                                               center=True, superposition=True))    for name in "CORE", "NMP", "LID":      data1 = results[name]      axs[4].plot((data1[:, 0])[1:], (data1[:, 1])[1:], linestyle="-",                  color=colors[name], lw=2, label=name)  axs[4].legend(loc="upper right",fontsize=17)  axs[4].set\_xlabel(r"time $t$ (ps)",fontsize=19)  axs[4].set\_ylabel(r"C$\_\alpha$ RMSD ($\AA$)",fontsize=19)  axs[4].tick\_params(axis='x', labelsize=13.5)  axs[4].tick\_params(axis='y', labelsize=13.5)  axs[4].yaxis.set\_ticks\_position('both')  axs[4].tick\_params(labeltop=False, labelright=True)  axs[4].yaxis.set\_major\_formatter(StrMethodFormatter('{x:,.1f}')) |

**Table (7):** Python script used for generating all graphs from section 3.3.1-3.3.4)