# Characterization of myosin VI phosphate departure by molecular dynamics simulations

Robin Manevy

## To cite this version:

Robin Manevy. Characterization of myosin VI phosphate departure by molecular dynamics simulations. Biomechanics [physics.med-ph]. Université Gustave Eiffel, 2023. English. NNT : 2023UEFL2034 . tel-04338607

**HAL Id: tel-04338607**
**https://theses.hal.science/tel-04338607**

Submitted on 12 Dec 2023

# UNIVERSITÉ GUSTAVE EIFFEL

## Ecole Doctorale SIE 531

## THÈSE

Pour obtenir le grade de :

**Docteur de l'Université Gustave Eiffel**

Discipline : Biomécanique

Présentée par :

**Robin MANEVY**

# Caractérisation du départ du phosphate de la myosine VI par simulations de dynamique moléculaire

Soutenue le 6 Mars 2023,
devant le jury composé de :

| | | |
|---|---|---|
| S. SACQUIN-MORA | Directrice de recherche - IBPC | Rapportrice |
| R. STOTE | Directeur de recherche - IGBMC | Rapporteur, Président du jury |
| M. CECCHINI | Maître de conférences - Unistra | Examinateur |
| A. HOUDUSSE | Directrice de recherche - Institut Curie | Examinatrice |
| I. NAVIZET | Professeure des universités - Univ. Gustave Eiffel | Directrice de Thèse |
| M. CARUEL | Maître de conférences - Univ. Paris-Est Créteil | Co-encadrant |
| F. DETREZ | Maître de conférences - Univ. Gustave Eiffel | Co-encadrant |

## Abstract

Myosins are a superfamily of proteins acting as molecular motors. This motor activity takes place in a cycle fueled by ATP hydrolysis. The ATP hydrolysis produces an ADP and an inorganic phosphate $P_i$. We study the cycle of myosin, particularly the step of powerstroke, the force-generating step, and its relationship with $P_i$ release.

In this thesis we aim at studying by molecular dynamics simulations the place of $P_i$ release in the myosin cycle. Our calculations use a crystallographic state of myosin VI solved in 2015 named $P_i$-Release because it exhibits characteristics favoring $P_i$ departure. Notably, this state presents one of the main $P_i$ exit door open. As this state also shows a mostly armed converter, it constitutes an argument in favor of $P_i$ departure before the powerstroke. However, models based on results from physiological studies propose that the powerstroke precedes the departure of $P_i$ while others would give no mandatory relative timing between $P_i$ release and powerstroke. These models are incompatible with a $P_i$-Release state before the powerstroke.

We have compared the $P_i$-Release state with two states prior and after the powerstroke, called Pre-Powerstroke and Strong-ADP respectively. Several methodologies were developed to perform the comparison of the states. The characterization of the three states is based on the comparison of crystallographic structures and snapshots from relaxation molecular dynamics without actin. We use the characterization of the states to try to transform the myosin from Pre-Powerstroke conformation to Strong-ADP and check whether the $P_i$-Release state is visited along the transition.

In order to study specifically the departure of $P_i$, we performed Umbrella Sampling simulations forcing the departure of $P_i$. These simulations were started from relaxed structures in the Pre-Powerstroke and $P_i$-Release states. These relaxed structures were obtained at the end of relaxation started with different orientation of protonated $P_i$ in order to study the impact of the orientation of $P_i$ on its departure.

Moreover, relaxation of Pre-Powerstroke and $P_i$-Release without $P_i$ were realized to study the impact of the absence of $P_i$ in the structure.

This thesis gives some elements to better understand the $P_i$-Release state and the departure of $P_i$.

# Résumé

Les myosines constituent une superfamille de protéines agissant comme moteur moléculaire au sein d'un cycle. Ce cycle est alimenté par l'hydrolyse de l'ATP comme source d'énergie, ce qui donne un ADP et un phosphate inorganique (Pi). Nous nous sommes intéressés au cycle de la myosine et particulièrement à l'étape concernant la production de force ainsi que sa relation au départ du phosphate.

Nous proposons dans cette thèse d'étudier par des moyens de simulations de dynamique moléculaire la place du départ du phosphate dans le cycle de la myosine. Nous utilisons notamment une structure cristallographique d'un nouvel état de la myosine VI résolue en 2015 et nommée $P_i$-Release du fait que certaines de ses caractéristiques laissent penser que cet état permet le départ du phosphate. Notamment, cette structure présente une des principales portes de sorties du phosphate ouverte. Comme ce nouvel état de la protéine a son convertisseur qui est encore pratiquement complètement "armé", il constitue un élément en faveur d'un départ du phosphate avant le "powerstroke".

Cependant des modèles basés sur des résultats d'études physiologiques tendent à décrire le départ du phosphate après le "powerstroke" quand d'autres ne donnent aucun lien entre départ du phosphate et powerstroke. Ces modèles ne sont pas en accord avec un état $P_i$-Release avant le powerstroke. Pour cela nous avons comparé l'état $P_i$-Release par rapport à deux états de la myosine avant et après le début du powerstroke : les états Pre-powerstroke et Strong-ADP respectivement. Plusieurs méthodologies ont été développées pour effectuer cette comparaison. La caractérisation des trois états se base sur la comparaison des structures cristallographiques et celles issues de relaxations tout atome en l'absence de l'actine.

Nous utilisons cette caractérisation pour essayer de transformer la myosine de l'état Pre-powerstroke à Strong-ADP afin de voir si l'on visitait l'état $P_i$-Release lors de la transition.

Afin d'étudier spécifiquement le départ du $P_i$, nous avons effectué des simulations d'Umbrella Sampling en forçant le départ du $P_i$ en partant de plusieurs structures relaxées de Pre-powerstroke et $P_i$-Release. Ces structures relaxées ont été choisies avec des orientations du $P_i$ protoné différentes pour étudier l'influence de l'orientation du $P_i$ sur son départ.

De plus, des relaxations spécifiques de Pre-powerstroke et $P_i$-Release sans $P_i$ ont été réalisées pour comprendre quel est l'impact de la présence du $P_i$ dans le site actif.

Cette thèse permet d'avoir des éléments pour mieux comprendre l'état $P_i$-Release de la myosine VI et comprendre le départ du $P_i$.

# Contents

# Introduction

In this introduction we present the context of our work and the subject of this thesis. We then summarize our methodology and list our main results.

## Contents

## Muscle contraction

### Anatomy: From the tissue to the actin-myosin interaction

Muscles are contractile tissue that can produce mechanical work. They are involved in essential physiological processes such as locomotion, blood circulation, breathing or digestion. Muscle is a multiscale material, see Fig. 1. The tissue is made of fascicles comprising fibrilar cells called myocytes or muscle fibers. The cytoplasm of the myocytes is essentially made of myofibrils. These consist in a chain of contracile units called sarcomeres.

Sarcomeres are mostly composed of two filaments: the actin (thin) filament, and the myosin (thick) filament. The thin filament is composed of an helical repetition of actin protein monomers. The thick filament is made out of myosin proteins attached by their tails all rolled on each other. Each myosin protein contains two globular domains (heads) that can interact with actin.

The sliding of the actin filament relative to the myosin filament is produced by the cyclical binding of these heads to the actin filament. The swinging of the myosin heads, when these latter are attached on actin, acts as a working stroke that pulls the actin filament toward the center of the sarcomere. The action of the multiple myosin heads produces the shortening of the sarcomere. The actin-myosin interaction is the fundamental force generation process. It is kept out of equilibrium by the enzymatic activity of the myosin protein.

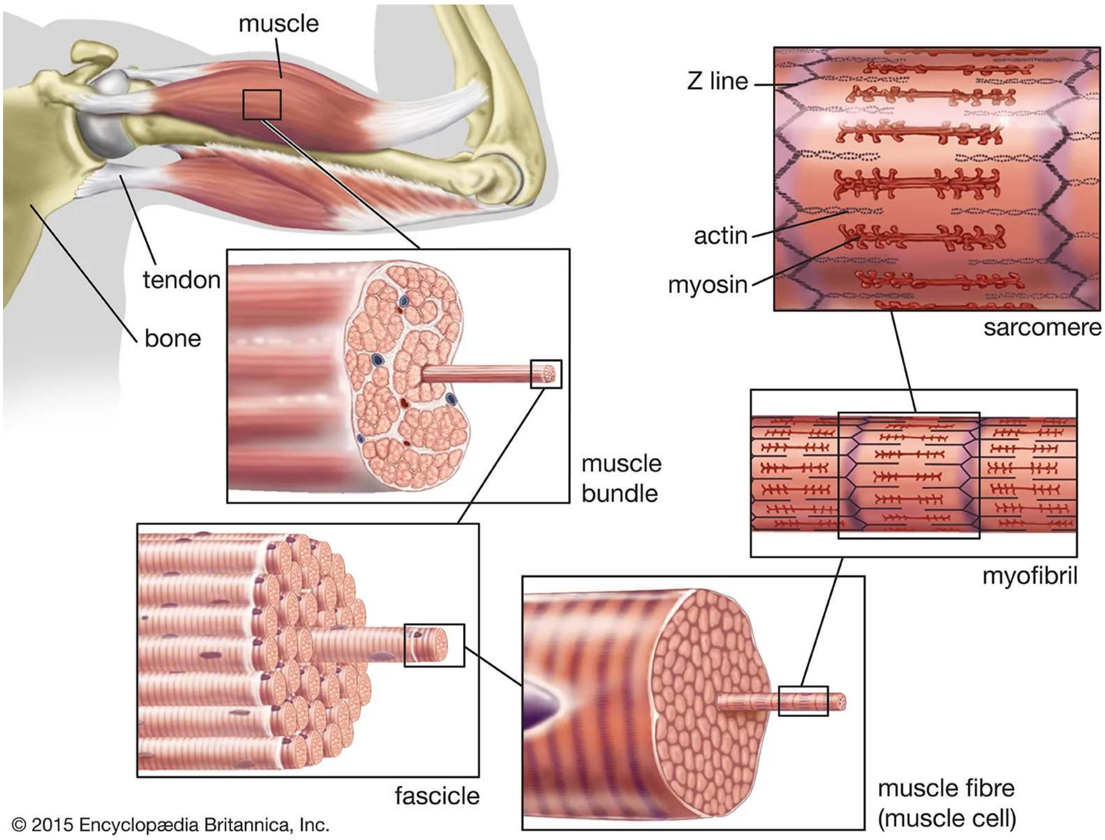


Figure 1: Figure showing the different scales of muscles. Adapted from (Gergely et al., 2022).

## ATP Hydrolysis: energy source of the actin-myosin cycle

The energy used for the functioning of numerous proteins such as myosin directly comes from the hydrolysis of Adenosine Tri-Phopshate (ATP), whose production relies on other primary sources of energy such as carbohydrates or fatty acids. Since ATP is permanently produced from multiple primary sources, it can thus be viewed as a vector of energy, like electricity. The ATP molecule is composed of an Adenosine and 3 phosphates. The bonds between these phosphates store each $\sim 100\,\mathrm{zJ}$ of chemical energy. This correspond to around $\sim 14\,\mathrm{kcal\,mol^{-1}}$. A portion of this energy is released by the hydrolysis of ATP which consists in the cleavage (separation, breakage of bond) of the terminal phosphate from the rest of the ATP with the intervention of a water molecule (see Fig. 2). The products of this reaction are an inorganic phosphate and an ADP (Adenosine Di-Phosphate). The reaction is written:

$$\mathrm{ATP} \xrightarrow{\mathrm{H_2O}} \mathrm{P_i} + \mathrm{ADP}$$

After hydrolysis, the generated $\mathrm{P_i}$ molecule remains trapped within the protein. It is its departure that is associated with the largest energy change along the myosin ATPase cycle (Geeves, 2016).

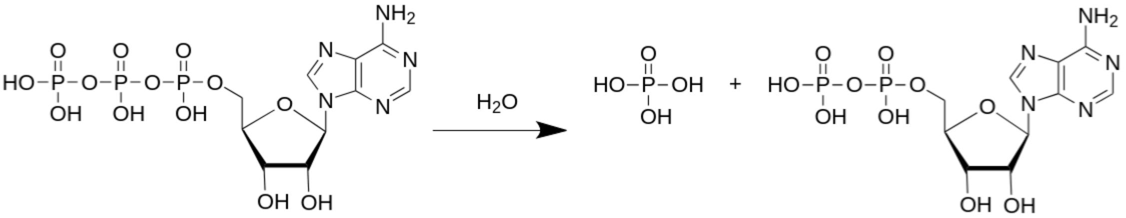


Figure 2: ATP hydrolysis. This reaction separates the terminal phosphate from the adenosine and its two other phosphate groups with the intervention of a water molecule. This reaction results in ADP and $P_i$. In this representation, all phosphates are completely protonated.

# Studied myosin structures

In this section we present the actin-myosin interaction cycle. It allows then to describe precisely the objectives of the thesis.

## Actin-myosin cycle

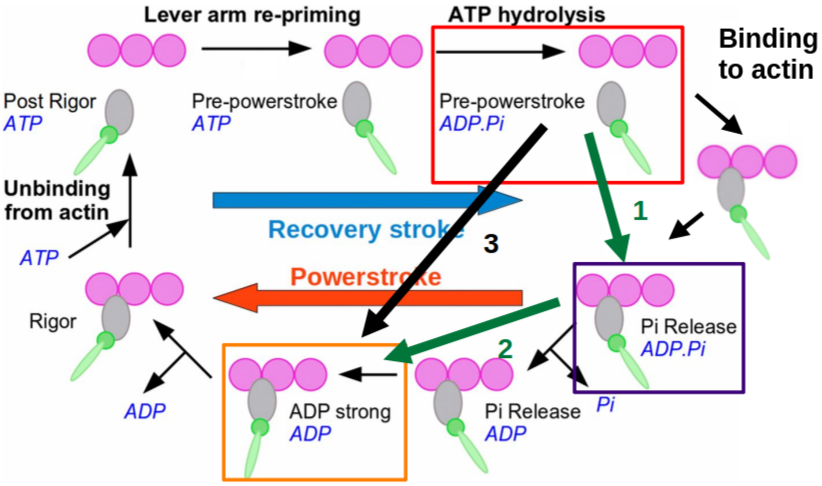


Figure 3: Representation of the catalytic cycle of Myosin interacting with actin adapted from (Blanc et al., 2018). This cycle relies on myosin structures solved by X-ray crystallography or cryo-EM methods. This cycle assumes that the state called $P_i$ Release is an intermediate after myosin binds to actin and before the powerstroke. The validity of this assumption is one of the questions to which this thesis aims at giving insights.

The ATPase activity of myosin in the presence of actin takes the form of a cycle of interaction, first identified by Lymn and Taylor in (Lymn and Taylor, 1971). A representation of the complete cycle based on structures of myosin is shown on Fig. 3.

Myosin is either attached or detached from actin, and, in each of these states, it can be either armed or disarmed. The change of the conformation by rotation of the myosin head compared to the tail (or the rocking of the tail compared to the head) lead to the two positions: armed before it produces force and disarmed after it has produced force. The terms refer to the loaded/cocked/armed gun before the shoot. When myosin is armed, it can produce force via a large conformational change called the powerstroke (PS). The PS happens when myosin is bound to actin. After the completion of the PS, the myosin is disarmed. The state is called Rigor state. In Rigor state, the myosin head is still attached to the actin filament but the products of the ATP hydrolysis have left the cavity.

The unbinding from actin is caused by the binding of an ATP. Myosin is then disarmed and detached from actin with a bound ATP. This state is called Post-Rigor. To produce force again, the myosin reverses the powerstroke while it is unbound. This reverse powerstroke is called recovery stroke and starts by a repriming of the lever arm followed by conformational changes stabilizing this reprimed position. Myosin transforms therefore into a state enabling ATP hydrolysis and consequently realises the hydrolysis and the recovery stroke. These events lead to a state called Pre-Powerstroke (PPS) with ADP and $P_i$. From this state myosin starts to interact with actin.

Following the first weak interactions with actin, myosin binds a bit more strongly to actin and transforms into a state called $P_i$-Release ($P_iR$) from which $P_i$ release can occur.

Finally myosin performs the powerstroke and visits two states in which it successively binds more strongly, called Strong-ADP and Rigor. The Strong-ADP state is a step after the loss of $P_i$ with ADP inside the cavity. In the Rigor state the myosin is nucleotide free.

# The relationship between $P_i$-Release and Powerstroke

Knowing whether $P_i$ release occurs before, along or after the powerstroke (*i.e.* the conformation change that produce mechanical force) is still an open question of primary physiological importance (Robert-Paganin et al., 2019; Debold, 2021). Understanding the interplay between $P_i$ release and the working stroke could indeed help to develop therapeutic strategies using drugs that would specifically target one of the two processes. Such treatments could serve as contraction enhancement or inhibitors depending on the disease. The question of whether the $P_i$ release or the powerstroke occurs before is thus crucial in order to define the best target for drugs.

To give insight to answer this question we have performed molecular dynamics and umbrella sampling simulations as long as developed tools to analyse the different structures. After a presentation of the different molecular structures (chapter 1), we present the results in two parts:

Part I corresponds to the study of phosphate release. A review of landmark studies on this topic is proposed in chapter 2. In chapter 3, we also studied more specifically the molecular mechanism of $P_i$ departure from myosin. In this context, we have raised a question on the orientation of $P_i$. In this study we have forced the departure of $P_i$ from the binding pocket using Umbrella Sampling simulations. We then studied the obtained trajectories. We particularly focused on the interactions formed by the $P_i$ with its environment and the state of the active site. This study allowed us to evaluate if there was a difference of behaviour of $\mathrm{P_i}$ in our simulations of PPS and $\mathrm{P_iR}$ without actin. This chapter is a draft of a paper to be submitted.

The second part aims at giving some structural insights in order to understand which of the $\mathrm{P_i}$ Release or Powerstroke happens before. By studying three states of myosin VI: PPS, $\mathrm{P_iR}$ and Strong-ADP, we want to add contributions to the question of whether $\mathrm{P_iR}$ can be an intermediate between PPS and Strong-ADP. The chosen approach is to characterize the PPS, $\mathrm{P_iR}$ and Strong-ADP states based on several relaxation molecular dynamics simulations starting from each state. These states were simulated without actin.

While these simulations were performed for PPS and $\mathrm{P_iR}$ with $\mathrm{P_i}$ in the active site, the simulations were also performed without $\mathrm{P_i}$ to distinguish the effect of the presence of $\mathrm{P_i}$ from the impact of the initial structures themselves on the dynamics of myosin (see appendix E). Some complement on the study of the $\mathrm{P_i}$ release are also put in appendix F.

We present in chapter 4 some collective variables for defining the state of the cleft and Switch II. A possible constraint to go from one structure to the other could be done by changing structural distances or angles. These collectives variables are presented in following chapters 5 and 6.

The characterization of PPS, $\mathrm{P_iR}$ and Strong-ADP was then used to define specific features of PPS and Strong-ADP. These specific features allowed us to propose collective variables that could be tested to transform the system from PPS to Strong-ADP (see Appendix D). We perform Umbrella Sampling simulations by changing the collective variable from their value observed in PPS to their value observed in Strong-ADP. The objective is to judge whether the $\mathrm{P_iR}$ state is visited along this transformation, setting it as an intermediate. This would correspond to the mechanism of arrow 1 (binding to actin and transformation into $\mathrm{P_iR}$ state) followed by arrow 2 (Departure of $\mathrm{P_i}$ in the $\mathrm{P_iR}$ state and then transformation into the Strong-ADP state) on Fig. 3. This mechanism is opposed to the one represented by arrow 3 (which should be in contrast seen as the transformation into the Strong-ADP state before $\mathrm{P_i}$ release) on Fig. 3.

Finally, the first appendix A gives the basis of the simulation presented in this thesis.

In this thesis we developed some tools for the characterization of the PPS, $\mathrm{P_iR}$ and Strong-ADP state in absence of actin. We have also given a description of the interactions of $\mathrm{P_i}$ with the residues of the protein in its pathway to leave the cavity.

# Chapter 1

# Molecular structure of Myosin

In this chapter we give a detailed description of the system we are studying in this thesis. After a general definition of the proteins, we present the protein family of actin and myosin, the two proteins forming a molecular motor by their interaction. We then present the three structures of myosin VI used in this thesis.

**Contents**

## 1.1 Molecular motors

In this section we present few basic information on the proteins. Then we briefly describe actin, forming with myosin the complete molecular motor. Subsequently we give general information on myosin, first by presenting the myosin superfamily, second by presenting the myosin structural elements.

### 1.1.1 Myosin and actin: examples of proteins

Proteins are a class of biopolymer whose monomers are amino acids (Lodish et al., 2008). Polymers are long chains of repeated unit elements. In proteins, these elements are the amino acids. The formed chains fold on themselves in specific ways to accomplish physiological functions. There exist 21 types of proteinogenic amino acids.

Proteins play fundamental roles in almost all physiological processes. For instance, the complex of two proteins — GroEL/GroES in bacteria (Ishii, 2017) or Hsp60/Hsp10 in eukaryotic cells (Gupta, 1995)— contributes to the correct folding of other proteins by confining them in a restrained space favoring compact folding. Proteins can also be transmembrane proteins like aquaporins, which help the crossing of cell membranes by water molecules. Cytoskeleton proteins notably help to maintain the shape of the cells. Actin contributes to the cytoskeleton of the cells together with microtubules and the intermediate filaments. Moreover, these proteins serve as tracks for molecular motors. The molecular motor proteins circulate on these tracks by successive attachment and detachment associated with conformational changes allowing displacement.

The molecular motors form a crucial class of proteins that use energy from the hydrolysis of ATP or GTP in order to produce work. Among the molecular motors, some representatives are the kinesin and dynein families, both using microtubule as track. Another fundamental family of molecular motors is myosin. The proteins of this molecular motor family interact with actin. In this work we are interested in a subset of this myosin family.

### 1.1.2 The actin filament: the myosin track

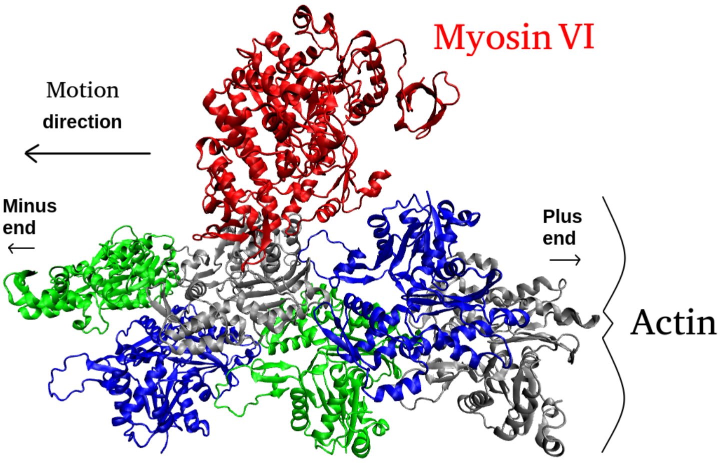


Figure 1.1: Cartoon representation of one myosin VI head in the Strong-ADP state on a portion of actin filament (from cryo-EM, pdb: 6BNQ). Blue, green and gray proteins are different actin monomers. The myosin motor domain is in red.

Actin represents a highly conserved family of proteins, which is the exclusive track on which myosins work (see Fig. 1.1). The sequence similarity of actin is more than 96% among all eukaryotes. The conservation of the actin sequence and its proportion in mammals (3% of the body weight) show its critical and universal role.

Actin exists in two main conformers: F-Actin and G-actin. G-Actin, which stands for "Globular Actin" is the conformer form taken by an isolated actin monomer. This monomer interact with a nucleotide (ATP or ADP) and divalent cation $Ca^{2+}$ or $Mg^{2+}$. F-Actin stands for "Filamentous Actin". This is the conformer taken by actin when they polymerize to form filaments of actin. The polymerisation is not done with covalent liaison but with interaction between the actin units. The form of the F-Actin unit is a bit flatter compared to the G-actin.

Actin binds to another actin with an ATP bound in its active site to form the filament. There is a preferred direction in the growth of F-Actin and therefore actin filament has two distinguishable ends. The plus-end of the actin filament is the region where more attachment of monomers occurs than detachment, while at the minus-end, more actin monomers detach than attach. Actin monomer detaches from F-Actin only after the hydrolysis of ATP and the release of the products. The difference between the two ends of the actin filament induces a polarity of the filament.

### 1.1.3 The myosin superfamily

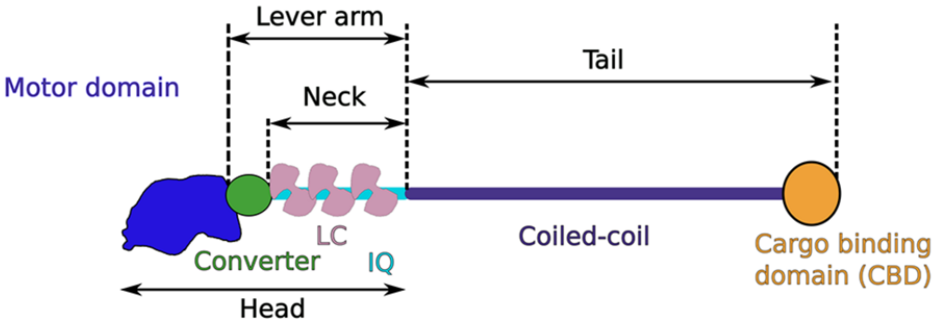


Figure 1.2: Schematic of the myosin anatomy. IQ stands for IQ motifs, which are motifs in a sequence of proteins that permits binding of calmodulin-like proteins, such as the light chains (LC). Taken from (Robert-Paganin et al., 2019).

Myosins represent a superfamily of motor proteins catalysing ATP hydrolysis, which also makes it an ATPase. Myosins recover energy from the ATP hydrolysis to fuel its functions. This superfamily of proteins is involved in numerous vital living processes. There exist a wide range of myosin families that are specifically tuned to accomplish all these functions.

In this section, we describe quite synthetically the system, the interested reader might refer to (Robert-Paganin et al., 2019) or (Houdusse and Sweeney, 2016).

The myosin protein is composed of the following general domains, represented on Fig. 1.2. The most important part of myosin is the motor domain. This motor domain is the active part of myosin, in which the hydrolysis of ATP is catalysed and the binding to actin occurs. The motor domain is composed of several subdomains, which are discussed in the following subsection. The motor domain is terminated by an $\alpha$-Helix which is an extension of the motor domain. This extension is called lever arm as it is assumed that its axial orientation relative to the rest of the motor domain amplifies the conformational changes of the motor domain. The lever arm is also called neck. The motor domain and the neck together form the head, also called subfragment 1 of myosin.

The neck usually has several IQ motifs, an amino acid sequence regularly found in proteins. These IQ motifs allow binding of calmodulin or calmodulin-like light chains (LC), which act as regulator of the myosin activity.

The lever arm is extended by a longer $\alpha$-Helix, called tail or subfragment 2. Depending on the myosin families, the size of this tail might change. Moreover, it can be terminated by a cargo binding site. In the case of two associated myosins, their tails are rolled on each other, forming a coiled coil of two $\alpha$-Helix.

Among the myosin families we can find functions like anchor or cargo transporter. This last function requires that the myosin transport components from one place to the other with a processive motility on their track. The processivity is permitted by the association of two heads attached by their tails. This processivity implies that each myosin head must spend the majority of its time bound to the track to allow in average permanent attachment of at least one of the two heads on actin. The processivity is not universal as other non processive myosin families spend most of their time detached from actin.

Most myosins move in the plus-end direction of the actin filament(Wells et al., 1999), making them plus-directed (see Fig. 1.1 for actin polarity). The exception to this rule is the myosin VI family, the one studied in this work, which is minus-directed. The difference in the direction of the movement of the myosin VI family is due to a supplemental part in its converter domain which is the insert-2 (Park et al., 2007).

In the sequel of the manuscript, we only discuss about the myosin motor domain, although we refer to it simply as “myosin”.

### 1.1.4 How molecular structures are determined

The molecular models (or molecular structures) discussed in this thesis come from two sources. Until recently, the high resolution molecular structures were mostly determined by X-ray crystallography (see the Protein Data Base statistics for more details in the https://www.rcsb.org). This method first necessitates to crystallize the proteins. This crystallization process involves some crystallizing agents and the cooling of the solution. The obtained crystals of protein are subjected to high energy synchrotron X-ray. The post treatment of the diffraction patterns allow to determine the position of every atoms except the hydrogen atoms, which cannot be sensed by X-rays, and the highly mobile atoms. Finally most atoms from the molecule can be positioned, which reveals the molecular structure of the protein.

More recently high resolution structures can be obtained from Cryo-Electron Microscopy. This method only necessitates to cool the protein, which then does not have to be crystallized. Cryo-EM was until recently way less precise than X-ray Crystallography. However recent improvement in the technology and especially post treatment techniques softwares now allows to reach resolution similar and sometimes better than X-ray Crystallography. For these reasons, it is now becoming the standard method for structure resolution.

### 1.1.5 Myosin subdomains and elements

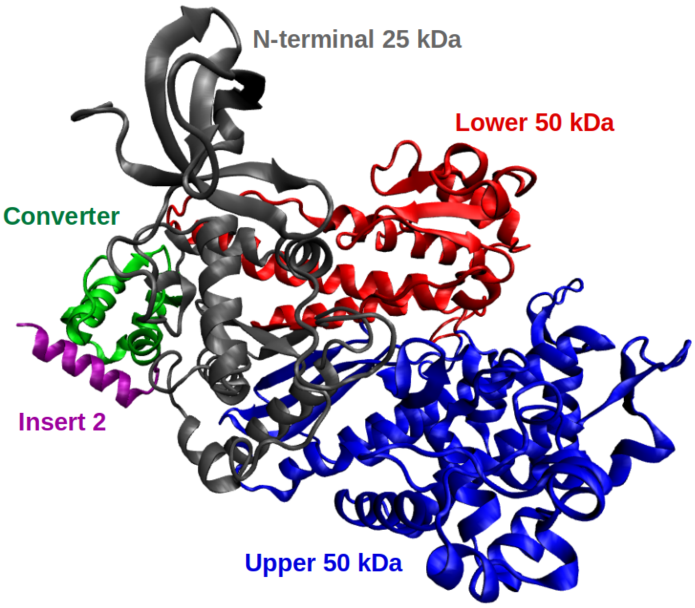


Figure 1.3: Representation of the myosin subdomains on the PPS state of myosin VI (PDB:2V26). The gray domain is the N-terminal 25 kDa. The blue domain is the Upper 50kDa. The red domain is the Lower 50kDa. The Green domain is the Converter. In purple is the Insert-2, a link between the motor domain and the lever arm, which starts in the case of myosin VI with an insert.

| Domain | Residue ranges | Number of residues |
|---|---|---|
| N-terminal 25kDa | 1-172, 671-705 | 207 |
| Upper 50kDa | 173-459, 600-631 | 319 |
| Lower 50kDa | 460-599, 632-670 | 179 |
| Converter | 706-770 | 65 |
| Insert-2 | 771-789 | 19 |

Table 1.1: Definition of the subdomain by residue ranges and number of residues for myosin VI.

Myosin is composed of four main subdomains. These subdomains are represented on Fig. 1.3. The definition of the subdomains by their residue ranges can be found in the table 1.1.

The first subdomain is the N-terminal, also sometimes called 25 kDa N-terminal subdomain (abbreviated 25kDa N-term). The second and third subdomains are the Upper 50 kDa and

Lower 50 kDa, respectively (abbreviated L50kDa and U50 kDa, respectively). These three domains were defined by proteolytic studies (Rayment et al., 1993; Mornet et al., 1981; Mahmood and Yount, 1984). This historical description can also be found in the name of the cleft, a space between Upper 50 kDa and Lower 50kDa, sometimes called 50kDa cleft (Ruppel and Spudich, 1996). The fourth and last subdomain is the converter. The converter is followed in the specific case of the myosin VI family by an element preceding the lever arm called insert-2.

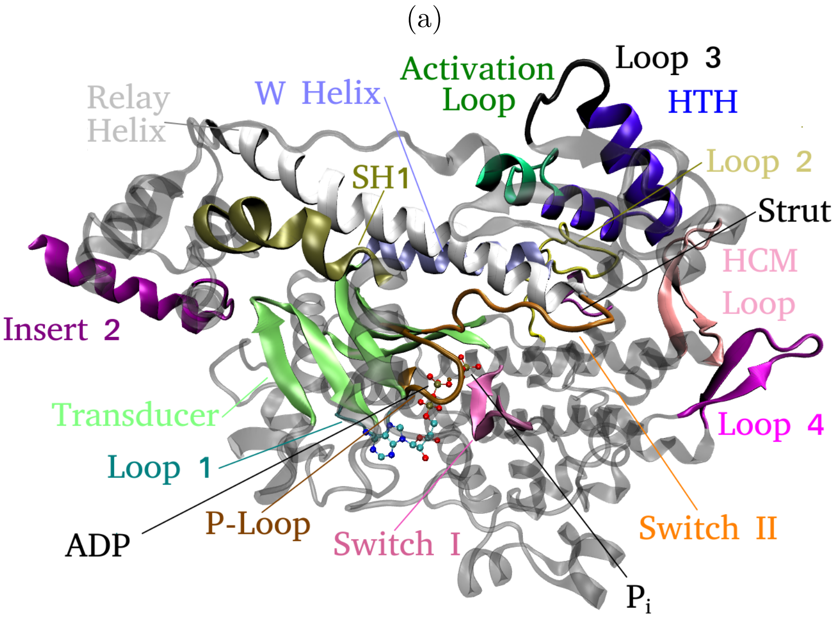


Figure 1.4: Cartoon representation of the myosin (pdb : 4pjm) with elements named. In the linear amino acid succession order: Transducer (lime), P-Loop (brown), Loop 1 (cyan), Switch I (mauve), Loop 4 (magenta), HCM Loop (pink), Switch II (orange), Relay helix (white), Activation Loop (turquoise), HTH (violet), Loop 3 (black), Strut (magenta), Loop 2 (yellow), W Helix (light blue), SH1 (tan). The colors used in these figures will be used all along the manuscript. Figure realised with VMD (Humphrey et al., 1996)

Elements of these subdomains playing major role in myosin function are represented on Fig. 1.4. The main characteristics of these elements are presented in the table included in the Fig. 1.5. These characteristics are: the name of the elements, the type of elements, the domain they belong to, the residues numbers and the functions.

We can define three specific types of elements. The first elements are the Nucleotide Binding Loops (NBL): the P-Loop, Switch I and Switch II. These elements are located at the active site in the center of the protein, and interact directly with the nucleotide (ATP or ADP). The NBL are actively participating in the hydrolysis. Their conformational change induced by the change of the nucleotide state is transmitted to distant regions of the protein through an allosteric pathway involving the transducer. This allostery is notably using the transducer to which the NBL are all connected at one end to communicate changes between distant part of myosin.

| Type of elements | Element | Subdomain | residue numbering | Function |
|---|---|---|---|---|
| Active Site | P loop | Nter | 150-158 | binds ADP |
| | Switch I | U50 | 195-205 | binds ADP & Mg |
| | Switch II | U50 & L50 | 457-467 | blocks Pi exit & controls lever arm position |
| Connectors | Loop1 | Nter & U50 | 174-180 | Transmit motor domain rearrangement to converter |
| | Relay helix | Converter & Nter | 467-500 | |
| | SH1 helix | Converter & L50 | 693-704 | |
| | Strut | U50 & L50 | 598-602 | assists cleft closure |
| | Loop2 | U50 & L50 | 622-642 | |
| Actin Binding Loops (ABL) | Activation loop | L50 | 515-525 | Interact either or both electrostatically and stereospecifically with actin |
| | HTH | L50 | 526-551 | |
| | Loop3 | L50 | 552-556 | |
| | Loop4 | U50 | 352-365 | |
| | HCM loop | U50 | 392-413 | |
| Others | Transducer | 4 Nter | 87-90, 93-98<br>145-150, 207-215 | deformations propagate changes between distant regions |
| | | 3 U50 | 220-229, 449-456, 662-670 | |
| | Cleft | between U50 & L50 | | Open or closed state related to affinity to actin |
| | W helix | L50 | 643-661 | ”hub coupling the structural changes of switch elements during ATP hydrolysis to temporally controlled interactions with actin that are passed to the central transducer and converter” |

Figure 1.5: Named elements of Myosin proteins playing a role in its catalytic cycle. We give the type of the element in which subdomain they are found, the numerical range of their residues (in the case of myosin VI) and their functions.

The transducer is a 7 stranded beta-sheet located in the interior along the length of the protein, see Fig. 1.4. The interactions between the strands of the transducer help transmit changes from the nucleotide binding pocket to the cleft (Coureux et al., 2003). Notably, it was described that an exchange of H-bonds between an upper strand and a lower strand is found in an intermediate state along the recovery stroke between Post-Rigor and Pre-Powerstroke that might impact the following changes leading to PPS (Blanc et al., 2018). A second type of elements are the connectors, among which are Switch II, Loop1, Relay Helix, the SH1 helix, the Strut and the Loop2. These elements connect the subdomains to each other.

The third type of elements are the actin binding loops, among which we find the Activation loop, Loop2, Loop3, Loop4, Helix-Turn-Helix (HTH) and the HCM Loop (Hyper cardiomyopathy Loop). These elements have residues that are interacting with actin at least momentarily during the myosin cycle in order to drive the conformational changes. Notably, the interactions they form might be responsible for the closure of the cleft, forming a strong binding (Llinas et al., 2015). This closure of the cleft then impacts the rest of the protein (Várkuti et al., 2015), (Gurel et al., 2017).

What can be seen in the table inserted in Fig. 1.5 is that Switch II is included as both connector of U50 and L50 and part of the Nucleotide Binding Loops. Similarly, the Loop2 is considered both as a connector and as Actin Binding Loop.

## 1.2 Comparison of the crystal structures of PPS, $P_iR$ and Strong-ADP states

In this section we show the main structural differences between the states PPS, $P_iR$ and Strong-ADP of myosin VI based on their crystal structures, see Fig. 3 in the Introduction. First, we define the cleft as it is used to compare the states. Secondly we compare PPS and $P_iR$. Finally we compare PPS and Strong-ADP.

### 1.2.1 Definition of the cleft of the myosin motor domain

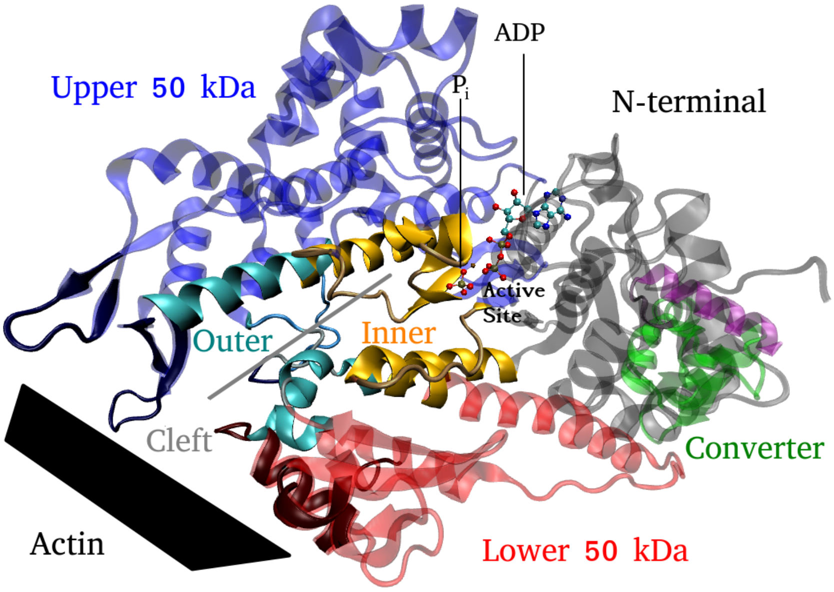


Figure 1.6: Representation of the myosin motor domain, with Upper 50 kDa in blue, Lower 50 kDa in red, Converter in green, insert-2 helix in purple, N-terminal domain in gray. The Cleft, which is the separation between the Upper 50 kDa and Lower 50 kDa subdomains is materialized by a gray line. It is composed of two parts from which surrounding residues are highlighted in two different colors. The Outer Cleft is highlighted in cyan, and the Inner Cleft in yellow. The Outer Cleft is on the side of the Actin Binding Surface, represented in black, while the Inner Cleft is on the part of the Active Site, in which the ADP-$P_i$ and $Mg^{2+}$are represented.

The cleft is the space between Upper 50 kDa and Lower 50 kDa, as represented on Fig. 1.6. It is an important feature used in the description of the myosin state that closes more and more along the progression of the PS. This linear view is however challenged by the still putative $P_iR$ state, in which the outer cleft is indeed observed as more closed than in PPS, but the inner cleft is more open than in PPS. It is the opening of the inner cleft that supposedly facilitate the escape of $P_i$ from the active site (Llinas et al., 2015).

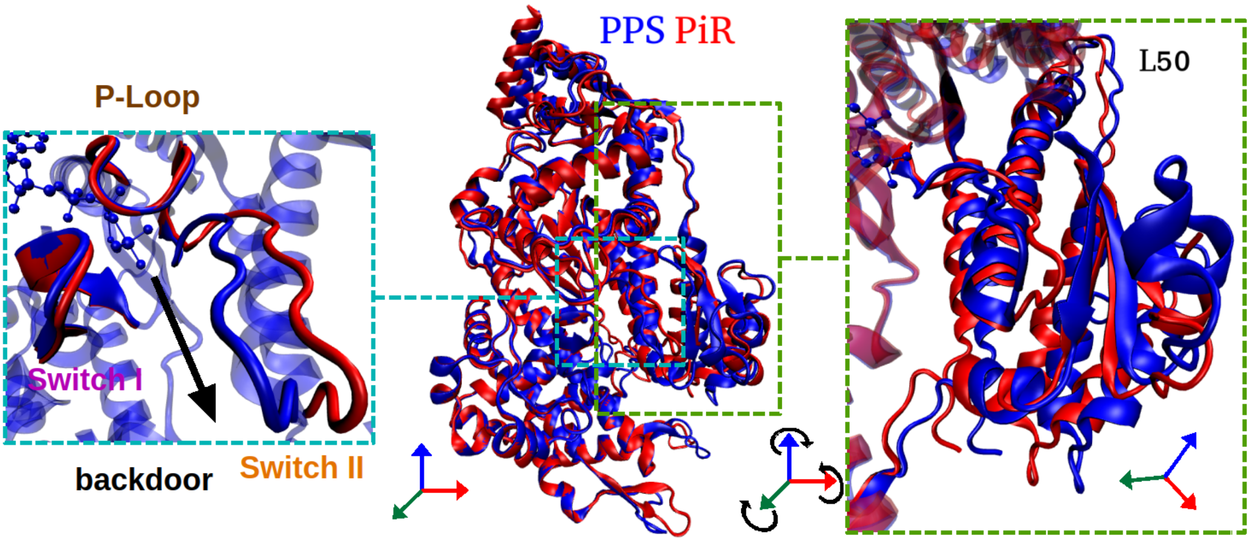


Figure 1.7: Representation of the superimposition on U50 of crystal structures of PPS (blue) and $P_iR$ (red). (middle) Whole protein, (left) Close up view on Switch II showing the difference of conformation in both crystals (on the right part of the figure) and (right) Close up view showing L50 rotation.

## 1.2.2 Comparison of PPS and PiR

The crystallographic structures of PPS and $P_iR$ are superimposed in Fig.1.7. From a macroscopic point of view, the two structures are almost identical, see figure in middle of Fig.1.7. The main difference is a slight rotation of the L50 subdomain, see right part of Fig. 1.7. This rotation leads to a slightly more closed outer cleft in $P_iR$ compared to PPS. However, the inner part of the cleft is more open. Indeed, the Switch II is slightly farther from Switch I in $P_iR$ compared to PPS, as represented on the left of Fig. 1.7. This farther distance opens a tunnel from the myosin ATPase catalytic site to the outside of the protein, which creates an exit route from the active site for the $P_i$, called backdoor (Yount et al., 1995).

### 1.2.3 Comparison of PPS and Strong-ADP

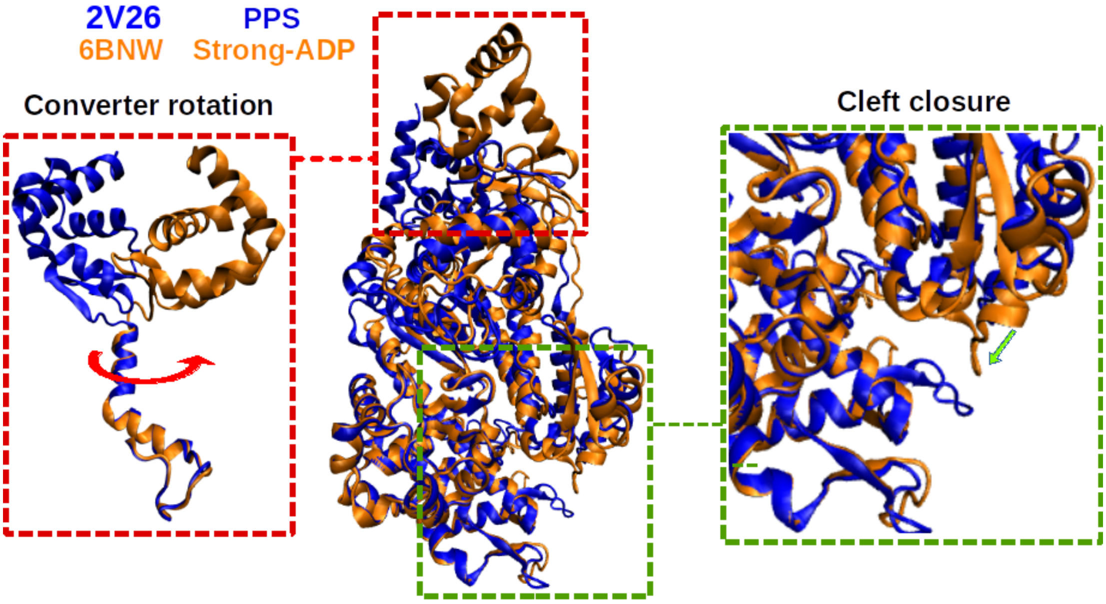


Figure 1.8: Superimpositions of the crystal structures of PPS (blue) and Strong-ADP (orange). (left) Representation of the rotation of the Converter with a superimposition on the SH1 helix. (middle and right) Superimpositions on U50 (middle) representation of the whole protein, (right) representation of the closure of the cleft.

The superimposition of PPS and Strong-ADP crystal structures is represented on Fig. 1.8. What can be seen is a clear difference of the converter position as emphasized in the left part of Fig. 1.8. This difference is related to the Powerstroke that is already mostly performed when the protein is in Strong-ADP. One other major difference is a closure of the cleft as represented on the right part of Fig. 1.8.

# Part I

# Study of phosphate release

# Chapter 2

# Bibliography of the departure of Pi

In this chapter we present previous studies that have given insights to the release of $P_i$ from the myosin active site. We particularly focus on the $P_i$R hypothesis. The $P_i$R hypothesis is the hypothesis that a state named $P_i$-Release is present in the actin-myosin cycle. The hypothetical state $P_i$R is assumed to reconcile experimental and structural data. We also present some recent challenges (already addressed in the literature or that we have addressed in this Thesis) to the $P_i$R hypothesis.

## Contents

## 2.1 Notable : some models exist without prior powerstroke or $P_i$ release

In (Caremani et al., 2013), the authors performed mechanical experiments on demembranated muscle fibers with different concentrations of $P_i$. They can explain their data with a model in which neither $P_i$-release nor Powerstroke must necessarily occur before the other. This means that in their model both powerstroke or $P_i$ release could happen before the other. However, this model includes a possible event that is difficult to explain structurally, namely a slipping of a myosin head from an actin to an adjacent one while the powerstroke is in progression.

## 2.2 What has brought to the $P_i$R hypothesis

### 2.2.1 Backdoor and other exit pathways

One of the first structural insights into the mechanism of the $P_i$ release came from the first solved crystallographic structures of myosin. Based on this structures, (Yount et al., 1995) fitted numerically an ATP into the protein pocket. They concluded from the position of ATP that $P_i$ could not leave the active site the same way ATP entered (called frontdoor) unless ADP leaves also the pocket. $P_i$ thus had to leave another way called backdoor in contrast with this frontdoor. The backdoor is characterized by a tunnel that can form between Switch I and Switch II, connecting the active site to the exterior of the protein.

(Lawson et al., 2004) have performed unconstrained molecular dynamics simulations on myosin II with ADP and $P_i$ ($HPO_4^{2-}$ form) inside the active site cavity and focused their study on the backdoor pathway. Although they did not observe $P_i$ departure, the simulation with a structure with open cleft led to a further closure of this pathway that only allowed water molecules entry.

(Kaliman et al., 2009) studied the $P_i$ departure mechanisms assuming the backdoor was the exit pathway in myosin II. They assumed that this exit might be blocked by the presence of the critical salt bridge, a salt bridge forming between GLU459 from Switch II and ARG238 from Switch I (myosin II numbering) identified as helping the ATP hydrolysis (Fisher et al., 1995; Smith and Rayment, 1996). They thus computed the free energy barrier associated with the opening of the backdoor by forcing the breaking of the critical salt bridge using Umbrella Sampling. Simulations were performed for two different myosin model one with protonated GLU459 and one with deprotaned GLU459. In the deprotonated model, the free energy barrier was estimated at 27 kcal/mol, while it was at 16 kcal/mol with protonated GLU459. The free energy barrier associated with the disruption of the salt bridge was estimated at 16 or 27 kcal/mol depending on the protonation state of GLU459.

(Cecchini et al., 2010) used molecular dynamics simulations aiming at identifying exit pathways on myosin II taken by the $P_i$ during release to solution. The study was performed using Pre-Powerstroke, Rigor and Post-Rigor structures as initial conditions. As a reminder, Pre-Powerstroke is the only one of these states in which $P_i$ is present, while Post-Rigor is a pre-hydrolysis state (*i.e.* with ATP), and Rigor is nucleotide free. This molecular dynamics study was of the type of Multiple Copies Enhanced Sampling. This method consists in having multiple virtual copies of the subsystem of interest (here the $P_i$) interacting with the rest of the system. These interactions result from the average of individual interactions of each copy of $P_i$ with the system. The authors used 30 copies of $P_i$ in each simulations. These copies of $P_i$ were given supplementary kinetic energy (temperature) using a local temperature increase, while the rest of the system was maintained at a temperature of 300K. The simulations were performed in increasing temperature of the $P_i$ from 300K up to 3000 K.

The author observed an increase of release probability of $P_i$ starting at 900K in Post-Rigor and Rigor-like, while it only increased slightly at 1500K and really increased starting at 2000K in PPS. This means that the energy barrier associated with $P_i$ departure is of several order of magnitude higher in PPS compared to Rigor and Post-Rigor.

The authors identified a total of 6 exit pathways : side door, backdoor, backdoor II, front door, rear and top. In PPS and Post-Rigor, the dominating path was the backdoor (between switch I and switch II) guiding the phosphate into the large cleft between the U50 and L50 subdomains. In Rigor, while the cleft is completely closed, the dominating path observed was the so-called "backdoor II" between P-loop and switch II, releasing $P_i$ in proximity to helix SH2 through a short tube between the N-terminal and the L50 subdomains.

One noticeable common feature of the Rigor and Post-Rigor states is an open Switch II. The authors then suggested that an open Switch II facilitates $P_i$ departure, regardless of the actual predominating path in both structures.

### 2.2.2 The hypothesis of an intermediate state called $P_i$-Release

In (Llinas et al., 2015), the authors present multiple structures of Myosin VI that they have crystallised and solved. The different structures presented exhibit different positions of the $P_i$ which are summarized in the table 2.1 and represented on Fig. 2.1.b.

| PDB code | Position of $P_i$ | How crystals were obtained |
|---|---|---|
| 4PFO | without $P_i$ | First structure solved/ Initial $P_i$R crystals |
| 4PJN | $P_i$ inside backdoor tunnel (position $P_i$R1) | $P_i$R crystals soaked very shortly in $P_i$ buffer |
| 4PJM | $P_i$ inside active site (position $P_i$R2) | $P_i$R crystals soaked slightly longer in $P_i$ buffer |

Table 2.1: Differences between the three main $P_i$R crystal structures, 4PFO, 4PJM and 4PJN.

First, the authors obtained crystals of myosin with MgADP but without $P_i$ in a new state (PDB: 4PFO). The comparison of the solved structures with a structure of myosin VI in Pre-Powerstroke state (PPS, PDB: 2V26) led them to characterize this state by the following

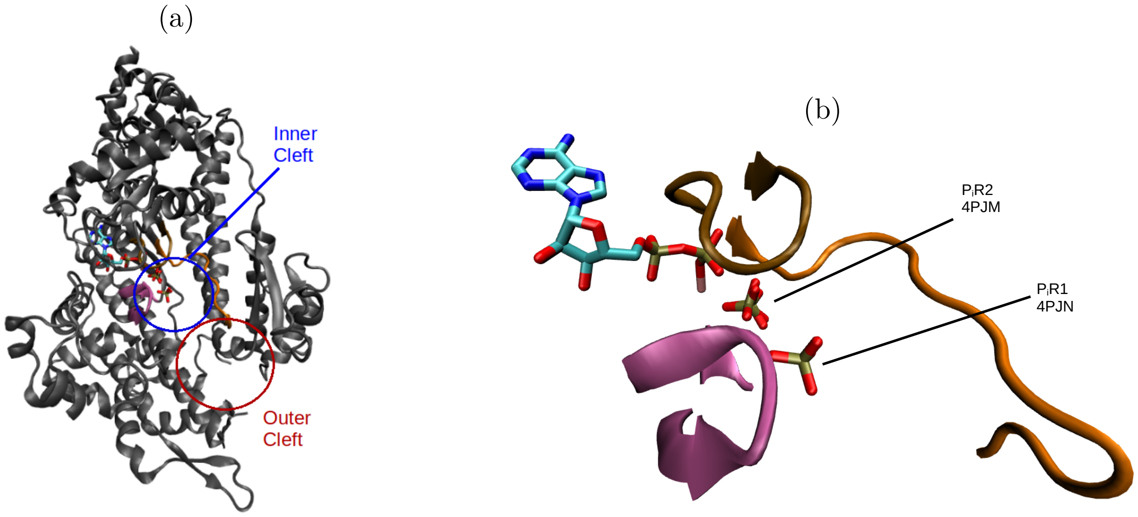


Figure 2.1: Representation of the two positions of the $P_i$ in the myosin. (a) In the context of the whole protein showing inner cleft and outer cleft. The global conformation of the structures 4PFO, 4PJM and 4PJN are very similar. (b) Close up view on the active site. The $P_i$R1 position is in the backdoor tunnel, between Switch I and Switch II. The $P_i$R2 position is in the active site, interacting with ADP. This last position is the position occupied after ATP hydrolysis.

differences: an open inner cleft, a slightly more closed outer cleft and a kink in the SH1 helix. The open inner cleft in the structures in the new state is notably characterized by the Switch II position far from Switch I. This opens a way for $P_i$ out of the active site between Switch I and Switch II that was called the backdoor tunnel. Moreover, this crystal structure is a proof that a Switch II movement can happen without a major lever arm movement, which was an argument against the backdoor hypothesis. Mostly because of the open inner cleft, the authors hypothesize that these structures are in a state favoring $P_i$ departure. For this reason, they called this state $P_i$-Release ($P_i$R).

As the obtained structure (PDB 4PFO) was seen as allowing exchange of $P_i$ with the exterior of the protein, the authors of (Llinas et al., 2015) soaked the same crystals, previously without $P_i$, in a phosphate buffer in order to evaluate how much the $P_i$ can enter in the protein in this state. With a soaking as quick as possible, they solved crystal structures in which the $P_i$ was located in a position at the end of the backdoor tunnel (position named $P_i$R1, PDB 4PJN) described as the main exit pathway. With a slightly longer soaking, the $P_i$ was located in the active site, close to the ADP/ the position after hydrolysis (position named $P_i$R2, PDB 4PJM).

The following mechanism of $P_i$ release was proposed : $P_i$ translocates from the $P_i$R2 position to the $P_i$R1 position before a complete release out of the protein. The active site undergoes a conformational change after $P_i$ departure. The translocation from $P_i$R2 to $P_i$R1 allows all the necessary molecular rearrangement for the powerstroke, namely essentially loops adaptation and transducer conformational change. With such a mechanism, $P_i$ remains inside the myosin, and therefore, cannot be detected in the solvent.

What should be kept in mind is that the $\mathrm{P_iR}$ state was also obtained with a myosin II (instead of myosin VI) after mutating the residues forming the critical salt bridge (structure PDB 4PJK). Additionally, crystals containing $\mathrm{P_iR}$ were obtained with a different crystal group (structure PDB 4PFP), meaning this state is not only due to crystal packing. However, the unit cell of the crystal contained two myosins. One of the myosin is in the $\mathrm{P_iR}$ state, while the other is in the PPS state.

All these crystal structures (4PFO, 4PFP, 4PJM, 4JPN and 4PJK) represent an ensemble of proofs of the existence of the state that the authors called $\mathrm{P_iR}$.

What was observed is that when $\mathrm{P_iR}$ crystals were soaked for a long time in the $\mathrm{P_i}$ buffer, the solved crystals were in a PPS state (structure PDB 4PK4). On contrary, Post-Rigor (PR) crystals with MgADP were soaked in a $\mathrm{P_i}$ buffer, and none of them allowed entering of $\mathrm{P_i}$ in the active site (structure PDB 4PJJ).

The study in (Llinas et al., 2015) was also based on functional experiments. These functional experiments consist in measuring the rate of $\mathrm{P_i}$ release from the myosins. In some cases (when the $\mathrm{P_i}$ release rate was deemed large enough, $\sim 100\ \mathrm{s}^{-1}$ ) the measurement of the rate of cleft closure was also performed. The measurements have been performed on three types of myosins (II, V and VI). This study was performed by comparing the rate of $\mathrm{P_i}$ release of the wild type proteins (protein without mutation) with proteins in which some residues were mutated. These mutations allow to evaluate the relative importance of each residues in the regulation of $\mathrm{P_i}$ release. They performed mutation on S203A, A458E, E152A, A422L and R521E (all in myosin VI numbering). Among these mutations, the S203A (Serine 203 transformed into Alanine) and A458E (Alanine into Glutamic Acid) from Switch I and Switch II respectively had large effects on $\mathrm{P_i}$ release. These mutations reduced by 2 and 3 fold the rate of release, respectively. The S203 guides the $\mathrm{P_i}$ away from the active site thus the mutation S203A impedes $\mathrm{P_i}$ entry into the backdoor tunnel, while A458E hinders and adds a charge into the tunnel.

The mutation R521E was also found to reduce the $\mathrm{P_i}$ release rate, which becomes 3 times slower than the WT. This mutation reverses the charge in activation loop which has the effect of slowing the formation of the $\mathrm{P_iR}$ state.

Mutation of A422 into leucine did not affect the $\mathrm{P_i}$ release rate but reduced the rate of cleft closure associated with the powerstroke, which is then expected to occur less quickly with this mutant. As the powerstroke is affected by this mutation while $\mathrm{P_i}$ release is not, this last result could support the hypothesis that $\mathrm{P_i}$ release occurs before the powerstroke.

## 2.3 Challenges to the $\mathrm{P_iR}$ hypothesis

In general functional studies tend to conclude that the $\mathrm{P_i}$ release occurs after the powerstroke as the apparent rate of $\mathrm{P_i}$ release in solution calculated with $\mathrm{P_i}$ binding proteins is slower than the rate of lever arm movement associated with the powerstroke. However these conclusions are based on the fact that the $\mathrm{P_i}$ is only detected when it arrives in solution. Thus it allowed to propose a mechanism in which the $P_i$ could translocate from the active site without reaching the solvent as it stays in the protein (Llinas et al., 2015). This is the source of the $P_i$R hypothesis. However, (Scott et al., 2021) performed single molecule study on myosin Va in which they monitored the powerstroke. The $P_i$ release was slowed by using a myosin with the mutation S217A (equivalent to S203A in myosin VI). The experiments did not result in a slowed rate or size of powerstroke between the control condition and an increased concentration of $P_i$. This would be clearly observed if $P_i$ release occurs before the powerstroke. Moreover in (Scott et al., 2021), the authors concluded by simulating their data that a model with $P_i$-Release occurring before cleft closure is impossible as it would have resulted in a pause after myosin binds to actin that was not observed (Debold, 2021). This is thus a big argument against the $P_i$R hypothesis.

## 2.4 What could be performed to support the $P_i$R hypothesis

In the study of (Llinas et al., 2015), it is assumed that the $P_i$R state is stabilized by actin although the state is solved without actin. In order to better establish the existence of this state, we shall address the following questions. For each question, we cite the previous simulations done and what simulation we have performed or could be performed to give insights for the answer.

### 2.4.1 Does the $P_i$R state really favor the $P_i$ departure?

In the study of (Mugnai and Thirumalai, 2021), the authors assumed the existence of the $P_i$R state. Then they ran simulations without checking whether their system still exhibit the $P_i$R characteristics.

In this thesis we have performed some simulations in order to characterize whether starting with a $P_i$R state actually lowers the potential of mean force of the $P_i$ departure from myosin compared to the simulation with a PPS state. During the simulations, structures are free to evolve, so we monitor when structures starting from PPS become $P_i$R or conversely when structures starting from $P_i$R become PPS.

### 2.4.2 Could the $P_i$R state exist between PPS and Strong-ADP?

In order to test this hypothesis, we shall drive the system to transform from the PPS state with and without $P_i$ in order to judge if this transformation can be favorable without $P_i$. This transformation should allow the visiting of the $P_i$R state by allowing the inner cleft to open. It should then be checked if the $P_i$R state is visited and possibly favorable. Some preliminary simulations in which we transformed the system by driving a collective variable from its value in PPS to its value in Strong-ADP were performed (see Appendix D).

(Preller and Holmes, 2013) performed a study in which they forced the closure of the cleft as in Rigor, but did not allow the inner cleft to open, which prevented the $P_i$ to be released and thus to explore the $P_iR1$ position.

### 2.4.3 Could the $P_iR$ state exist with actin?

The $P_iR$ state is supposed to be present in the actin-myosin cycle in interaction with actin (see Fig.3). Therefore the influence of the actin on this state is a legitimate question. Could the $P_iR$ state be stabilized by interactions with actin? Could the $P_iR$ state exist with actin? What would stabilize the $P_iR$ state?

(Llinas et al., 2015) made a mutation on the Activation Loop that resulted in a decrease of the $P_i$ release rate. They conclude that the Activation Loop had a major role in stabilizing the $P_i$ release state.

To answer the questions on the role of actin on the $P_iR$ state, simulations with the presence of actin and myosin could be performed. However these simulations would necessitate assumptions on the actomyosin ensemble as there is no yet available structure related to PPS or $P_iR$. An alternative would be to restrain actin binding loops and see if it stabilizes the $P_iR$ state. This study was not performed in the present work.

### 2.4.4 What could block or help the $P_i$ departure?

The main purpose of our study of the $P_i$ departure is focused on the interactions of the $P_i$ with surrounding residues along its departure and the study of the interactions between the Nucleotide Binding Loops (see chap. 3).

### 2.4.5 What could trigger the $P_i$ departure?

(Mugnai and Thirumalai, 2021) realised multiple simulations lasting several microseconds of unconstrained molecular dynamics of myosin VI using the $P_i$-Release ($P_iR2$) structure as a starting point. In this study, the $P_i$ is in the form with two protonated oxygen atoms $H_2PO_4^-$.

They found that $P_i$ release might be favored when $Mg^{2+}$ is coordinated by four water molecules. Moreover, $P_i$ was only released when they rotated it in the binding pocket from the orientation foud in the crystal, which means that $P_i$ orientation have a strong impact on the release molecular mechanism.

As the way the authors of (Mugnai and Thirumalai, 2021) rotated the $P_i$ was unclear, the study presented in the following chapter is based on the definition of clearly defined orientation of the $P_i$, and in the comparison of results obtained starting from structures in PPS and $P_iR$, which evolution are monitored.

### 2.4.6 Is the $P_iR$ crystal structures obtained by soaking the crystal in $P_i$ solution an intermediate in the $P_i$ release?

The $P_iR1$ structure obtained by (Llinas et al., 2015) is an experimental structure obtained artifially by soaking myosin VI in a concentrated $P_i$ solution. We have tried to check whether we can find the same stable position of the $P_i$ as found in crystals along its departure ($P_iR1$ position) (see chap. 3).

### 2.4.7 What is the consequence of stabilizing the $P_iR$ state along a simulations?

(Mugnai and Thirumalai, 2021) have restrained every $C_\alpha$ atom in their initial position in the $P_iR$ state except the ones from SER204 and THR158 for a simulation of $P_i$ release from $P_iR$ state. SER204 and THR158 were chosen because they coordinate $Mg^{2+}$. This simulation did not result in $P_i$ departure. One reason is that it could be expected that other parts of myosin should be able to evolve freely for the $P_i$ departure. Moreover, this simulation was performed without rotating $P_i$, although the authors have shown with their other simulations that this rotation was useful to accelerate the departure of $P_i$.

It could be interesting to perform a similar study by restraining less residues, particularly the position of Switch II. This study was not performed by lack of time.

# Chapter 3

# Molecular Interactions Leading to Phosphate Release in Pre-Powerstroke and Pi-Release Structures

## Contents

## 3.1 Introduction

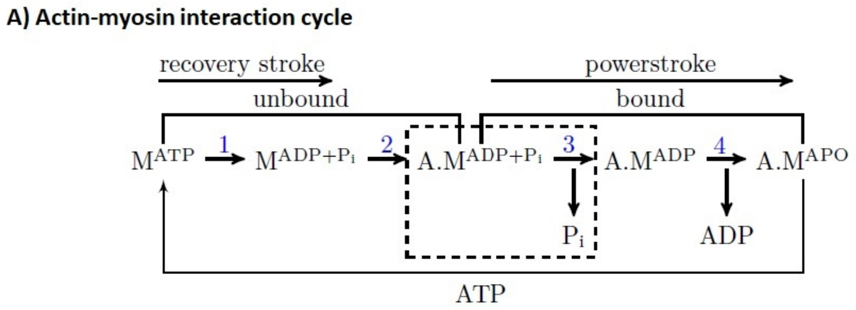
Muscle contraction is produced by a highly organized structure of microscopic contractile units, each of which comprising antagonists arrays of myosin II proteins pulling on actin filaments. Myosin II, is an ATPase that is able to convert metabolic energy harvested from adenosine triphosphate (ATP) hydrolysis into mechanical work in the presence of actin (Alberts et al., 2014; Woledge et al., 1985; Robert-Paganin et al., 2019). This energy transduction takes the form of an out-of-equilibrium cycle of structural conformational changes, during which, one ATP molecule gets hydrolyzed into adenosine diphosphate (ADP) and an inorganic phosphate ($\mathrm{P_i}$) within the myosin active site (Robert-Paganin et al., 2019; Blanc et al., 2018), see Fig. 3.1.

From a mechanical perspective, the so-called powerstroke (see steps 3 and 4 in Fig. 3.1) constitutes the most important step of this cycle. It consists in a large rotation of a lever-arm-shaped domain of the protein that produces, in the absence of external load, a $\sim$10 nm displacement of the actin filament to which the myosin is bound. This large conformational change is associated with the departure of the inorganic phosphate from the active site and the release of ADP.

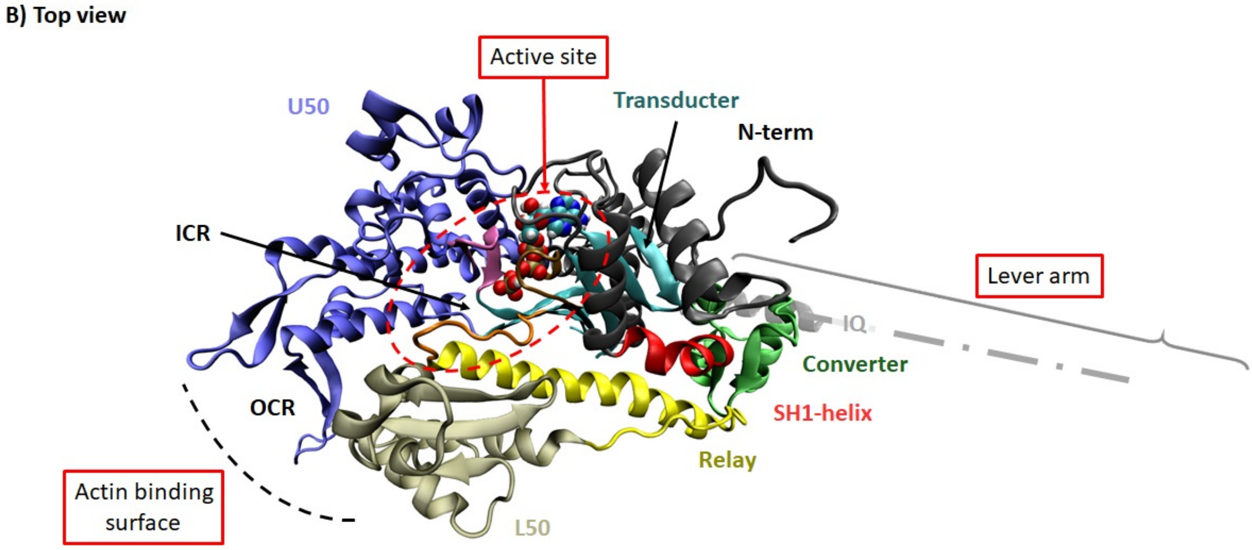
Since the release of $\mathrm{P_i}$ (step 3) is associated with the largest free energy drop in the biochemical cycle, it has been hypothesized that this step should precede the powerstroke (Geeves, 2016; Sweeney and Houdusse, 2010). This hypothesis is however challenged by mechanical experiments on muscle suggesting that the mechanical conformational pathway could be uncorrelated from the biochemical step (Caremani et al., 2013, 2015).

To elucidate this controversy, numerous structural studies have been conducted aiming at identifying the molecular mechanisms that promote the release of inorganic phosphate from the active site, reviewed in (Geeves and Holmes, 2005; Robert-Paganin et al., 2019). In this context, the interplay between working stroke and the $\mathrm{P_i}$ release was found to involve dynamic changes in the interactions between the following key secondary structures of Myosin II. These secondary structures are the nucleotide binding elements: P-loop (residues 150 to 157), Switch I (residues 195 to 205) and Switch II (residues 456 to 467), which are all in direct interaction with $\mathrm{P_i}$. A notable contribution of these structural studies was the identification of the so-called backdoor and frontdoor escape pathways using X-ray crystallography (Yount et al., 1995), see Fig. 3.2. The frontdoor is the entry point of ATP in the active site, located between P-loop and Switch I, and leading to the space between the N-terminal and the Upper 50kDa subdomains. An escape of phosphate through the frontdoor is judged unlikely, since the pathway is obstructed by ADP. In this configuration, $\mathrm{P_i}$ release would be preceded by ADP release, which is not compatible with an actin strongly bound state containing ADP (state named Strong-ADP). Hence, the backdoor is considered the main exit point of $\mathrm{P_i}$ (Yount et al., 1995; Robert-Paganin et al., 2019).

The hypothesis that the inorganic phosphate escapes the active site almost exclusively through the backdoor is supported by molecular dynamics simulation studies performed on



**A) Actin-myosin interaction cycle**



**B) Top view**

Figure 3.1: **(A)** Diagram representation of the actin-myosin interaction cycle. Starting with myosin (M) detached from actin (A) and, having an ATP bound in its active site (post-rigor state or $M^{ATP}$), the cycle can be decomposed into 5 steps. Step 1: ATP hydrolysis into ADP and $P_i$ in myosin active site ($M^{ADP+Pi}$). Step 2: Myosin binding to actin leading to the pre-powerstroke state (PPS or $A.M^{ADP+Pi}$). Step 3: release of $P_i$ from the active site leading to the strong-ADP state ($A.M^{ADP}$). Step 4: release of ADP from the active site leading to the rigor state ($A.M^{APO}$). The rigor state has no ligand inside its cavity. Step 5: Binding of a fresh ATP molecule and detachment from actin ($M^{ATP}$). The steps 3 and 4 are concomitant with the powerstroke, a large amplitude conformational change that generates force on the actin filament. The hydrolysis of ATP occurs after the recovery stroke, a large exploration of conformational states that myosin undergo with ATP bound. **(B)** Essential elements of the myosin VI in the Pre-powerstroke state (PPS) (PDB 2V26) motor domain after protonation and minimisation. Myosins are allosteric machines containing three major sites: the actin binding surface, the active site that binds and hydrolyzes ATP, and the mechanical element called the lever arm. The colors represent the the subdomains of the motor domain: N-terminus (N-term) black; upper 50 kDa (U50) purple; lower 50 kDa (L50) tan; Transducer cyan. The lever arm includes the converter (green) and the IQ helical region (grey, partially shown and materialized by a dot-and-dash line). The SH1-helix (red) and the Relay (yellow) are two structural elements that are cooperatively linked to the converter. These elements are organized around the internal cleft, which can be subdivided in two regions: the inner cleft region (ICR) near the active site and the outer cleft region (OCR) near the actin binding surface.

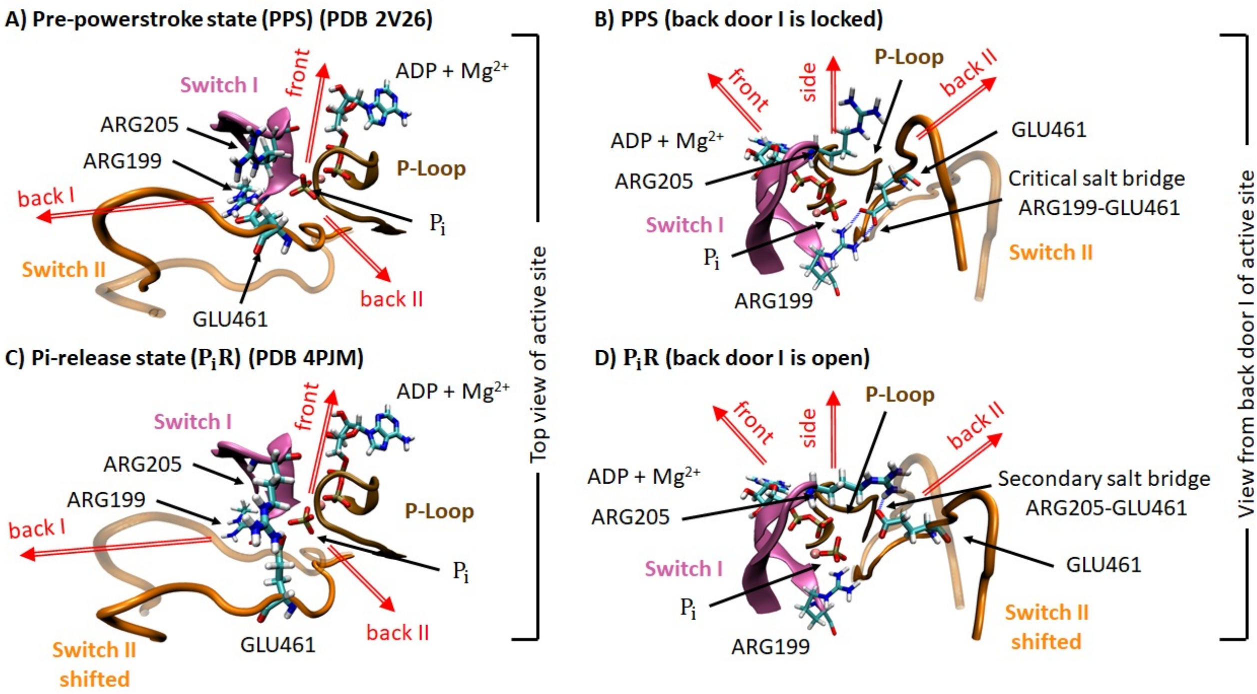


Figure 3.2: Active site domain of myosin VI studied structures after protonation and minimization. Two views of the Pre-powerstroke state (PPS) (PDB 2V26) **(A)** and **(B)** and two views of the Pi-release state ($P_i$R) (PDB 4PJM) **(C)** and **(D)** are shown. The P-Loop (residues 150 to 157) in brown, Switch I (residues 195 to 205) in pink and Switch II (residues 456 to 467) in orange are represented in cartoon. $P_i$, ADP, Mg$^{2+}$, ARG199 and ARG205 of Switch I, GLU461 of Switch II are represented in licorice with color code C cyan, H white, O red, N blue, P brown, and Mg$^{2+}$ light pink. Switch II is represented in solid for the represented structure and in transparent corresponding to the other state (corresponding to PPS for $P_i$R figures **(C)** and **(D)** and $P_i$R for the PPS figures (**(A)** and **(B)**) to illustrate the Switch II displacement and the opening of the phosphate release tunnel at the back door I in $P_i$R state. The $P_i$ release paths are represented by red double arrows following the terminology of (Cecchini et al., 2010). The backdoor I route (back I) goes through Switch I and Switch II, and the $P_i$ is released through the cleft; the frontdoor route (front) goes through the P-Loop and Switch I and the $P_i$ is released *via* the ATP-binding pocket; the backdoor II path (back II) goes through the P-Loop and Switch II and releases the phosphate in proximity to helix SH2; the side path (side) is perpendicular to the top view (see (**A**) and (**C**) and and its goes through the three nucleotide-binding elements (P-Loop, Switch I and Switch II). Note that the top and the rear routes defined in (Cecchini et al., 2010) are not shown for clarity reasons. The critical salt bridge (between GLU461 and ARG205) is formed in PPS state and close the backdoor I pathway. The secondary salt bridge (between GLU461 and ARG199) is partially formed in $P_i$R state and the backdoor I is open.

structures representing different states of myosin (Lawson et al., 2004; Cecchini et al., 2010). Three pathways were detected in these studies: backdoor I, backdoor II, and sidedoor (see Fig. 3.2 and review (Robert-Paganin et al., 2019)).

In simulations starting with ADP and $P_i$ in Pre-PowerStroke (PPS) and Post-Rigor (PR) structures, the pathway is principally through the tunnel named backdoor I. This pathway is characterized by interactions of the phosphate with switch I and switch II. The open or closed state of the tunnel depends on the formation or dissociation of the so-called critical salt bridge (ARG205-GLU461), which were first identified in (Yount et al., 1995). The second backdoor pathway (backdoor II) was observed in simulations involving Rigor structure, together with a less predominant pathway through the front door, characterized by interactions with switch I and P-Loop. The backdoor II pathway is characterized by interactions of $P_i$ with switch II and P-Loop and presents an energy barrier higher than for the backdoor I (Geeves and Holmes, 2005; Cecchini et al., 2010; Mugnai and Thirumalai, 2021). The third pathway (sidedoor) was observed by (Cecchini et al., 2010), where $P_i$ leaves the active site in between the three nucleotide-binding elements.

The crystallographic structures called $P_iR$ (4PJN, termed $P_iR1$ hereafter) published in (Llinas et al., 2015), also support the hypothesis of an escape of $P_i$ through backdoor I tunnel. Indeed, in these structures, the inorganic phosphate is found inside or at the end of the backdoor I tunnel (4PJN, termed $P_iR1$ hereafter) or close to ADP (4PJM, termed $P_iR2$ hereafter). These two positions could be obtained by varying the soaking time in a 25 mM to 100 mM $P_i$ solution, a higher concentration of $P_i$ in the solution leading to a $P_i$ located closer to the active site in the structure ($P_iR2$).

In these states, the lever arm is observed in the upward position alike the PPS conformation. The main characteristics of the $P_iR$ structure is an open Switch II, as opposed to the PPS structure, where Switch II is closed. In contrast, Switch I and P-loop remain close to the position they occupy in PPS. Another difference between $P_iR$ and PPS structures is a small rotation of the L50 domain (see Fig. B.1 in appendix B). The large opening of Switch II reveals an escape route through the backdoor I while the outer cleft, near actin, remains closed.

The structural model of force generation suggested by these observations is that the inorganic phosphate leaves the active site prior to the powerstroke, since the absence of the critical salt bridge allow the movement of Switch II to an open state. The $P_i$ can however remain trapped at the end of the tunnel ($P_iR1$) before it becomes detectable in the solution, a scenario that could be compatible with a release of $P_i$ following the working stroke.

More recently, (Mugnai and Thirumalai, 2021) ran unconstrained microsecond molecular dynamics simulations starting from the $P_iR$ crystallographic structure 4PJM ($P_iR2$). Simulations were performed with oriented or rotated $P_i$ to destabilize its interactions with the magnesium among other interactions and presumably ease the $P_i$ departure. The authors concluded on the fact that magnesium hydration by four water molecules precedes or at least facilitates the $P_i$ Release.

The goal of the present study is to describe the release pathway of $P_i$ from the active site, emphasizing the sequences of interactions occurring between phosphate, Switch I, Switch II and P-loop. Constrained molecular dynamics simulations were performed using the Umbrella Sampling (US) method (Torrie and Valleau, 1977) to force the departure of $P_i$ starting from crystallographic structures of myosin in Pre-Powerstroke conformation (PPS, PDB : 2V26) and $P_i$-Release conformation ($P_i$R2, PDB : 4PJM). In particular, we studied the effect of the $P_i$ protonation state on both its relaxed configuration in the active site, on the molecular interactions occurring during its release and on the associated energy landscape. Our results support the conclusion that $P_i$ release from the active site is easier from the $P_i$R structure compared to the PPS structure. In addition, our study seems to confirm that, from the $P_i$R state, the favored escape pathway is through backdoor I.

## 3.2 Results

### 3.2.1 Effect of Pi protonation state on its orientation within the active site

In our study, we used the phosphate form $H_2PO_4^-$, which is characterized by the presence of two protonated oxygen atoms. This form corresponds to the dominant protonation state in neutral pH solution with low ionic-strength (Powell et al., 2005). In addition, previous studies have shown that this form is preferred as product of ATP hydrolysis by myosin (Schwarzl et al., 2006; Kiani and Fischer, 2014), and that it is prevalent during phosphate release (Grigorenko et al., 2007; Wriggers and Schulten, 1999; Okazaki and Hummer, 2013). It is also the form used in (Cecchini et al., 2010; Mugnai and Thirumalai, 2021), where molecular dynamics simulations similar to the ones presented in this paper are performed.

To study the effect of the orientation of $P_i$ on its release process, six configurations have been constructed by protonating two out of its four oxygen atoms. The six configurations are labeled as PPS-(A,B,...,F) and $P_i$R-(A,B,...,F) for the PPS and $P_i$R states, respectively, see Tab. 3.1. The orientation of the four oxygen atoms are numbered from 1 to 4, according to the convention taken in Fig. 3.3 (B and D). The position number of the protonated oxygen atoms for a given state are indicated in parentheses in Tab. 3.1. For instance, in the PPS-A configuration, the orientation (3,4) corresponds to a phosphate where the oxygen atoms pointing in direction 3 and 4 are protonated. The change of $P_i$ observed during relaxation simulations are indicated by the arrows in the second column of Tab. 3.1.

#### 3.2.1.1 Relaxation of PPS structures

The structure of the active site after relaxation of the six initial PPS configurations are represented in Fig. 3.4, with the inserts showing the $P_i$ orientation before relaxation. In these inserts,

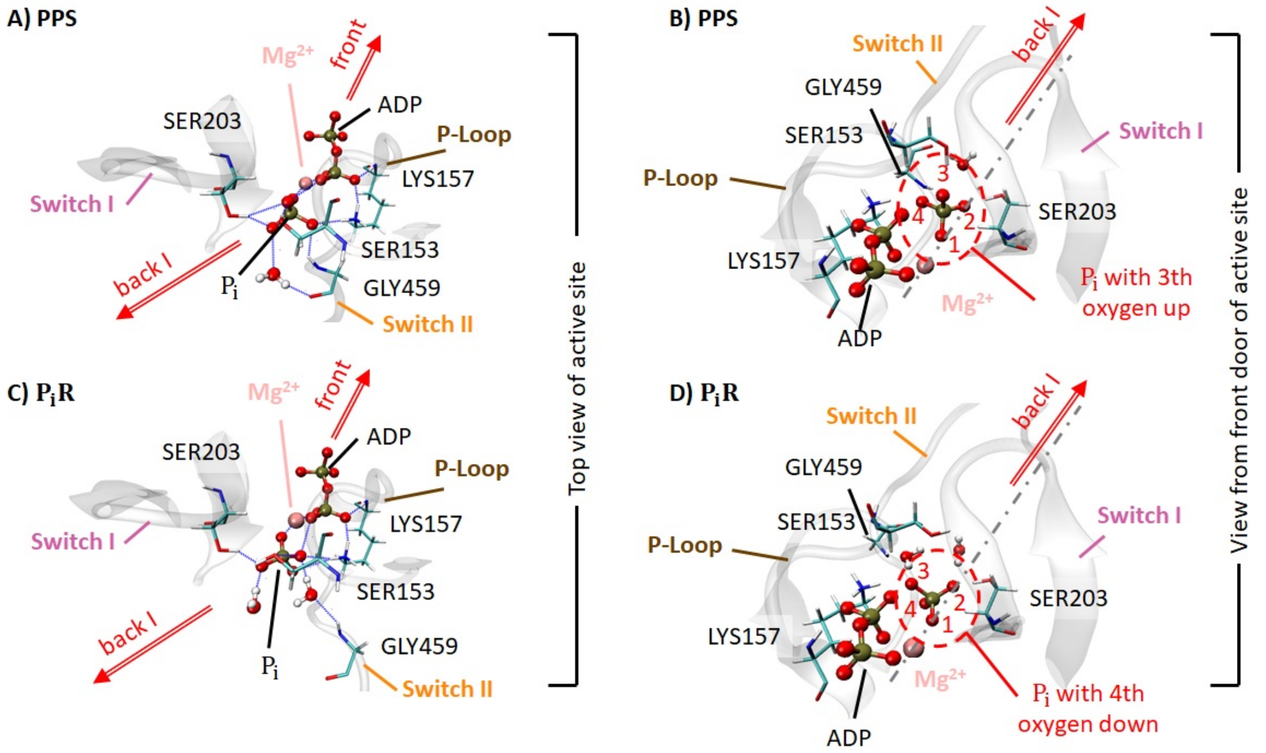


Figure 3.3: Definition of the $P_i$ orientation and main starting interactions in active site domain of myosin VI. studied structures after protonation and minimization. **(A)** and **(B)** NBL of PPS after protonation and minimization and **(C)** and **(D)** NBL of $P_iR$ after protonation and minimization. The $P_i$ is not protonated as this represents the starting structures of all simulations. The nucleotide binding loop P-Loop, Switch I and Switch II are represented in ghost cartoon. SER153 and LYS157 of P-Loop, SER203 of Switch I and GLY459 of Switch II are represented in licorice with color code C cyan, H white, O red, N blue. $P_i$, ADP without the nucleoside part, $Mg^{2+}$ and water molecule close to $P_i$ are represented in CPK with color H white, O red, P brown and $Mg^{2+}$ light pink. In **(A)** and **(C)** (showing the same orientation of the ADP for PPS and $P_iR$) the possible H-bonds and interactions with $Mg^{2+}$ are represented with blue dotted lines. In **(B)** and **(D)** (showing the same orientation of the ADP for PPS and $P_iR$) the convention used to describe the orientation of the $P_i$ is highlighted. Note that the orientation of the $P_i$ is slightly different in $P_iR$ from PPS. We have defined an axis which starts from the $Mg^{2+}$ in the direction of backdoor I, presented by a gray dot-and-dash. This axis goes through the oxygen position O1 and position O2 of $P_i$. The oxygen position O1 corresponds to the oxygen closest to $Mg^{2+}$. The oxygen position O3 is pointing to the SER153 residue. When looking at the $P_i$ with $Mg^{2+}$ in front (facing O1-P bond) and SER203 on the right side (P-O2 toward right side) as it is the case in panel **(C)** and **(D)**, position O3 is pointing up and position O4 points back down. Panel **(B)** and **(D)** are view with the same orientation of ADP.

Table 3.1: Interactions (cf. details on interactions analysis in section §3.4.5) between the four oxygen atoms of $P_i$ and the surrounding residues of the active site, from the six initial protonation states of PPS and PiR. The protonated atoms of the initial twelve configurations and in the corresponding final relaxed configurations are recalled between parenthesis in the second column, the arrow meaning the transformation. The initial interactions before relaxation are given in the lines PPS-*crystal* and $P_i$R-*crystal*. Residue's number is given in bracket when the residue does not form H-bond (criteria of distance X-H-Y less than 3 Å and angle 180+/-45°) with the oxygen but is in position very close. "–" means that no residues are found close enough to the oxygen to be in interaction.

| *Structure* | *configuration* | O1 | O2 | O3 | O4 |
|---|---|---|---|---|---|
| PPS-*crystal* | – | $Mg^{2+}$ | $H_2O$ | 153, 203 | 157, 459 |
| PPS-A | (3,4)→(3,4) | $Mg^{2+}$ | 203 | ADP | 153, (157), (459) |
| PPS-B | (2,4)→(2,4) | $Mg^{2+}$, 204 | $H_2O$ | 153, 203 | 157, ADP, (459) |
| PPS-C | (2,3)→(2,3) | $Mg^{2+}$, 204 | 203, $H_2O$ | ADP, 153 | 157, 459 |
| PPS-D | (1,4)→(1,2) | $Mg^{2+}$, ADP | – | 153, 203 | 157, 459 |
| PPS-E | (1,3)→(1,2) | $Mg^{2+}$, ADP | 204, $H_2O$ | 153, 203, $H_2O$ | 157, 459 |
| PPS-F | (1,2)→(1,2) | $Mg^{2+}$, ADP | – ,(204) | 153, 203 | 157, (459) |
| $P_i$R-*crystal* | – | $Mg^{2+}$ | $H_2O$, (203), (204) | ADP, 157, (153) | (157), (458) |
| $P_i$R-A | (3,4)→(3,4) | $Mg^{2+}$ | 203, $H_2O$ | ADP, 153 | (157), $H_2O$ |
| $P_i$R-B | (2,4)→(2,3) | $Mg^{2+}$, 204 | (203), (204) | ADP, (153), 203 | 157, (459) |
| $P_i$R-C | (2,3)→(2,3) | $Mg^{2+}$ | 203, (204) | ADP, 153 | 157, 459 |
| $P_i$R-D | (1,4)→(1,2) | $Mg^{2+}$, ADP | – | 153, 203 | 157, 459 |
| $P_i$R-E | (1,3)→(1,2) | $Mg^{2+}$, ADP | 204, $H_2O$ | 153, 203 | 157, $H_2O$ |
| $P_i$R-F | (1,2)→(1,2) | $Mg^{2+}$, ADP | 204, $2\,H_2O$ | 153, 203 | 157, $H_2O$ |

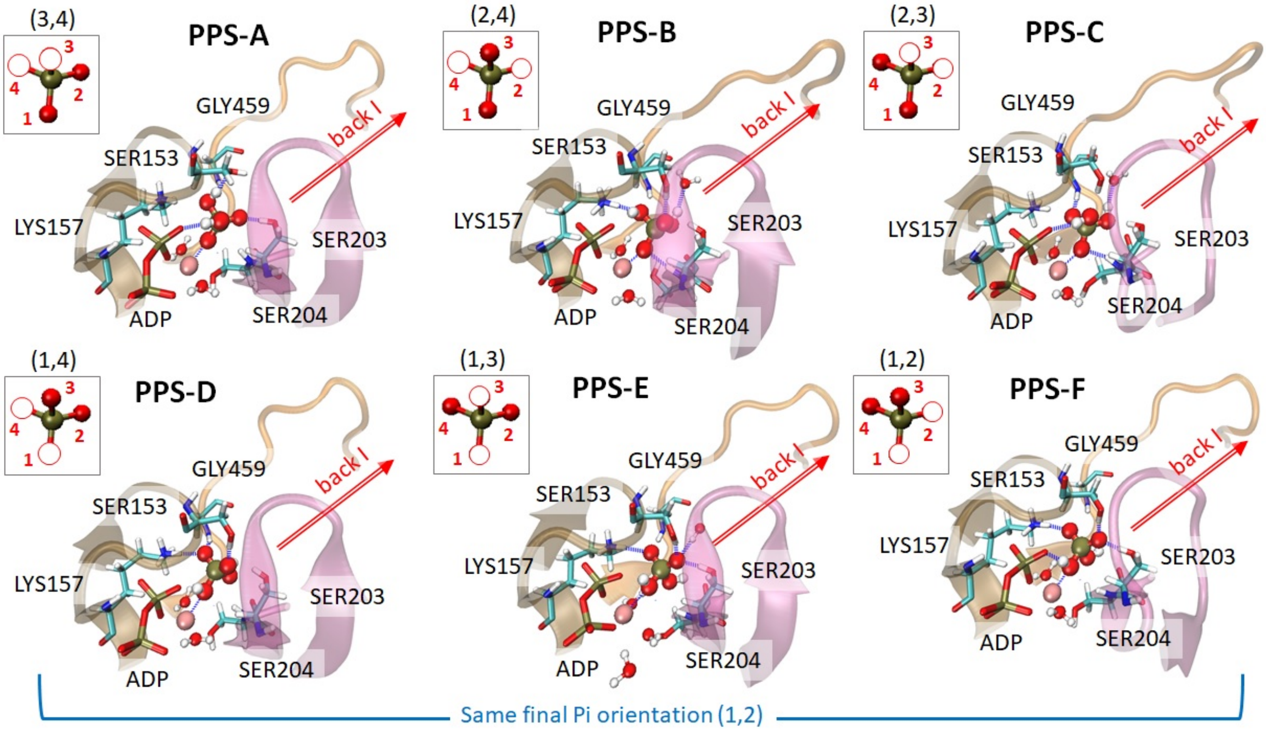


Figure 3.4: Representation of the Nucleotide Binding Loops at the end of relaxation for six initial PPS configurations, depending on the protonation state of $P_i$. For each configuration (PPS-A to PSS-F), the relaxed active site structure is shown with the coloring of the different structural element similar to Fig. 3.3. The P-Loop (brown), Switch I (pink) and Switch II (orange) are represented in cartoon. ADP without the nucleoside part, $Mg^{2+}$, SER153 and LYS157 of P-Loop, SER203 and SER204 of Switch I, GLY459 of Switch II are represented in licorice with color code C cyan, H white, O red, N blue, P brown, and $Mg^{2+}$ light pink. The $P_i$ and water molecules at a distance less than 4 Å from the $P_i$ or $Mg^{2+}$ are represented in CPK with the same color code. For comparison between the $P_i$ orientation before and after relaxation for the six considered models PPS-(A-F), the initial $P_i$ orientation is shown in the inserts in the same orientation as Fig. 3.3**B**. In the insert, the protonated atoms are indicated by white circles. The numbers 1,2,3 and 4 of the insert correspond to the oxygen position described in Fig. 3.3. The hydrogen bonds formed between $P_i$ and the surrounding residues, which are detected by interactions analysis (cf. section §3.4.5) are indicated by the blue dotted lines. The interactions between $P_i$ and $Mg^{2+}$ are also shown in blue dotted lines. The backdoor I route (back I) is represented by a red double arrow.

the two protonated atoms are indicated by a white circle. The hydrogen bonds (H-bonds) are indicated by the blue dotted lines.

**Change in $P_i$ orientation after relaxation** The main observation is that the orientation of $P_i$ is the same at the beginning and at the end of relaxation for configurations PPS-(A, B, C, and F) while the configurations PPS-(D and E) change during the relaxation to a configuration close to PPS-F, see Fig. 3.4 and Tab. 3.1. The configurations PPS-(D, E and F) are the configurations starting with a protonated O1 oxygen atom, the one that is in direct interaction with $Mg^{2+}$. The configuration PPS-D changes to be alike configuration PPS-F only at the end of the 200 ns dynamics while the configuration PPS-E shows the rotation of the $P_i$ already at the beginning of the dynamics to be alike PPS-F.

**Changes in the $Mg^{2+}$ coordination** One peculiarity of the PPS-E configuration is the fact that the coordination of $Mg^{2+}$ changes: a water molecule replaces SER204 and ADP $\alpha$-Phosphate replaces a water molecule in the coordination of $Mg^{2+}$, see Fig. B.2 and Fig. B.3 in appendix B. This leads to a double coordination of the $Mg^{2+}$ by two oxygen atoms of ADP.

Looking at the Table 3.1, we also observe that the dynamics from configuration PPS-(B and C), which have no protonated O1 and a protonated O2 lead to O1 interacting with $Mg^{2+}$, like in the others relaxed structure, but also with SER204. Their O2 forms a H-bond with a water molecule.

**Changes in the $P_i$ orientation and interactions along the dynamics** The observation of the protein motion during the relaxation simulation shows that the $P_i$ maintains the same orientation for configurations PPS-(A, B, C and F), see Video 1[1]. The $P_i$ orientation only changes once during the relaxation simulations PPS-(D and E), at the beginning for PPS-D and at the end for PPS-E, as mentioned above. Furthermore, the analysis of interactions (see §3.4.5 for details) demonstrates that for all configurations the $P_i$ interacts with oxygen of ADP $\beta$-Phosphate and five residues: SER153 and LYS157 of P-Loop, SER203 and SER204 of Switch I and GLY459 of Switch II. There is also one weaker interaction with the ASN200 residue of Switch I. The color chart of Fig. B.5 in appendix B shows the evolution of these interactions along the relaxation simulation. In particular, the interactions of $P_i$ with SER153 and SER203 residues disappear once for a few dozens of nanosecond for PPS-A configuration, at around times 60 ns and 150 ns of the dynamics, respectively. The interaction of $P_i$ with SER153 also disappears once for a few dozens of nanosecond for PPS-C configuration (around time 80 ns).

**Evolution of the state of the salt-bridges along the dynamics** The evolution of critical and secondary salt bridges can be vizualized in Video 2[2]. The color chart (see Fig. B.7 in

---

[1] Video 1 available online: https://www.youtube.com/watch?v=FGezG-ft0g0

[2] Video 2 available online: https://www.youtube.com/watch?v=SOsQwnB90wY

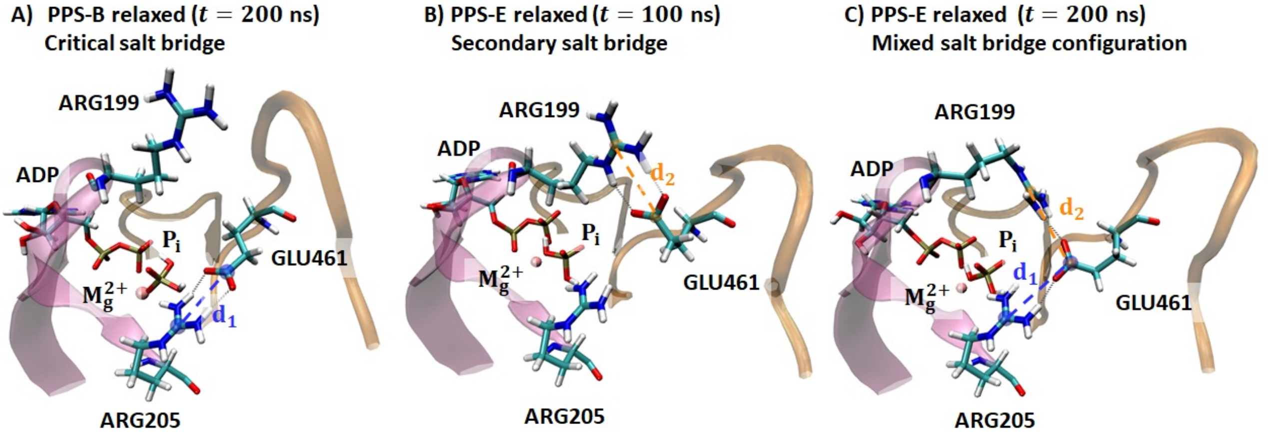


Figure 3.5: Definition of distance $d_1$ characterizing the state of the critical salt bridge (205@CZ - 461@CD) (blue) and $d_2$ the secondary salt bridge (199@CZ - 461@CD) (orange). The P-Loop (brown), Switch I (pink) and Switch II (orange) are represented in cartoon in the view from backdoor I of active site. $P_i$, ADP, $Mg^{2+}$, ARG199 and ARG205 of Switch I and GLU461 of Switch II are represented in licorice with color code C cyan, H white, O red, N blue, P brown, and $Mg^{2+}$ light pink. The hydrogen bonds formed between ARG199, ARG205 and GLU461 residues are indicated by the gray dotted lines. **A)** Critical salt bridge ($d_1 \approx 4$ Å and $d_2 \approx 8$ Å). **B)** Secondary salt bridge ($d_1 \approx 8$ Å and $d_2 \approx 4$ Å). **C)** Mixed salt bridge configuration ($d_1 \approx 4$ Å and $d_2 \approx 4$ Å).

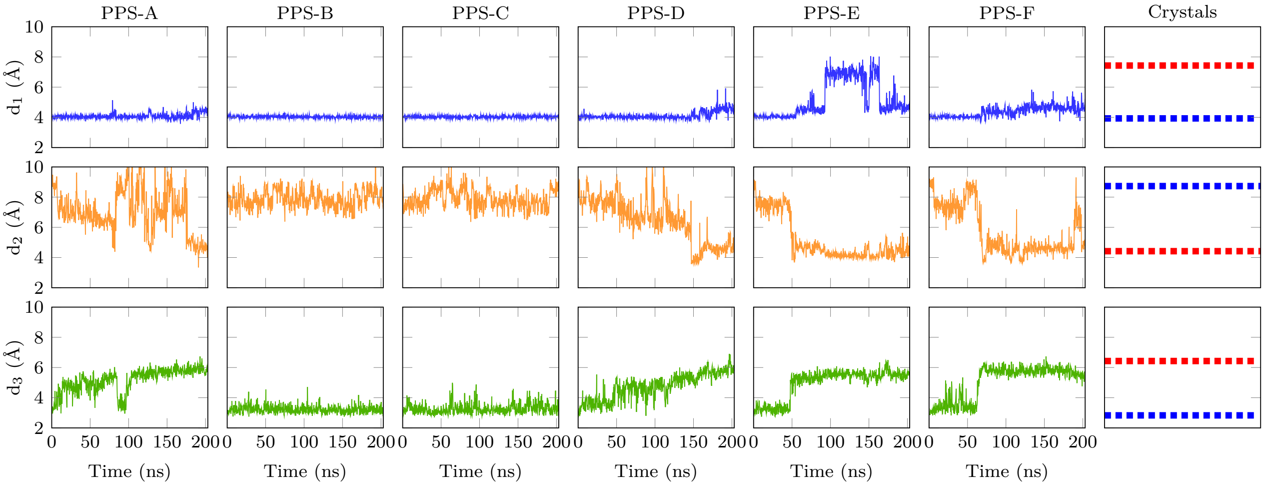


Figure 3.6: Evolution of the distances characterizing the states of the critical salt bridge $d_1$ corresponding to the distance (205@CZ - 461@CD) (blue) and of the secondary salt bridge $d_2$ corresponding to the distance (199@CZ - 461@CD) (orange) along the relaxations of the PPS structure (see Fig. 3.6 for illustration of distances $d_1$ and $d_2$). The distance between Switch II and P-Loop is monitored by the distance $d_3$ (460@O - 153@N) (green), shown on Fig. B.9 in appendix B. In the last column is marked the corresponding distances in the $P_iR$ crystal (red) and in the PPS crystal (blue).

appendix B), which provides the analysis of the interactions of NBL residues, highlights the breaking of the critical salt bridge in favor of the creation of the secondary salt bridge or a mixed salt bridge configuration. These three configurations are illustrated in Fig. 3.5.

The states of the critical and secondary salt bridges are further investigated by monitoring the distance ARG205@CZ - GLU461@CD, $d_1$, and ARG199@CZ - GLU461@CD, $d_2$, respectively, along relaxation dynamics, see Fig. 3.5. During the 200 ns of relaxation, the critical salt bridge remains closed (distance 205@CZ - 461@CD, $d_1 \approx 4\,\text{Å}$) and secondary salt bridge remains open (distance 199@CZ - 461@CD, $d_2 \approx 8\,\text{Å}$) in configurations PPS-B and PPS-C. In these cases, the values of the characteristic distances coincide with the values measured in the PPS crystallographic structure. In the dynamics starting with configurations A, D and F, the critical salt bridge is less stable than in B and C, as small fluctuations appear during relaxation.

We observe a decrease of the distance 199@CZ - 461@CD ($d_2$) from 8 Å to around 4 Å in configuration PPS-(A, D, E and F). This decrease coincides with the onset of fluctuations of both $d_1$ and $d_2$ distances, though the latter fluctuates more than the former.

For these four protonation states, the “mixed salt bridge” conformation involving residues ARG195, ARG205 and GLU461 (see Fig. 3.5C) is observed. It appears when $d_1 \approx d_2 \approx 4\,\text{Å}$. This intermediate conformation between the critical salt bridge and the secondary salt bridge can be detected on the color chart of Fig. B.7 in appendix B, when ARG205-GLU461 and ARG199-GLU461 interactions are present simultaneously. The secondary salt-bridge ($d_1 \approx 8\,\text{Å}$ and $d_2 \approx 4\,\text{Å}$) is observed in PPS-E during around 60 ns.

Furthermore, the apparition of $d_1$ and $d_2$ fluctuations is systematically coupled to an increase of the 460@O - 153@N distance, $d_3$. This last distance gives an estimate of the distance between Switch II and P-Loop (see Fig. B.9 in appendix B), therefore the observed increase suggests an opening of Switch II.

#### 3.2.1.2 Relaxation of $\mathrm{P_iR}$ structures

The same six protonation configurations have been tested with the $\mathrm{P_iR}$ structure. The corresponding structures obtained after relaxation are shown in Fig. 3.7 and the interactions of the oxygen atoms of $\mathrm{P_i}$ are given in Table 3.1.

**Coordination of $\mathbf{Mg^{2+}}$** Our first observation is that the coordination of $Mg^{2+}$ is do not change during relaxation, except for the $\mathrm{P_iR}$-A configuration. In most of the relaxed structures as in the initial structures, the first coordination sphere of $Mg^{2+}$ is composed of 2 water molecules, SER204, THR158, the $\mathrm{P_i}$ and one oxygen from ADP $\beta$-Phosphate. We also observe that, in $\mathrm{P_iR}$-A relaxed structure, ADP is coordinated twice to $Mg^{2+}$ (with $\beta$- and $\alpha$-Phosphate), while SER204 is not coordinated to $Mg^{2+}$ anymore (see Fig. B.4 in appendix B. This special coordination, where SER204 is not anymore coordinated to $Mg^{2+}$, is the same as observed in PPS-E relaxed structure.

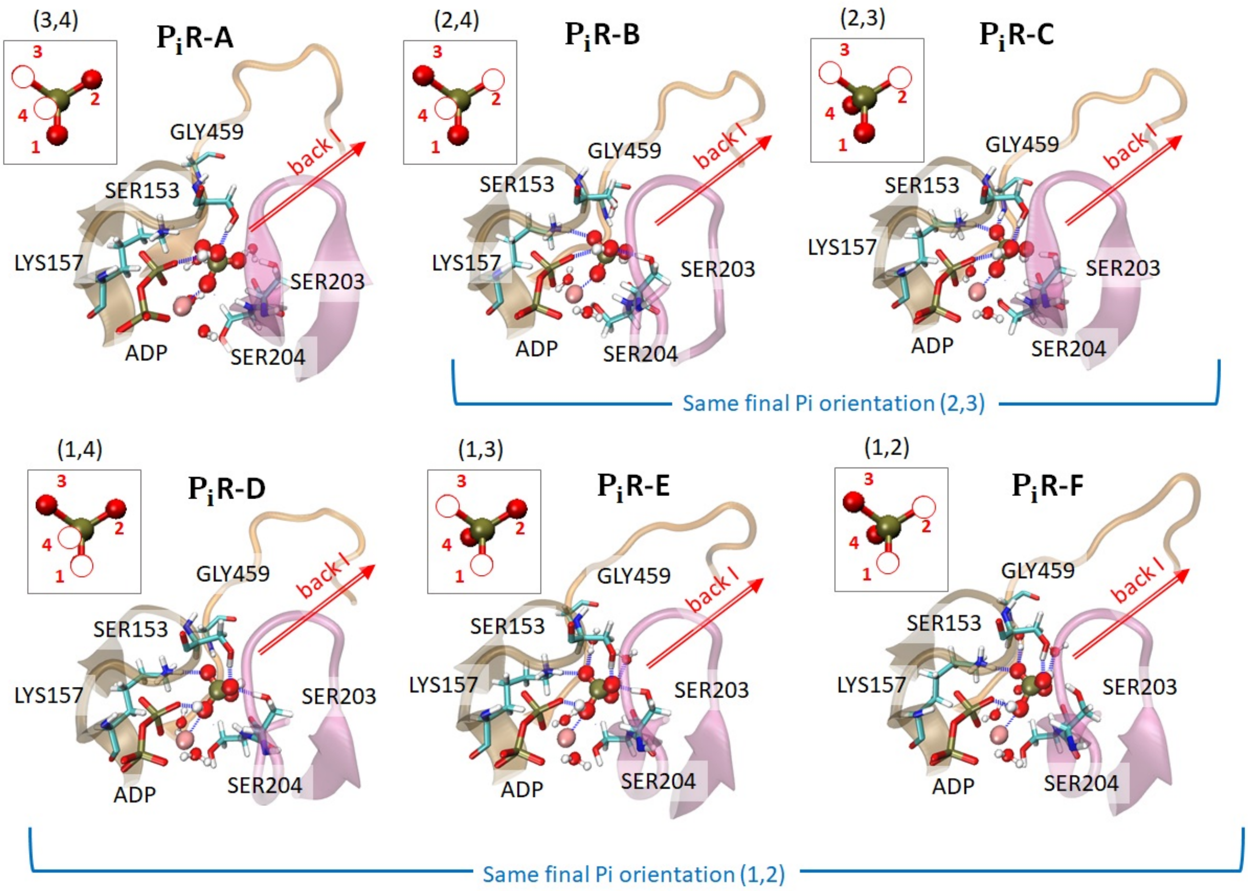


Figure 3.7: Representation of the Nucleotide Binding Loops at the end of relaxation for the six initial $P_i$R configurations ($P_i$R-A to $P_i$R-F).For each configuration, the relaxed active site structure is shown with the coloring of the different structural element similar to Fig. 3.3. The P-Loop (brown), Switch I (pink) and Switch II (orange) are represented in cartoon. ADP without the nucleoside part, Mg$^{2+}$, SER153 and LYS157 of P-Loop, SER203 and SER204 of Switch I, GLY459 of Switch II are represented in licorice with color code C cyan, H white, O red, N blue, P brown, and Mg$^{2+}$ light pink. The $P_i$ and water molecules at a distance less than 4 Å from the $P_i$ or Mg$^{2+}$ are represented in CPK with the same color code. For comparison between the $P_i$ orientation before and after relaxation for the six considered configurations $P_i$R-(A-F), the initial $P_i$ orientation is shown in the inserts in the same orientation as Fig. 3.3**D**. In the insert, the protonated atoms are indicated by white circles. The numbers 1,2,3 and 4 of the insert correspond to the oxygen position described in Fig. 3.3. The hydrogen bonds formed between $P_i$ and the surrounding residues, which are detected by interactions analysis (cf. section §3.4.5) are indicated by the blue dotted lines. The interactions between $P_i$ and Mg$^{2+}$ are also shown in blue dotted lines. The backdoor I route (back I) is represented by a red double arrow.

**Changes in $P_i$ orientation** A change in $P_i$ orientation is seen for half of the simulations: $P_i$R-(B, D and E). After relaxation, the orientation of $P_i$ in $P_i$R-B is the same as in $P_i$R-C, and the orientations of $P_i$ in $P_i$R-(D and E) are the same as in $P_i$R-F. Finally, only three different orientations are observed after relaxation: (3,4) adopted by $P_i$R-A; (2,3) adopted by $P_i$R-(B and C); and (1,2) adopted by $P_i$R-(D, E and F).

In the three final orientations [(1,2), (2,3), and (3,4) ], one of the two protonated oxygen points toward an oxygen from ADP $\beta$-Phosphate, suggesting that this interaction participates in the stabilization of the inorganic phosphate.

The final orientation (1,2) is obtained from the initial configurations $P_i$R-(D, E and F), which are characterized by an initial structure with the protonated O1 oxygen atom pointing at the $Mg^{2+}$ atom. After relaxation, the same protonated oxygen atom still points toward the $Mg^{2+}$ atom, but the other protonated oxygen is rotated such as to point toward the backdoor I (in opposite direction of the $Mg^{2+}$ like it is in the initial structure of $P_i$R-F).

The final orientation (2,3) is reached from the initial configurations $P_i$R-(B and C). In this final configuration, in contrast to orientation (1,2), the oxygen atom pointing toward the $Mg^{2+}$ atom is not protonated. Instead, the two protonated oxygen atoms point in the direction of the SER204 backbone oxygen and in the direction of an oxygen from ADP $\beta$-Phosphate, respectively, see Fig. 3.7.

Finally, the orientation (3,4), which is obtained only from configuration $P_i$R-A, is also characterized by a non-protonated oxygen atom pointing toward the $Mg^{2+}$ atom, see Fig. 3.7. The protonated oxygen atoms point toward an oxygen from the ADP $\beta$-Phosphate, and toward a water molecule in the direction of ILE457.

These observations suggest that when a protonated oxygen points toward $Mg^{2+}$, $P_i$ always equilibrates toward orientation (1,2), independently of the initial orientation of the other protonated oxygen atom. Moreover, the protonation of O1 favors the re-orientation of the other protonated oxygen in the direction of the backdoor (position 2), *i.e.* leading to the orientation (1,2). Conversely, in the case where a non-protonated oxygen coordinates with $Mg^{2+}$, the phosphate is less inclined to turn.

**Evolution of interactions along the dynamics** The analysis of the relaxation dynamics shows that the $P_i$ maintains the same orientation along the simulation of configurations $P_i$R-(A, C and F), see Video 3[3]. The $P_i$ orientation only changes once at the beginning of the relaxation simulations of $P_i$R-(B, D and E) , and remains constant afterward. The interactions' analysis (see §3.4.5 for details) demonstrates that for all structures the $P_i$ interacts with the oxygen atom of ADP $\beta$-Phosphate and at least four residues: SER153 and LYS157 of P-Loop, SER203 and SER204 of Switch I. A weaker interaction with ASN200 residue of Switch I is also observed for all configurations. The color chart of Fig. B.6 in appendix B shows the evolution of these interactions along the relaxation simulation. It shows in particular that an interaction of $P_i$

---

[3] Video 3 available online: https://youtu.be/zYb2s9wJphY

with GLY459 residue of Switch II appears at the middle of the relaxation corresponding to configurations $P_i$R-(B, C and D), which causes a slight closure of Switch II (see $d_3$ decrease on Fig. 3.8 and on Video 3[4]).

It is interesting to note that GLY459 residue interacts with GLY151 residue of P-Loop in the other configurations, $P_i$R-(A, E and F) (see B.8 in appendix B). An interaction of $P_i$ with ARG205 also appears for configurations $P_i$R-(B and C), but we do not observe any significant correlation between the presence of this interaction and the state of the critical salt bridge (present in $P_i$R-C) and secondary salt bridge (present in $P_i$R-B) (see Fig. 3.8 and color chart in Fig. B.8 in appendix B).

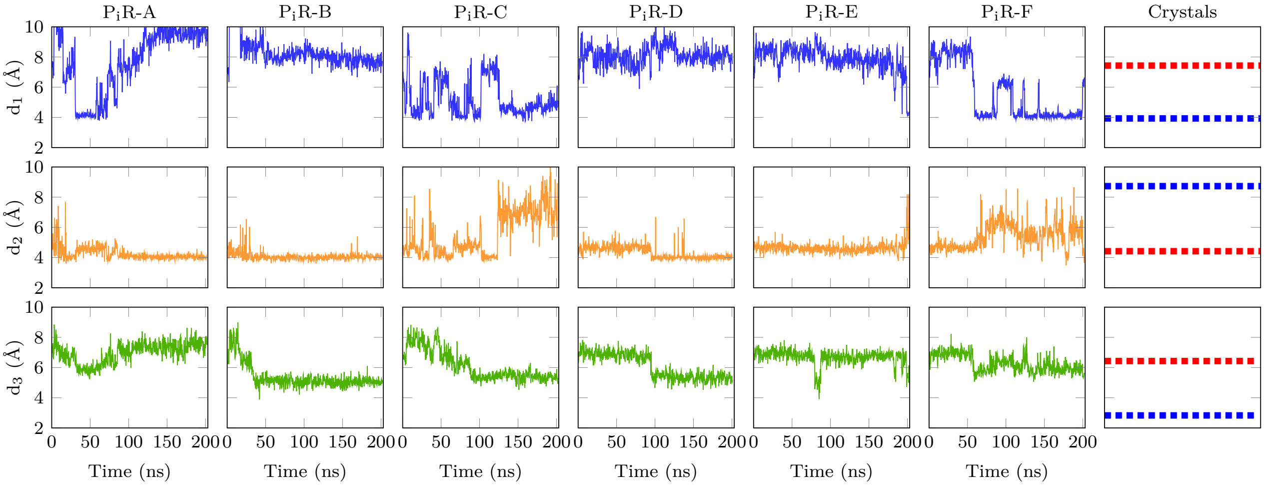


Figure 3.8: Evolution of the distances characterizing the state of the critical salt bridge (205@CZ - 461@CD) and the secondary salt bridge (199@CZ - 461@CD) along the relaxations of the PiR structure, and see Fig. 3.6 for illustration of distances $d_1$ and $d_2$. The distance between Switch II and P-Loop is monitored by the distance (460@O - 153@N) (green), $d_3$ shown on Fig. B.9 in appendix B.

**State of the salt bridges along the dynamics** The evolution of the state of the critical and secondary salt bridges are illustrated in Fig. 3.8 and on Video 4[5]. The main observation is that the state of the salt bridges in $P_i$R configurations is less stable than in PPS configurations. The critical salt-bridge remains mostly in an open configuration for configurations $P_i$R-(A, B, D and E). For configuration $P_i$R-(A, C and F), the interactions of GLU461 fluctuate between the two arginines ARG205 and ARG199, which is the signature of a mixed salt bridge configuration ($d_1 \approx 4\,\text{Å}$ and $d_2 \approx 4\,\text{Å}$). During the last quarter of the simulation of $P_i$R-C configuration, the critical salt bridge ($d_1 \approx 4\,\text{Å}$ and $d_2 \approx 8\,\text{Å}$) is formed, although less stable than for PPS structures. Like observed for the PPS configurations, the value of the 205@CZ - 461@CD distance, $d_3$, characterizing the open and close states coincide with the value measured in the

---

[4] Video 3 available online: https://youtu.be/zYb2s9wJphY

[5] Video 4 available online: https://youtu.be/2IHEFeia-go

$\mathrm{P_i}$R crystal for all configurations, despite the slight decrease of $d_3$ observed during the $\mathrm{P_i}$R-(B, C and D) relaxation simulations.

### 3.2.2 Effect of $\mathrm{P_i}$ protonation state on the release mechanism and energetics

The effect of the $\mathrm{P_i}$ orientation on its escape from the active site was investigated by Umbrella Sampling (US) using the distance between the center of mass of $\mathrm{P_i}$ and $\mathrm{Mg}^{2+}$— denoted $d_{P_\mathrm{i}-\mathrm{Mg}^{2+}}$ — as the control collective variable. The $d_{P_\mathrm{i}-\mathrm{Mg}^{2+}}$ distance is increased by steps of 0.3 Å from 3.5 to 20 Å. Each window lasts 3 ns during which the system relaxes freely. Twelve simulations were performed starting from the final structures resulting from the unconstrained relaxation dynamics discussed in the previous section. For each simulation PPS-(A-F) and $\mathrm{P_i}$R-(A-F), the associated potential of mean force (PMF) is computed (see Fig.3.9), and the interactions analysis is carried out during the escape process (see color charts in Fig. B.11 to Fig. B.14 in appendix B).

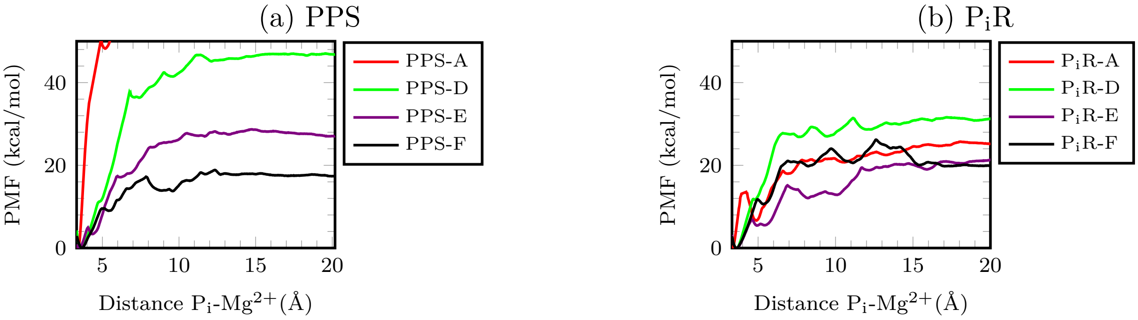


Figure 3.9: Potential of mean force (PMF) for $\mathrm{P_i}$ departure for each obtained structure after relaxation of different initial orientations starting from PPS crystallographic structure PPS-(A to F) (a) and $\mathrm{P_i}$R crystallographic structure $\mathrm{P_i}$R-(A to F) (b). The PMF of simulation PPS-B, PPS-C, $\mathrm{P_i}$R-B and $\mathrm{P_i}$R-C are not shown as the $\mathrm{P_i}$ is not released. The PMF of simulation PPS-A is very rapidly above 50 kcal/mol. The complete curve is available in the in appendix B on Fig. B.10

#### 3.2.2.1 $\mathrm{P_i}$ Release from the PPS configurations

The PMF profiles for the six PPS simulations are shown in Fig. 3.9**a**. The PMF associated with PPS-B and PPS-C are not represented in Fig. 3.9**a** since in these two simulations the departure of $\mathrm{P_i}$ from the active site was not observed. For the four other simulations, the free energy value rapidly increases then shows a series of local minima in the path of the $\mathrm{P_i}$ departure. Then, the PMF profile reaches a plateau for large $\mathrm{P_i}$-$\mathrm{Mg}^{2+}$ distance ($d_{P_\mathrm{i}-\mathrm{Mg}^{2+}} \geq 12$ Å). The free energy barrier for PPS-A is above 50 kcal mol$^{-1}$, the one of PPS-D around 48 kcal mol$^{-1}$, the one of PPS-E around 28 kcal mol$^{-1}$, and the one of PPS-F around 18 kcal mol$^{-1}$. We notice that the three lowest PMF profiles correspond to US simulations starting with structures PPS-(D,E,F), which all have O1 and O2 atoms protonated (structures (1,2)). This (1,2) configuration also shows an interaction of unprotonated O3 oxygen with SER203, and an interaction of hydrogen of O1 with oxygen of $\beta$-Phosphate, contrary to the others (see Table 3.1).

In contrast, the configurations where O1 is unprotonated and O2 is protonated (PPS-(B,C)) are very stable, which could be explained by an energy barrier much higher than in the other configurations in the condition of the present US simulations. In addition, in the relaxed PPS-(B,C) structures, O1 is coordinated with Mg$^{2+}$ and makes a H-bond with the N-H of the backbone of SER204. Beside, the protonated O2 makes a H-bond with an oxygen of a water molecule whose hydrogen atoms are involved in H-bonds with the carboxylate group of GLU461 and backbone C=O of GLY459. These last interactions are not seen in the other structures, see table 3.1.

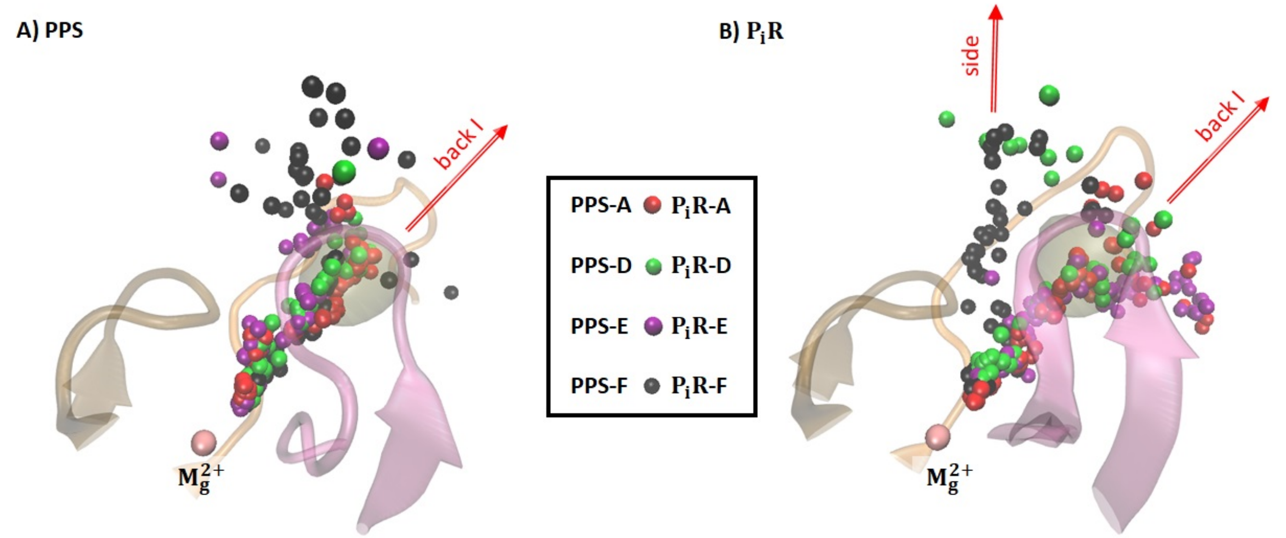


Figure 3.10: Structural representation of the calculated $P_i$ release pathways by colored spheres for the US simulations in the PPS-(A, D, E and F) and $P_i$R-(A, D, E and F) structures. The pathways of structures PPS-B, PPS-C, $P_i$R-B and $P_i$R-C are not shown as the $P_i$ is not released. The P-Loop (brown), Switch I (pink) and Switch II (orange) are represented in cartoon for PPS-A (right) and $P_i$R-A (left). Mg$^{2+}$ is represented in light pink. The volume of the $P_i$ in $P_i$R1 crystallographic structure (4PJN) (Llinas et al., 2015) is represented in transparent tan surface to show the passage of $P_i$ by this position. All structures are superposed on U-50 (179 to 193 and 227 to 431 residues), on Switch I (194 to 204 residues) and on L-50 (510 to 597 residues). The backdoor I route (back I) and sidedoor route (side) are represented by red double arrows.

The $P_i$ release pathway is shown in Fig. 3.10A) for the PPS-(A, D, E, F) US simulations. The color dots representing the $P_i$ trajectory during the US simulations bring to light the tube from active site to the inner cleft, which corresponds to the backdoor I route. At the end of the tube, the $P_i$ passes through its position in $P_i$R1 crystallographic structure (4PJN) of Llinas et al (Llinas et al., 2015), which is represented by the transparent tan surface in Fig. 3.10. Looking at the Mg$^{2+}$ from backdoor I, the $P_i$ goes through a window formed by Switch I on the left, ARG199 residue on the top, GLY461 on the right and ARG205 on the bottom (in concordance with the view used in Fig. 3.5). The movements of critical salt bridge and secondary salt bridge during the $P_i$ release are represented on Video 5[6]. A mixed salt bridge is observed for all configurations where $P_i$ is released, *i.e.* PPS-(A, D, E, F) at the beginning of the simulation. For the two simulations that do not allow the leaving of the $P_i$ (PPS-(B,C)) the critical salt bridge is stable all along the simulation. In PPS-(A, D, E, F), $P_i$ interacts with ARG199 residue before leaving the cavity leading to a partially formed critical salt bridge.

The results of interactions analysis are summarized in Table 3.2 and represented on the colors charts in Suppl. Info. and on Video 6[7]. At the beginning of US simulation, the $P_i$ interacts with $Mg^{2+}$, $\beta$-Phosphate, SER153 and LYS157 residues of P-Loop, SER203 and SER204 of Switch I, and GLY459 residue of Switch II. We define the active site detachment when the interaction between $P_i$ and $Mg^{2+}$ is broken. At the same moment, the $P_i$ interactions with LYS157, ASN200, SER205 and GLY459 residues also fade. This event corresponds to the highest jump of free energy on the PMF (see Fig. 3.9**a**). The jump is more important for the PPS-A where the O1 is unprotonated. The two other simulations with unprotonated O1, PPS-(B,C) do not let the $P_i$ leave the cavity. These two simulations have in common with PPS-A the presence of ASN200 not far from $P_i$ at the beginning of the simulation.

After the $P_i$ detachment, the $P_i$ goes through a tube to the PiR1 structure (see Fig. 3.10A), where it interacts with SER153 of P-Loop, THR197, ARG199, SER203, ARG205 of Switch I and GLU461 of Switch II. There is also an interaction with ASN200 in PPS-D, which could explain the high free energy barrier of PPS-D. The H-bond between ASN200 and $P_i$ occurs around $d_{P_i-Mg} \approx 8\,\text{Å}$, corresponding to the first decline of PMF.

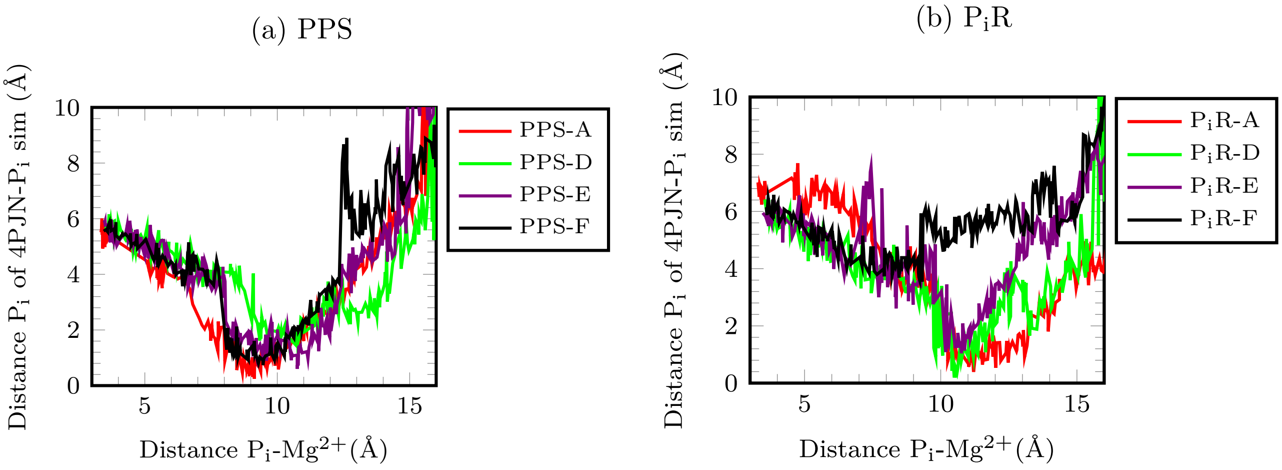


Figure 3.11: Distance between the phosphorus atom of $P_i$R in 4PJN and the phosphorus atom of $P_i$R along the US simulations after superimposition of the NBL of the structures. (a) for PPS US simulations, (b) for $P_i$R US simulations.

To identify if the $P_i$ is leaving the active site by the backdoor I and explore the position $P_i$R1 of (Llinas et al., 2015) more quantitatively, we have superimposed the NBL of the 4PJN

---

[6] Video 5 available online: https://www.youtube.com/watch?v=F7ZQabnIHI8

[7] Video 6 available online: https://youtu.be/5iENrbbYXCg

Table 3.2: Residues interacting with $P_\mathrm{i}$ identified in the PPS and $P_\mathrm{i}$R for different protonation states of investigated myosin VI structures during the US simulations. The residues close to the $P_\mathrm{i}$ in the active site at the beginning of US simulations are written in bold and those close when $P_\mathrm{i}$ is near the position it occupies in 4PJN (*i.e.* $d_{P_\mathrm{i}-Mg} \in [8\,\text{Å} \text{ to } 10\,\text{Å}]$ for PPS and $d_{P_\mathrm{i}-Mg} \in [10\,\text{Å} \text{ to } 12\,\text{Å}]$ for $P_\mathrm{i}$R) are underlined.

| | | Protonation | | |
|---|---|---|---|---|
| State | Elements | A | B | C |
| PPS | P-Loop | **SER153**,<br>**LYS157** | **SER153**,<br>**LYS157** | **SER153**,<br>**LYS157** |
| | Switch I | THR197, ARG199,<br>**ASN200**,<br>**SER203**,<br>**SER204**, ARG205 | **ASN200**,<br>**SER203**,<br>**SER204**,<br>**ARG205** | **ASN200**,<br>**SER203**,<br>**SER204**,<br>**ARG205** |
| | Switch II | **GLY459**, GLU461 | **GLY459** | **GLY459** |
| | other | GLU242 | | |
| | Path | back I | – | – |
| $P_\mathrm{i}$R | P-Loop | **SER153**,<br>**LYS157** | **SER153**,<br>**LYS157** | **SER153**,<br>**LYS157** |
| | Switch I | THR197, ARG199,<br>**ASN200**,<br>**SER203**,<br>**SER204**, ARG205 | **ASN200**,<br>**SER203**,<br>**SER204**,<br>**ARG205** | **ASN200**,<br>**SER203**,<br>**SER204**,<br>**ARG205** |
| | Switch II | – | **GLY459** | **GLY459** |
| | other | GLU242 | – | – |
| | Path | back I | – | – |
| State | Elements | D | E | F |
| PPS | P-Loop | **SER153**,<br>**LYS157** | **SER153**,<br>**LYS157** | **SER153**,<br>**LYS157** |
| | Switch I | THR197,<br>ARG199, ASN200,<br>**SER203**,<br>**SER204**, ARG205 | THR197,<br>ARG199, ASN200,<br>**SER203**,<br>**SER204**, ARG205 | THR197, ARG199,<br>**SER203**,<br>**SER204**, ARG205 |
| | Switch II | **GLY459**, GLU461 | **GLY459**, GLU461 | **GLY459**,<br>PHE460, GLU461 |
| | other | – | – | – |
| | Path | back I | back I | back I |
| $P_\mathrm{i}$R | P-Loop | **SER153**,<br>**LYS157** | **SER153**,<br>**LYS157** | **SER153**,<br>**LYS157** |
| | Switch I | THR197, ARG199,<br>**SER203**,<br>**SER204**, ARG205 | THR197, ARG199,<br>**SER203**,<br>**SER204**, ARG205 | ARG199,<br>**SER203**,<br>**SER204**, ARG205 |
| | Switch II | **GLY459** | – | GLY459, PHE460 |
| | other | GLU242, GLN470 | GLU242 | – |
| | Path | back I | back I | side |

crystal structure of $\mathrm{P_iR1}$ on the NBL of 336 snapshots along US simulations and monitored the distance between the phosphorus of the snapshots with the one of $\mathrm{P_iR1}$. This distance is plotted against $d_{P_{\mathrm{i}}-Mg}$ in Fig. 3.11. For each configuration the minimum of the curve indicates the value of $d_{P_{\mathrm{i}}-Mg}$ for which the similarity between the simulation snapshot and the 4PJN structure is highest. This minimum is observed at $d_{P_{\mathrm{i}}-Mg} \approx 8\,\text{Å}$ to $10\,\text{Å}$ for PPS-(A,F), at slightly larger values ($d_{P_{\mathrm{i}}-Mg} \approx 8\,\text{Å}$ to $11\,\text{Å}$) for PPS-E, and $9\,\text{Å}$ to $11\,\text{Å}$ for PPS-D, see Fig. 3.11(a) By monitoring the interactions of $\mathrm{P_i}$ with the residues along the exit tunnel and the most frequent interactions between residues of the NBL, we find that the interactions of $\mathrm{P_i}$ with residues SER153, THR197, SER203, ARG205, and GLU461 described by Llinas in the $\mathrm{P_iR1}$ appear as the predominant interactions in the four US simulations PPS-(A, D to F) where the $\mathrm{P_i}$ leaves the cavity (see Fig. B.11 and Fig. B.13 in appendix B). In addition, the interactions seen in the $\mathrm{P_iR1}$ of Llinas are all present when the $\mathrm{P_i}$ is around 8 to 11 Å from $\mathrm{Mg^{2+}}$ which is in concordance with the results concluded from Fig. 3.11. By comparing the PMF profiles of Fig. 3.9(a) with the minimum of curve in Fig. 3.11a, we obtain that for PPS-F, $\mathrm{P_iR1}$ structure found by Llinas et al corresponds to a minimum in the PMF profile. The same observation is found for PPS-A (see Fig B.10 in appendix B).

Finally, the $\mathrm{P_i}$ leaves the cavity defined in $\mathrm{P_iR1}$ structure to reach the cleft ($d_{P_{\mathrm{i}}-Mg} \approx 10\,\text{Å}$), when the interactions with SER153 and SER203—initially present in active site—stop. Once released the $\mathrm{P_i}$ continues to interact with ARG199 residue outside the myosin.

These observations suggests that, for the four PPS US simulations effectively leading to $\mathrm{P_i}$ escape from the active site, the $\mathrm{P_i}$ leaves the cavity through backdoor I, transiently exploring the $\mathrm{P_iR1}$ structure obtained by (Llinas et al., 2015).

#### 3.2.2.2 $\mathrm{P_i}$ release from the PiR configurations

The PMF obtained for the $\mathrm{P_iR}$ umbrella sampling simulations are presented in Fig. 3.9. The four structures starting from $\mathrm{P_iR}$ and showing a release of $\mathrm{P_i}$ have a barrier bellow $30\,\mathrm{kcal\,mol^{-1}}$. As in the PPS case, the simulations starting with $\mathrm{P_iR}$-(B, C), which both have orientation (2,3), show no release of the $\mathrm{P_i}$. The configuration $\mathrm{P_iR}$-A characterized by orientation (3,4), i.e. without a protonated O1, shows a low barrier. As PPS-A, $\mathrm{P_iR}$-A relaxed structure has the configuration (3,4) but the $\mathrm{Mg^{2+}}$ coordination has changed during the relaxation, leading, like PPS-E, to a different coordination than the other simulations (see Fig. B.4 in appendix B). This could explain the low barrier in $\mathrm{P_iR}$-A simulation.

The $\mathrm{P_i}$ release pathway is shown in Fig. 3.10B) for the $\mathrm{P_iR}$-(A, D, E, F) configurations. The color dots representing the $\mathrm{P_i}$ trajectory during the US simulations bring to light the tube from active site to the inner cleft, which corresponds to the backdoor I route for $\mathrm{P_iR}$-(A, D, E) configurations. At the end of the tube, the $\mathrm{P_i}$ passes through its position in $\mathrm{P_iR1}$ crystallographic structure (4PJN) of (Llinas et al., 2015), which is represented in transparent tan surface in Fig. 3.10. For these structures, the $\mathrm{P_i}$ takes the backdoor I route through the THR197, ARG199, ARG205, GLY461 residues. The evolution of the critical salt bridge and the secondary salt bridge during the $P_i$ release are represented on Video 7[8]. A secondary salt bridge is observed for all these configurations, $P_i$R-(A, D, E) with an open Switch II. Then the mixed salt bridge turns into a partially formed critical salt bridge just before the release of $P_i$, which interact with ARG199 residue.

The interactions analysis (see Table 3.2, the colors charts Fig. B.12 and Fig. B.14 in appendix B and Video 8 [9]) provides similar results for $P_i$R-(A, D, E) US simulations as for PPS-(A, D, E, F) simulations, *i.e.* with a strong interaction of $P_i$ with THR197, ARG199, SER203 and ARG205 residues of Switch I just before $P_i$ explores the position of $P_i$R1.

Among these simulations, $P_i$R-D shows an higher free energy barrier on PMF for $P_i$R-D structure (see Fig. 3.9b). This can be explained by the fact that, contrary to the other $P_i$R-(A, E and F) and alike the PPS simulations, $P_i$ is in interaction with GLY459 at the beginning of US simulation ( and until $d_{P_i-Mg} \approx 8\,\text{Å}$ ). In this simulation the Switch II is less open than in the others $P_i$R simulations.

In $P_i$R-F the $P_i$ releases by the sidedoor (see Fig. 3.10B) without passing through the $P_i$R1 structure. In this case, which is the most hydrated (interaction with 3 $H_2O$) at the end of relaxation,the $P_i$ does not interact with THR197 and it goes towards GLY459 and PHE460 residues of the Switch II to exit. For this particular configuration, we found that the distance between ARG205 and GLU461 stays small along the US simulation, the critical salt-bridge is reformed, like in the PPS US simulations, (see Fig. B.14 in appendix B). Given that the $P_i$ orientation (1,2) after relaxation in $P_i$R-F is the same as $P_i$R-(D and E) structure, we can ask ourselves if the presence of the critical salt bridge favors the exit by the sidedoor.

## 3.3 Discussion

This study allowed us to compare the influence of the protonation state on the $P_i$ release from the PPS (PDB 2V26) and $P_i$R (PDB 4PJM) structures of myosin VI.

Starting from six different positions of the protonated oxygen of the $P_i$ in PPS and in $P_i$R, the relaxation leads to four main orientations in PPS and three in $P_i$R. The nature protonated or deprotonated of the oxygen coordinated to $Mg^{2+}$ (O1) stay along the dynamics. The obtained final structures have most of the time a protonated O2. The only relaxed structures without protonated O2 are PPS-A and $P_i$R-A. These seems to show that the protonated state of position O2 is more favorable than the deprotonated one. It seems that, except for the final structure PPS-B, when the final structure shows a protonated position O1 then the position O3 is deprotonated and similarly a deprotonated position O1 is with protonated O3. The $P_i$ is less prone to reorientation in the PPS structure compared to the $P_i$R structures, especially if the O1 is not protonated. This observation supports the conclusion that $P_i$ release from the active site is facilitated in $P_i$R because of an increase in $P_i$ mobility.

---

[8] Video 7 available online: https://youtu.be/2ZPqOfjERyU

[9] Video 8 available online: https://youtu.be/xQt_VEkDWhI

Our study does not address the question to know what is the *in vivo* $P_i$ orientation just before the release. This question remains an open question. However, (Kiani and Fischer, 2014) have found a $P_i$ protonation state with an orientation (2,3) similar to the ones of our simulations where $P_i$ does not release: PPS-C and $P_i$R-(B and C) (see Table 3.1). The study of Kiani and Fischer on catalytic mechanism of ATP hydrolysis was conducted by combined quantum-classical (QM/MM) simulations on Dictyostelium discoideum Myosin II structure. Therefore, if we assume that the hydrolysis reaction takes place in the same way in myosin VI and that the orientation (1,2) is the privileged one as our results suggest, a rotation or an exchange of protons may occur after the hydrolysis and before the departure of the $P_i$. For instance, we can imagine a proton transfer from oxygen (O3 or O4), which interacts with the $\beta$-Phosphate towards the O1 which coordinates the $Mg^{2+}$. Taking into account the protonation changes in the simulations could maybe lower the obtained high values of the free energy barrier. The jump of the PMF is particularly due to the detachment of $P_i$ from $Mg^{2+}$.

Simulations with much higher computational cost, such as QM/MM simulations (Sun et al., 2017; Zhou et al., 2018; Berraud-Pache et al., 2018) or reactive force fields like ReaxFF potential force field (Chenoweth et al., 2008; Senftle et al., 2016; Shin et al., 2021), could help us to investigate these questions.

We can also imagine other explanations why our simulation result in high value of the free energy barriers. Allosteric effect of actin or the hydration of $Mg^{2+}$ were not taken in our simulations. Indeed, (Mugnai and Thirumalai, 2021) observed in most of their MD simulations of myosin VI that the $Mg^{2+}$ coordinates with four water molecules before the $P_i$ release. However, this phenomenon was not observed during our US simulations. It can be assumed that the addition of two other water molecules close to Mg could lower the value of the free energy barriers.

Despite these reservations about free energy barriers, we are quite confident that certain facts seem to favor the $P_i$ release, such as the orientation (1,2) which is found for the PPS-(D, E and F) and $P_i$R-(D, E and F) structure. This could be due to the fact that protonated O1 interacts with $Mg^{2+}$ and its hydrogen with the $\beta$-Phosphate for all these structures. For PPS-(A, B and C) and $P_i$R-(A, B and C), we observe that the O1 oxygen interacts with $Mg^{2+}$ is unprotonated and that another oxygen of $P_i$ interacts with $\beta$-Phosphate. For all these structures, except $P_i$R-A, the $P_i$ release is much more difficult or impossible. Although the $P_i$R-A configuration exhibits the same interactions with $Mg^{2+}$ and $\beta$-Phosphate, the free energy barrier obtained in the US simulation of $P_i$R-A is of the same order of magnitude as the ones of configurations (PPS-(D, E and F) and $P_i$R-(D, E and F)), that start with a (1,2) orientation. This could be due to the change in $Mg^{2+}$ coordination observed during the relaxation (see Fig. B.4 in appendix B. A lower energy barrier is also observed for PPS-F that present the same change of $Mg^{2+}$ coordination.

From the point of view of $P_i$ release pathway, the $P_i$ exits seven time out of height through the backdoor I route and once through the sidedoor route (see Fig. 3.10). For $P_i$ releases pathways through the backdoor I, the structure explores the crystalline structure $P_iR1$ and $P_i$ interacts with THR197 of Switch I and SER153 of the P-Loop. The $P_i$ reaches the cleft as soon as it has finished crossing the $P_iR1$ structure, when the interactions with SER153 initially present in the active site stop. In the $P_iR$-F case, the $P_i$ does not pass through the $P_iR1$ structure. In this simulation, the $P_i$ does not interact with THR197 residue and it goes towards GLY459 and PHE460 residues of the Switch I to release through the sidedoor. It could be noticed that in the starting structure $P_i$ form H-bonds with three water molecules, see Table 3.1).

At first glance, the backdoor I route seems more favorable for the $P_i$ release from PPS and $P_iR$ structure as also observed by (Cecchini et al., 2010) for PPS and post-rigor structures of dicty myosin II. Conversely (Mugnai and Thirumalai, 2021) observed the release by four routes for the release of $P_i$ in myosin VI structure : frontdoor, sidedoor, backdoor I and backdoor II (see Fig. 3.2 for the definition of these pathways). The quantification of the relative probabilities of each $P_i$ release route would require several simulation for each protonation state that was not conducted in this study due to their computational costs.

## 3.4 Materials and Methods

### 3.4.1 Model construction

Simulation for the PPS state started from the crystal structure PDB code 2V26 (Gourinath et al., 2003) of myosin VI containing ADP, Vanadate and $Mg^{2+}$. Vanadium atom was replaced by Phosphorus atom. Simulation for the $P_iR$ state started from conformation A of the crystal structure PDB code 4PJM, which was proposed to be an intermediate structure following the PPS in the myosin cycle and bound to actin, allowing $P_i$ departure before the powerstroke (Llinas et al., 2015). This structure contains ADP, $P_i$ and $Mg^{2+}$ and is characterized by SER153 forming an hydrogen bond with an oxygen from $P_i$. Both structures were resolved without the presence of actin.

In order to model missing loops (residues 1-4, 175-179, 397-406, 565-566, 622-637), the program Modeller (Webb and Sali, 2017) was used to generate ten initial models, and the best model was taken based on the DOPE (Discrete Optimized Potential Energy) score (Shen and Sali, 2006). All water molecules present in the crystallographic structure were kept, except those making clash with reconstructed loops's residues. In particular, two water molecules are already coordinated to $Mg^{2+}$ in both crystallographic structures.

Concerning the protonation state of the residues (except $P_i$), the protonation states for each structure were calculated by the H++ server (available online) (Anandakrishnan et al., 2012) using neutral pH, a salinity of 0.15 M and a dielectric constant of 4 and 80 for the interior of the protein and the solvent respectively. For $P_i$, for each structures, six models were constructed by protonating two out of the four oxygen atoms of inorganic phosphate in the crystal structures of PPS and $P_iR$, respectively. The numbering of the oxygen atoms of the $P_i$ are given in Fig. 3.3. Oxygen position 1 is the oxygen closest to $Mg^{2+}$ and Oxygen position 2 is the one on the downright when looking at the direction $Mg^{2+}$-O1-P, i.e. the closest to SER203. The position 3 is the one up when looking at the phosphate facing with position 1 on the left of P and position 2 on the right of P, and the position 4 is the one down or back. The description of orientation of $P_i$ giving the oxygen atoms involved in H-bonds with the residues are given in the lines PPS-Initial and $P_i$R-Initial of Tab. 3.1 for the two starting structures PPS and $P_iR$ respectively. The initial protonation state of each model is drawn in the inserts of the Fig. 3.4 and Fig. 3.7 and described in the first column of Tab. 3.1.

The models were solvated using the tleap program from the AMBER suite (Case et al., 2005) in a cuboid box of TIP3P water with a minimum distance of 12 Å between the protein and the edges of the box. $Na^+$ and $Cl^-$ were added to neutralize the system and to obtain a salt concentration of 0.15 M.

### 3.4.2 Force field

Protein was parameterized using the AMBER ff14SB force field (Maier et al., 2015), parameters for ADP and $Mg^{2+}$ were taken from the Bryce database[10]. These parameters correspond to the parameters ANV in (Mugnai and Thirumalai, 2021). Force field parameters and charges for the phosphate were respectively taken from GAFF2 (the second version of the General Amber Force Field (Wang et al., 2004)) and (Kashefolgheta and Vila Verde, 2017).

### 3.4.3 Molecular Dynamics Simulation

#### 3.4.3.1 Relaxation dynamics

In the whole paper, relaxation stands for the unconstrained (free) dynamics run for 200 ns after construction of the models.

Minimization was performed in a sequence of four minimizations. In the first minimization, only hydrogen atoms are free to move while heavier atoms are kept under harmonic restraints on cartesian positions. This step allows to remove clashes involving hydrogen atoms. In second minimization water molecules and reconstructed loops are free to to adjust, while heavy atoms from the rest of the protein remain under positional restraints. The third minimization allows all atoms except heavy atoms from backbone to move. Finally, the whole system was submitted to a fourth minimization without any restraint. Each minimization has been carried out starting with 1000 steps of Steepest Descent followed by 4000 steps of Conjugate Gradient and using a force constant of $5\,\mathrm{kcal\,mol^{-1}\,\mathring{A}^{-2}}$ for the restraints.

After the minimization, the system was heated up linearly from 0 K to 300 K using a Langevin thermostat with a collision frequency of 1.0 $ps^{-1}$ during 1 ns in NVT. Short range

---

[10] Bryce database available online: http://amber.manchester.ac.uk/

interactions were cut off at 12 Å and a time step of 2 fs was used. The SHAKE algorithm (Ryckaert et al., 1977) was used to constrain all bonds involving hydrogen atoms. Following heating, the system was equilibrated during 2 ns in NPT, pressure being controlled at 1 bar using a Berendsen barostat (Berendsen et al., 1984) and a relaxation time of 1 ps. Dynamics was finally performed using the same conditions (pressure and temperature) for 200 ns. Coordinates were saved every 10 ps. Dynamics were run using the pmemd program and analyzed with the Cpptraj program both from the AMBER suite (Case et al., 2005). Trajectories were visualised with VMD (Humphrey et al., 1996). The final structures were used as initial structures for the Umbrella Sampling simulations.

### 3.4.4 Umbrella Sampling protocol

In order to compute the leaving of the $\mathrm{P_i}$ from the cavity we employed Umbrella Sampling (US) simulations with harmonic restraints. The distance between the center of mass of the heavy atoms from the phosphate and the magnesium $d_{P_{\mathrm{i}}-\mathrm{Mg}^{2+}}$ was used as collective variable (CV). An harmonic potential was used, with a force constant of 20 kcal/mol/Å$^2$. Distance restraints were taken every 0.3 Å in the range spanning from 3.5 Å to 20.0 Å. Starting from the last frame of the previous step/windows, a minimization with new restraint of 5000 steps, decomposed in 1000 steps of Steepest Descend followed by 4000 steps of Conjugate Gradient, was run followed by 300 ps of equilibration. Production runs of 3 ns were then simulated for each of the 56 windows.

The data were analyzed using 40000 structures saved every 0.1 ps. The potential of mean force (PMF) profiles were constructed using the weighted histogram analysis method (WHAM) (Kumar et al., 1992) taking the CV values distribution resulting from the US simulations.

### 3.4.5 Interactions analysis

In order to follow the interactions, and especially the H-bond formations and breaks along the dynamics, colour charts are produced. First, four Umbrella Sampling Simulations (2 PPS and 2 $\mathrm{P_i}$R) are used to determine which interactions are relevant to follow. For each Umbrella sampling simulation, 6 snapshots from each 3 ns equilibration of the 56 Umbrella Sampling windows (simulation under one values of the constraint) were taken (one snapshot every 0.5 ns) leading to a total of 336 snapshots.

The Cpptraj program of AMBER suite of programs was used to identify (i,j) residues couples having hetero-atoms Xi and Yj belonging respectively to residues i and j, with atom Xi protonated. The couple (i,j) is selected if i or j is $\mathrm{P_i}$ or a residue of the Nucleotide Binding Loops and if at least one of the Xi-H Yj pairs forms an H-bond in at least 4 snapshots of the 4*336 snapshots . H-bond is considered to be formed when the distance X · · Y is 3 Å or less and the angle X-H · · Y is 180° more or less 45° (as it is as default in Cpptraj). This allows to select 14 couples of residues (i,j), where i and j belongs to the Nucleotide Binding Loops and

14 couples of residues (i,j), where i or j is the $P_\mathrm{i}$. The couples (i,j) are not ordered, i.e. (i,j) and (j,i) correspond to the same residue couple where at least one H-bond was detected during one Umbrella sampling. A similar analysis is done for distances between residues and $Mg^{2+}$ with the only criteria of distance (without considering any protonated or deprotonated heteroatom) and reported in the color charts in the Suppl. Info..

For the interaction analysis of the 200 ns relaxation simulations, 100 snapshots were taken every 1 000 000 simulated steps, which corresponds to 1 snapshot every 2 ns simulated time. For the interaction analysis of US simulations colour chart analysis, 112 snapshots were taken every 1 500 000 simulated time steps, which corresponds to 2 snapshots per US windows/values of the constraint.

## 3.5 Glossary

In the present chapter, we use the following abbreviations :

- ATP: Adenosine triphosphate
- $P_\mathrm{i}$: Inorganic phosphate (in this study : $H_2PO_4{}^-$)
- US: Umbrella Sampling
- NBL: Nucleotide Binding Loops
- PPS: Pre-Powerstroke state
- $P_\mathrm{i}$R: $P_\mathrm{i}$-Release state
- $d_{P_\mathrm{i}-Mg}$ : Distance $P_\mathrm{i}$-$Mg^{2+}$, between the heteroatoms of $P_\mathrm{i}$ and $Mg^{2+}$, used as reaction coordinate in US simulations
- $d_{Mg-O}$ : Distance $Mg^{2+}$-O, between the $Mg^{2+}$ and one oxygen from a water molecule, used as reaction coordinate in US simulations bringing water molecules

# Part II

# Characterization of PPS, PiR & Strong-ADP

# Chapter 4

# Study of the state of the Cleft and Switch II in unconstrained dynamics of PPS, $P_iR$ & Strong-ADP

The objective of this chapter and the following ones is to characterize the PPS, $P_iR$ and Strong-ADP states using both their crystallographic structures and relaxation simulations. The crystallographic data provide static information, and relaxation simulations provide information on the dynamical behaviour of these structures. In this chapter we study specifically the states of the cleft and the switch II loop from the active site with some selected distances.

**Contents**

## 4.1 Introduction

The PPS and $P_iR$ states of myosin correspond to the beginning of the powerstroke. PPS corresponds to a state where myosin starts interacting with actin without being fully bound. These interactions are electrostatic and involve the flexible actin binding loops located on the actin binding surface of myosin (Robert-Paganin et al., 2019). These interactions are weaker than stereospecific interactions developed later in the cycle, for example in the Strong-ADP state.

The $P_iR$ state is hypothesized to represent one of the first step of actin bound myosin, with the noticeable characteristics that the open inner cleft allows $P_i$ departure from the active site (Llinas et al., 2015).

In Strong-ADP the myosin cleft is completely closed, allowing myosin to form strong interactions with actin. The myosin also exhibits a different Converter orientation, compared to PPS and $P_iR$. This change of orientation is related to the powerstroke, which is almost completely performed in Strong-ADP. Moreover, the Converter is known to have large fluctuations in its position (Blanc et al., 2018).

In this chapter, we focus the study on the state of the cleft and the Switch II loop that is part of the inner cleft. A coarser and/or more global description of the whole protein are presented in the next chapters.

## 4.2 Choice of structures and simulation protocol

The chosen structures for PPS, $P_iR$ and Strong-ADP are from PDB 2V26 (Ménétrey et al., 2007), 4PJM (Llinas et al., 2015) and 6BNQ (Gurel et al., 2017). Among these structures, 2V26 and 4PJM are X-ray crystal structures of myosin without actin. 6BNQ contains an actomyosin complex solved with cryo-EM with a fragment of actin filament with 8 actin monomers and 6 bound myosin heads. The chosen Strong-ADP structure is one of the myosin heads (chain I) of 6BNQ.

The myosin structures extracted from the three PDB files, were completed with their missing loops using MODELLER (Webb and Sali, 2016). The Strong-ADP state is different from the Rigor state notably because of the presence of ADP without $P_i$ in Strong-ADP. The ADP molecule is released during the transition toward the Rigor state. However, the ADP is not present in the 6BNQ PDB file. This is due to limited resolution resulting in non explicit placement of ADP. We thus had to add it in order to get a correct behaviour of the conformation. This was performed by superposing residues 150 to 170 (P-Loop and the following helix) of $P_iR$ 4PJM on those of Strong-ADP 6BNQ chain I structure, and using the coordinates of ADP in the $P_iR$ structure to position it in the Strong-ADP structure.

The most probable protonation states for each structure were calculated by the H++ server (Anandakrishnan et al., 2012) (available online) using neutral pH, a salinity of 0.15 M and a dielectric constant of 4 and 80 for the interior of the protein and the solvent respectively. $Na^+$ and $Cl^-$ ions were added to neutralize the system and obtain a salt concentration of 0.15 M. Protein was parameterized using the AMBER ff14SB force field (Maier et al., 2015). The parameters for ADP and $Mg^{2+}$ were taken from the Bryce database. The parameters for $Mg^{2+}$ correspond to the parameters ANV in ref (Mugnai and Thirumalai, 2021). Force field parameters, including charges, for the phosphate were respectively taken from (Kashefolgheta and Vila Verde, 2017) and GAFF2 (the second version of the General Amber Force Field (Wang et al., 2004)).

The obtained models were then hydrated with TIP3P water molecules in a rectangular box with at least 12 Å between the protein and the edge of the box using the LEaP program of AMBER (Case et al., 2005). The systems composed of the protein and the box of water are represented on Fig. 4.1. On this figure, the PPS structure is represented in blue, the $P_iR$ structure is in red and the Strong-ADP is in gray.

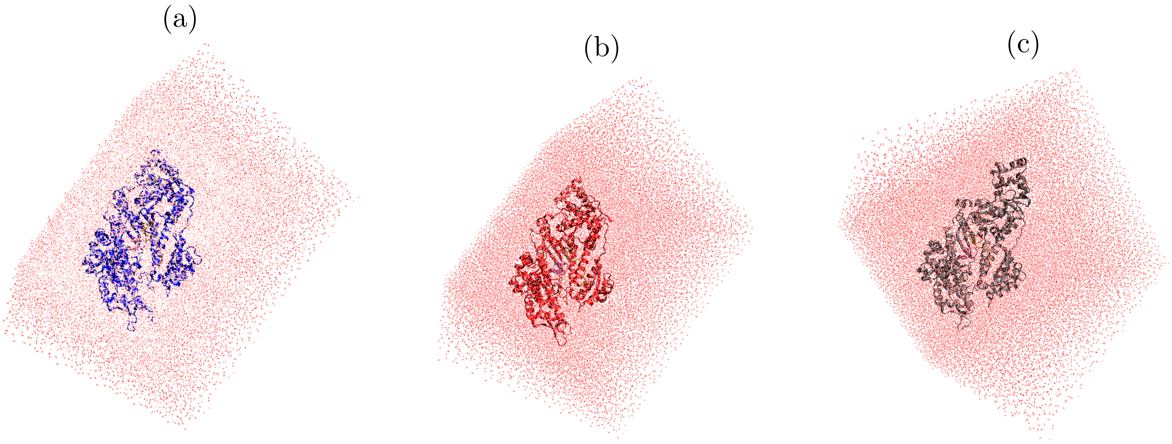


Figure 4.1: Simulations box before relaxations of (a) PPS in blue, (b) $P_iR$ in red and (c) Strong-ADP in gray.

The system was then minimised in three steps all starting with 1000 steps of Steepest Descent followed by 4000 steps of Conjugate Gradient and using a force constant of 5 kcal.mol$^{-1}$.Å$^{-2}$. First, only the hydrogen atoms were allowed to move, while the other atoms were restrained. Second, the water molecules were allowed to move, together with the side chains of the residues. Third, the whole system was minimized without restraint. Finally, the system was heated linearly from 0K up to 300 K in a 1 ns NVT simulation with a Langevin thermostat using a friction coefficient of 1 ps$^{-1}$ , a time step of 2 fs (this time step is kept for all the simulations) with the SHAKE algorithm to restrain hydrogen bonds. After that, the system was equilibrated in NPT with a Berendsen barostat at 1 bar during 2 ns with a time constant set at 1 ps. The system was then simulated in NPT for 300 ns, keeping the same thermostat and barostat. This last simulation is called unconstrained simulation or relaxation in what follows.

## 4.3 Study of the state of the cleft

In this section we present the criterion used to define the state of the cleft. We chose distances between $C_\alpha$ from residues assumed to define the degree of cleft closure. These distances are represented on Fig. 4.2.

### 4.3.1 State of the cleft in the crystallographic structures

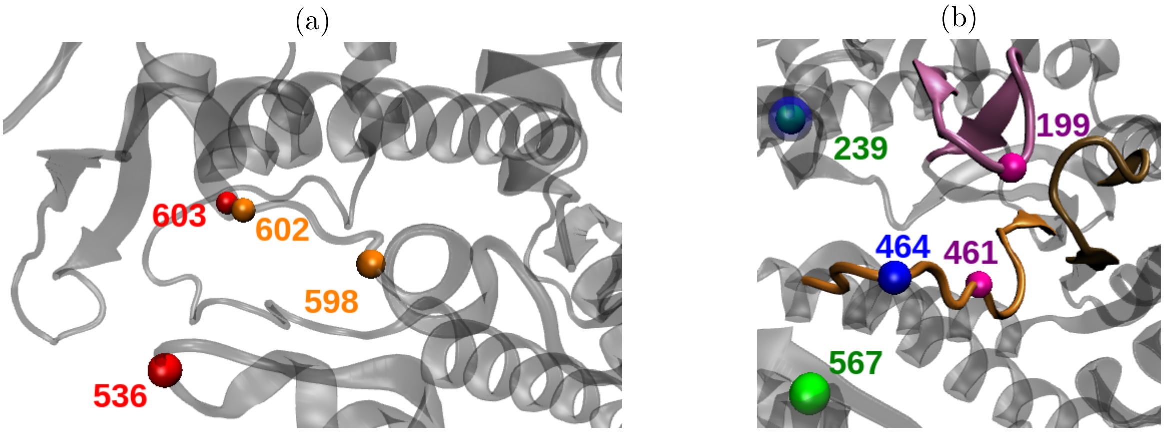


Figure 4.2: Representation of the $C_\alpha$ from the residues, whose distances are chosen to reflect the state of the cleft. (a) Residues describing the Outer Cleft. Red balls correspond to distance PRO536-MET603. Orange balls correspond to distance ASN598-HIS602. (b) Residues describing the Inner Cleft. Blue balls correspond to distance GLY239-GLU464. Green balls correspond to distance GLY239-567. Pink balls correspond to GLU461-ARG199.

| Distance | PPS | $\mathrm{P_iR}$ | Strong-ADP | Diff. $\mathrm{P_iR}$-PPS | Diff. $\mathrm{P_iR}$-SA |
|---|---|---|---|---|---|
| Outer Cleft | | | | | |
| PRO536-MET603* | 20.6 | 16.0 | 20.3 | 4.6 | 4.3 |
| ASN598-HIS602* | 11.1 | 9.9 | 11.8 | 1.2 | 1.9 |
| Inner Cleft | | | | | |
| GLY239-GLU464 | 11.2 | 13.7 | 11.6 | 2.5 | 2.1 |
| GLY239-ILE567 | 17.2 | 19.6 | 17.1 | 2.4 | 2.5 |
| GLU461-ARG199 | 7.8 | 11.3 | 8.1 | 3.5 | 3.2 |
| GLU461-THR197 | 10.0 | 13.4 | 10.3 | 3.4 | 3.1 |
| GLU461-ARG205 | 11.6 | 13.0 | 12.0 | 1.7 | 1.0 |

Table 4.1: Values in Å of the chosen distances describing the cleft in crystal states of PPS, $\mathrm{P_iR}$ and Strong-ADP and difference between $\mathrm{P_iR}$ and PPS or between $\mathrm{P_iR}$ and Strong-ADP. The distances are separated between the distances describing the outer cleft, near the actin binding surface and the one describing the inner cleft.

The values of the chosen distances in the selected crystallographic structures of PPS, $\mathrm{P_iR}$ and Strong-ADP are presented in Tab. 4.1. Two groups can be identified. The first group corresponds to the description of the outer cleft, which is the part close to the actin binding surface. These associated distances are ASN598-HIS602 and PRO536-MET603. These distances are shorter in the $\mathrm{P_iR}$ crystal structure compared to PPS and Strong-ADP. The second group corresponds to the inner cleft closer to the active site. It regroups all the remaining distances shown in Tab. 4.1. These distances are larger in the $\mathrm{P_iR}$ crystal structure compared to PPS and Strong-ADP. The differences observed between the two groups reflect the opening of the inner cleft, and the closure of the outer cleft.

From the analysis of the different structures the PRO536-MET603 was selected as the best descriptor of the closure state of the Outer Cleft.

The distance GLU461-ARG199 (or GLU461-THR197) could be used to discriminate $P_iR$ from PPS or Strong-ADP as the differences for these distances between $P_iR$ and PPS or SA (see the 2 last columns of table 4.1), are more than 3 Å.

### 4.3.2 State of the cleft along relaxation simulations

The evolution of the selected distances along unconstrained MD are represented on Fig. 4.3. The average observed value for the distance 239-464 is increasing by ∼3 Å from PPS (10 Å) to Strong-ADP (13 Å) with an intermediate value for $P_iR$ (12 Å), see the blue traces in Fig.4.3. This distance could therefore be a good descriptor for the evolution of the system from PPS to Strong-ADP. The value of the distances in the crystal structures are represented by the dotted lines in Fig. 4.3. The distance 239-464 in the crystal structure is larger for $P_iR$ than for PPS and Strong-ADP (see also table 4.1). This is not in line with the average values during the simulation since we observe that it is Strong-ADP that evolves around the value seen in the $P_iR$ crystal structure.

The distance 598-602 (Fig.4.3 in orange) seems quite good for discriminating PPS and Strong-ADP, with a difference of 2 Å between the median of the values observed during simulations starting from PPS and Strong-ADP. Comparing the crystal structures, this distance is shorter for PPS than for Strong-ADP, as in the dynamics. The difference between PPS and $P_iR$ is less obvious looking at the average of the dynamics compared to the crystal structures.

The crystal structure of PPS corresponds to a value of the distance 598-602 on the top of the upper quartile of this distance during the dynamics. This can be due to crystal packing that are relaxed during the simulation. Analysis of other PPS crystal structures could be done to better understand the origin of this observation.

Another distance that could be used is 536-603 (in red in Fig.4.3) with a difference of 3 Å between the median values in PPS and Strong-ADP. The distance 536-603 is lower in $P_iR$ than in both PPS and Strong-ADP in average along the dynamics like distance 598-602. It could then help define $P_iR$. However distributions overlap between PPS and $P_iR$ is quite large. Moreover, as for the previous distance, the dynamics from the PPS structure exhibits values similar to the one from $P_iR$, which evolve around the value in $P_iR$ crystal structure, while the distance in the PPS crystal structure is similar to the one in Strong-ADP.

In contrast to all preceding distances, the distance 239-567 (in green in Fig.4.3) is not discriminating since all boxplots are overlapping.

Finally, the two last distances defining the inner cleft evolve as expected in the PPS simulation, and in the first part of the $P_iR$ simulation, however, by the end of the simulation of $P_iR$ they shorten significantly, which is correlated with a closure of Switch II.

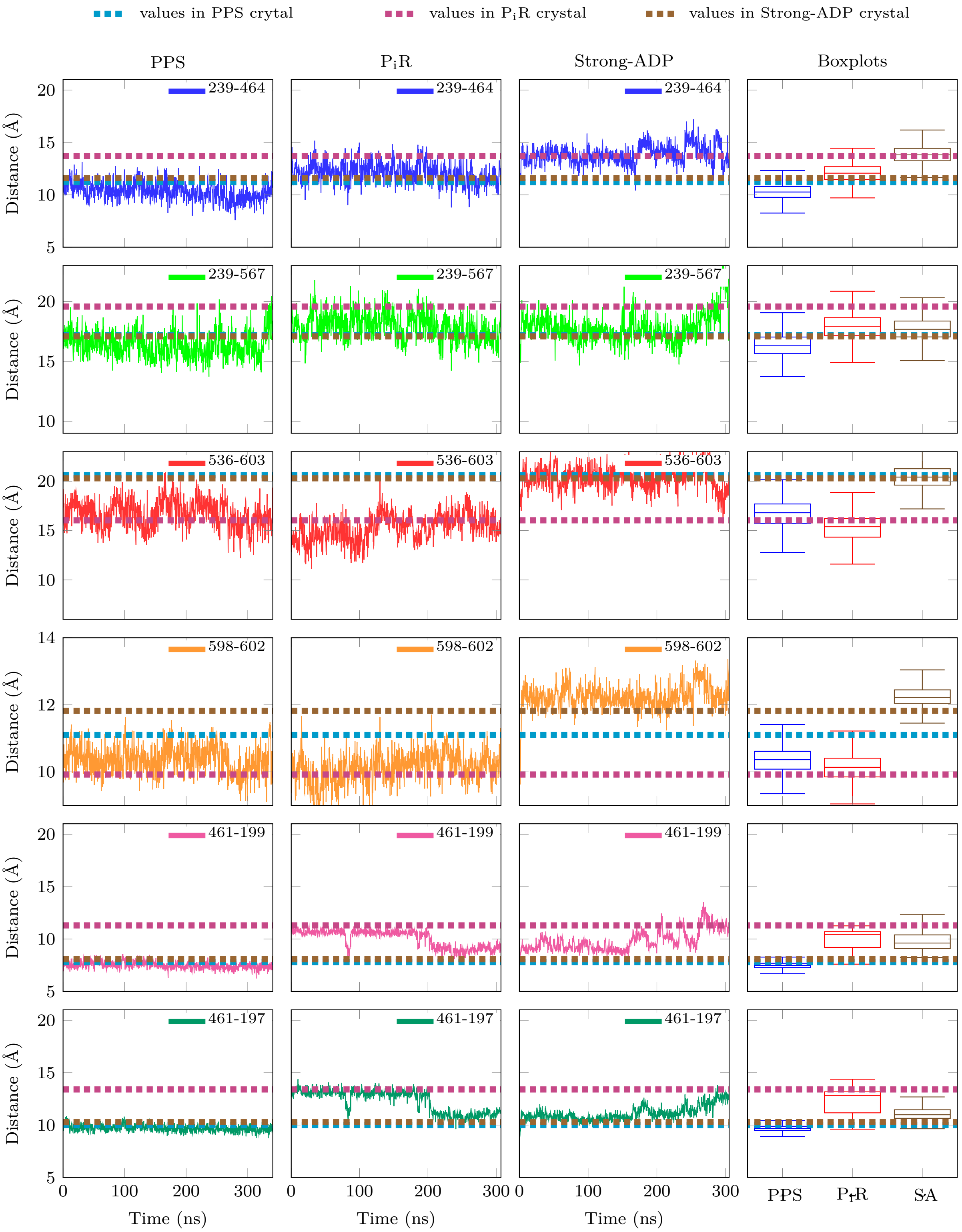


Figure 4.3: Distances between $C_\alpha$ of residues representing the cleft closure along unconstrained dynamics of PPS, P$_i$R and Strong-ADP. (blue) GLY239 and GLU464, (light green) GLY239 and ILE567, (red) PRO536 and MET603, (orange) ASN598 and HIS602, (pink) GLU461-ARG199, (dark green) GLU461-THR197. Boxplots are represented on the last column, with the median, the upper and lower quartile and the upper and lower whisker. (lower/upper whisker are the smallest/largest data values which are larger/smaller than lower/upper quartile -1.5 x IQR where IQR is the "interquartile range", i.e. the difference between upper quartile and lower quartile. Dotted lines correspond to the values in the crystal structures for comparison. (dotted blue) PPS, (dotted red) P$_i$R and (dotted orange) Strong-ADP.

Both the study of the crystal structures and the evolution during unconstrained dynamics lead to take the distances 536-603 to distinguish a characteristics of $P_iR$ compared to Strong-ADP or PPS, specifically as it is outside of the range between the median values in PPS and in Strong-ADP. This distance defines the outer cleft. In order to add a criterion defining the Inner Cleft we could take the distance 461-197 that is larger of approximately 2 Å during most of the simulation for $P_iR$ compared to PPS or Strong-ADP.

## 4.4 State of Switch II

The Switch II loop is one of the three nucleotide binding loops (NBL) together with Switch I and P-Loop. The interactions these NBL form with the ligands and with surrounding residues lead to allosteric communications between distal parts of the myosin. One of the main characteristic of the $P_iR$ state is an open Switch II. We thus describe its state quantitatively with a set of distances represented on Fig. 4.4.

In particular, the distance 205@CZ-461@CD defines the formation of the critical salt bridge necessary for efficient ATP hydrolysis (Onishi et al., 1998). The formation of the critical salt bridge might be helped by the formation of a H-bond that is defined by the distance 460@O-153@N. Another distance of interest is 199@CZ-461@CD, which defines the formation of another salt bridge. It can be seen as an antagonist to the first (205@CZ-461@CD) salt bridge, as both ARG199 and ARG205 can interact with GLU461, but cannot be formed exactly simultaneously, this salt bridge is then defined as the *secondary salt bridge* in (Blanc, 2018). This secondary salt bridge is often seen as forming during simulations while being rarely observed in crystals. Monitoring these interactions might help to explain why this opening can be important and how it can be opened and maintained as such by the first interactions with actin. However, the results should be taken with caution because of the absence of actin in our simulations.

### 4.4.1 State of Switch II in crystal structures

| Distance | PPS | $P_iR$ | Strong-ADP | concerns the ... |
|---|---|---|---|---|
| 205@CZ-461@CD | 3.9 | 7.4 | 5.6 | critical salt bridge |
| 460@O-153@N | 2.8 | 6.4 | 6.9 | H-Bonds favoring critical salt bridge |
| 461@O-474@ND2 | 2.9 | 2.9 | 3.4 | ? |
| 199@CZ-461@CD | 8.7 | 4.4 | 5.4 | secondary salt bridge |

Table 4.2: Table of the value in Å of the chosen distances from Switch II loop and P-Loop, Switch I and Relay helix in crystal states of PPS, $P_iR$ and Strong-ADP.

Tab. 4.2 shows the values of selected distances defining the state of the nucleotide binding site and, in particular, of Switch II in the crystal structures of PPS, $P_iR$ and Strong-ADP. The criterion used to determine whether a salt bridge or H-bond is formed or not is a distance of around 4 Åand 3 Å, respectively.

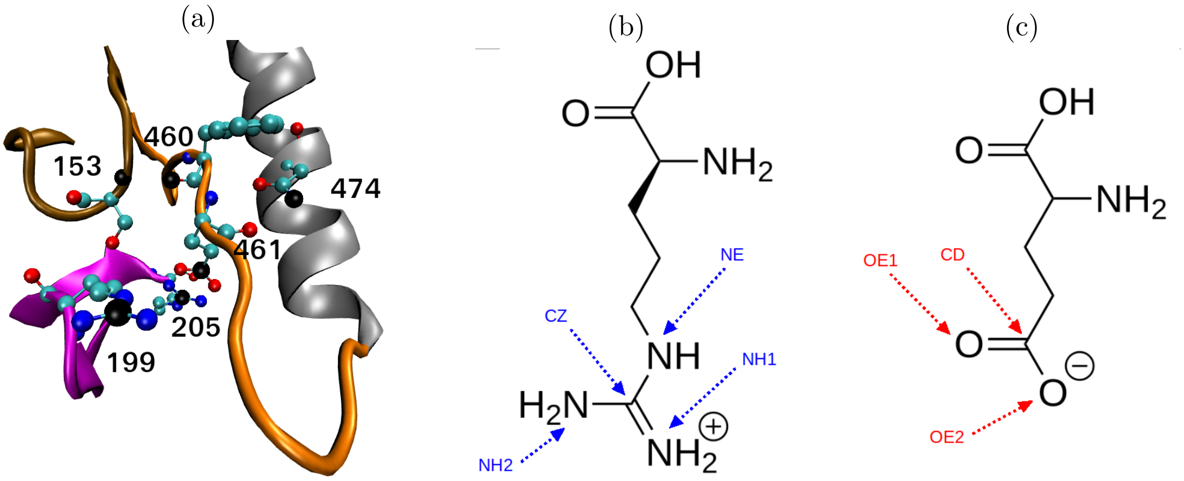


Figure 4.4: Representation of the distances chosen to reflect the state of Switch II. Switch II in orange, Switch I in magenta, P-Loop in ochre and part of Relay helix in gray. Distances are actually between the black balls from the represented residues on (a). The actual distances are SER153@N-PHE460@O, ARG199@CZ-GLU461@CD and ARG205@CZ-GLU461@CD. The corresponding atoms of the residues are represented on (b) for the arginine and (c) for the glutamate. For SER153 and PHE460, the residues are not represented as the atoms implicated in the distance simply are their backbone nitrogen and oxygen respectively.

We can observe that the distance between 205@CZ and 461@CD is small enough to allow the critical salt bridge formation in the crystal structures of PPS ($\sim$ 4 Å), while not formed in $\mathrm{P_i}$R (205@CZ-461@CD equal $\sim$ 7 Å), which instead has a formed secondary salt bridge (199@CZ-461@CD equal $\sim$ 4 Å). The H-Bond 460@O-153@N is formed only in PPS with a distance of 2.83 Å.

### 4.4.2 State of Switch II during relaxations

The evolution of characteristic distances of switch II during the relaxation simulations is shown in Fig. 4.5. All distances are stable along the first part of the simulation, up to 200 ns. For PPS and $\mathrm{P_i}$R the distances fluctuate in the first part of the simulation near the values measured in the crystal structure. However, for Strong-ADP the distances evolve around the values of the PPS crystal and not the ones of the Strong-ADP crystal. We recall that for the dynamics, ADP was artificially put in the Strong-ADP crystal structure which was initially solved without the coordinate of the nucleotide.

In the second part of the simulations (after 200 ns of the simulation of $\mathrm{P_i}$R) the critical salt bridge forms, while the secondary salt bridge breaks and the distance between residues 460 and 153 reduces. In addition, in Strong-ADP, the critical salt bridge and the H-bond between residue 460 and 153 initially formed break in the second part of the relaxation simulation.

This observation could be explained by the absence of actin, which would stabilize the position of Switch II in an open configuration that could thus help $\mathrm{P_i}$ departure. The claim that actin could impact the stability of Switch II is relevant by the fact that only the PPS

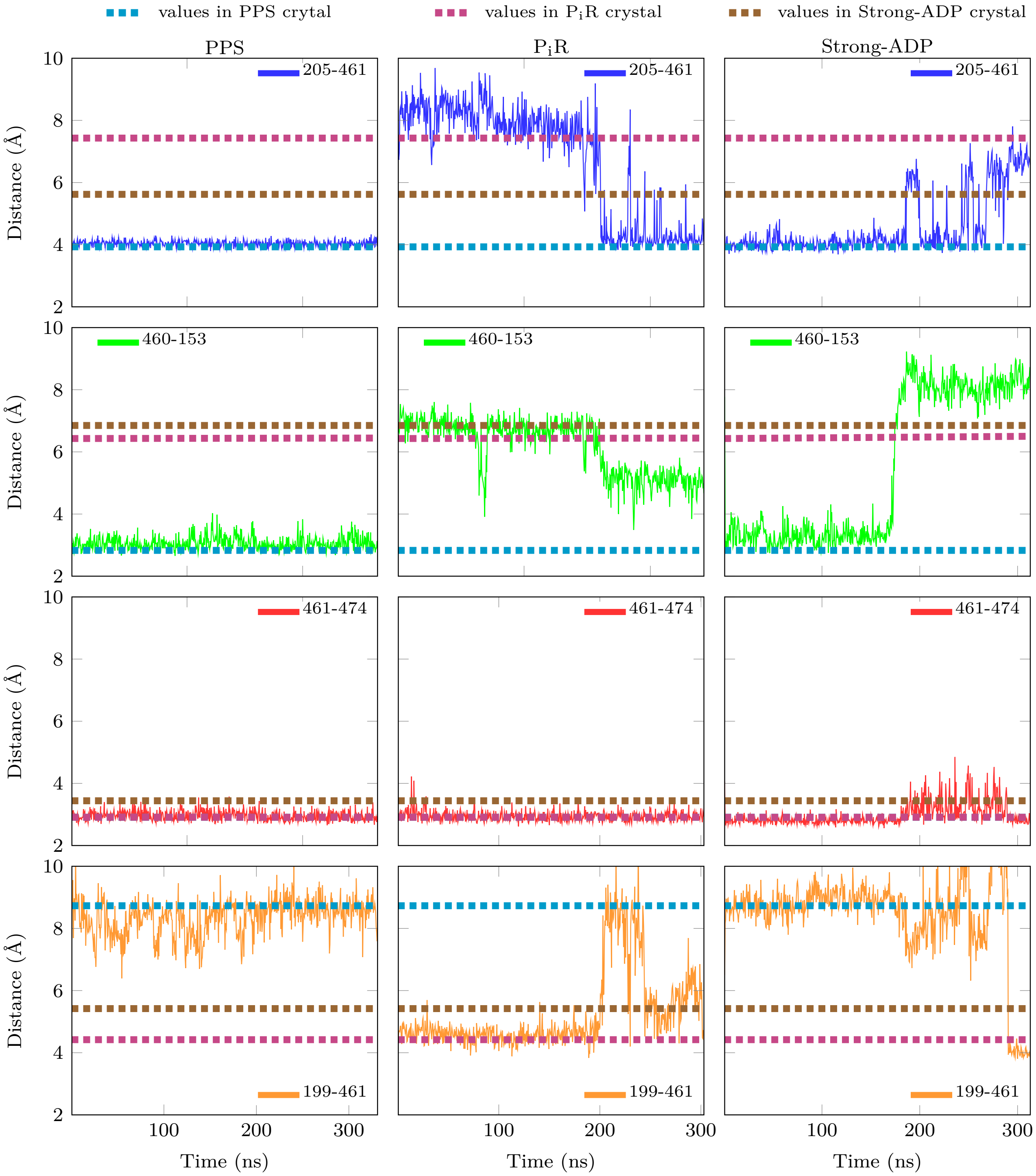


Figure 4.5: Distances between $C_\alpha$ of residues representing the state of Switch II along unconstrained dynamics of PPS, $P_iR$ and Strong-ADP. (blue) 205 and 461, (light green) 460 and 153, (red) 461 and 174, (orange) 199 and 461. Dotted lines correspond to the values in the crystal structures for comparison. (dotted blue) PPS, (dotted red) $P_iR$ and (dotted orange) Strong-ADP.

crystal structure is rightfully without actin, while $P_iR$ is assumed to be an intermediate bound to actin and Strong-ADP is with actin.

## 4.5 Conclusion

In this chapter we studied distances describing the state of the cleft and Switch II. We performed a comparison between their values in the crystal structures of PPS, $P_iR$ and Strong-ADP and their evolution during MD simulation of relaxation from these structures. The values of some distances evolved far from the one in the crystal structures during the MD simulations. It is particularly the case of the distance ARG205-GLU461 (and ARG199-GLU461) between Switch I and Switch II in the simulations of $P_iR$ and Strong-ADP. Their instability might be caused by the absence of actin in our model, even though it could stabilize these states.

# Chapter 5

# Study of the dynamics of PPS, $P_iR$ and Strong-ADP with helix

In this chapter we study the same simulations as in chapter 4. While the previous chapter was about the study of specific parts of the protein (cleft and switch II), in this chapter we aim at characterizing the whole protein. This developed methodology is based on the relative position of helices, using the distance between their centers of mass, the 3-points and 4 points ("dihedral") angles between the $C_\alpha$ from the first and last residue of each helix.

### Contents

## 5.1 Presentation of the helix methodology

Helices of more than 10 amino acids are assumed rigid so that they would not be subjected to large deformation. The choice of the helices and the corresponding amino acids was based on their appearance as helix in the VMD cartoon representation (thus following the associated convention for the definition of helices). The chosen helices are represented with different colors on the Fig. 5.1, and their residue numbers are shown in the table 5.1.

Their relative positions is characterized by three parameters chosen for their simplicity of computation and the fact that they can readily be used as a collective variable for biased simulations. The parameters were computed directly with the cpptraj program of the AMBER suite (Case et al., 2005). The first parameter is the distance $L^{i-j}$ between the center of mass of

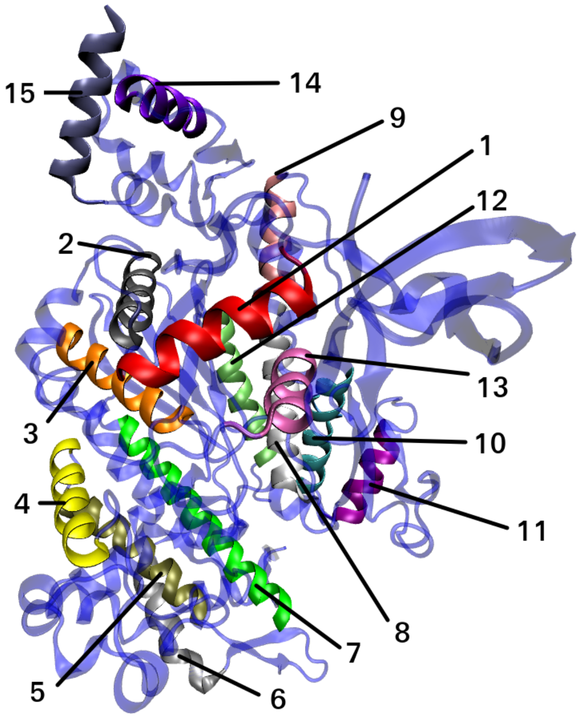


Figure 5.1: Representation of the myosin with the 15 helices chosen to describe the protein conformation highlighted.

| Helix number | Residue range | Number of residues | Domain | color |
|---|---|---|---|---|
| 1 | 69-84 | 16 | Nter | red |
| 2 | 126-142 | 17 | Nter | gray |
| 3 | 157-173 | 17 | Nter | orange |
| 4 | 313-328 | 16 | U50 | yellow |
| 5 | 331-349 | 19 | U50 | tan |
| 6 | 366-379 | 14 | U50 | silver |
| 7 | 412-441 | 30 | U50 | green |
| 8 | 467-486 | 20 | L50 | white |
| 9 | 490-500 | 11 | L50 | pink |
| 10 | 515-534 | 20 | L50 | cyan |
| 11 | 539-551 | 13 | L50 | purple |
| 12 | 643-660 | 18 | L50 | lime |
| 13 | 681-691 | 11 | Nter | mauve |
| 14 | 730-743 | 14 | Conv | ochre |
| 15 | 773-789 | 17 | Insert2 | iceblue |

Table 5.1: Description of the 15 helices. The range of the residue numbers, the number of residues contained in the helices as well as the domain the helices belongs to and the corresponding color in Fig. 5.1 are given.

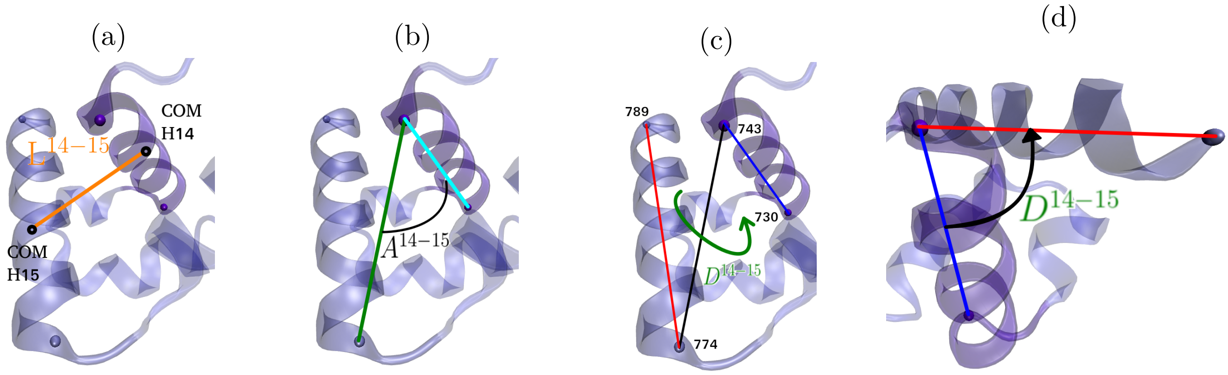


Figure 5.2: Representation of the three defined parameters used to represent the relative positionning of two helices. (a) Distance between the center of mass (COM) of both helix. (b) Angle defined by the $C_\alpha$ from the first and last residues of the first helix (helix 14) and the $C_\alpha$ from the first residue of the second helix (helix 15). (c) Representation of the dihedral angle that would correspond to a (there fictionnal) torsion between two helices. The three segments allowing the definition of the dihedral angle are explicitly shown. The three segments connect the $C_\alpha$ from the first and last residue of the first helix and the $C_\alpha$ from the first and last residue of the second helix. (d) Representation of the dihedral angle by turning the representation in (c) so that we see it in the direction of the black central segment, explicitly showing the dihedral angle. The dihedral angle $D^{i-j}$ is defined between the $C_\alpha$ of the residues at the ends of the helices in the order i followed by j and in the order of the residue numbers inside helix. This results in the dihedral angle between the $C_\alpha$ in this order : 730-743-774-789. If the dihedral 15-14 was computed, it would be different as the order would then be : 774-789-730-743, which is related but not trivially to the dihedral 14-15.

two helices $i$ and $j$. The coordinate $\underline{X}_i$ of the center of mass of helix $i$ is computed by averaging the positions $\underline{x}_{i,k}$ of the $N_i$ the backbone atoms of that helix (C, C$_\alpha$, N and O):

$$\underline{X}_i = \frac{1}{N_i} \sum_{k=1}^{N_i} \underline{x}_{i,k} \tag{5.1}$$

The second parameter is the angle $A^{i-j}$ formed by the C$_\alpha$ from the first and last residues of the first helix (helix i) and the C$_\alpha$ from the first residue of the second helix (helix j). This angle is illustrated on Fig. 5.2(b). An alternative angle $A'^{i-j}$ is also defined using the same two C$_\alpha$ of the first and last residues of the first helix and the last C$_\alpha$ of the second helix.

The third parameter is the dihedral angle $D^{i-j}$ between the axes of the helices corresponding to the segment between the C$_\alpha$ from the first and last residue of each helices in the increasing order of the residue numbers, as represented on Fig. 5.2.c.

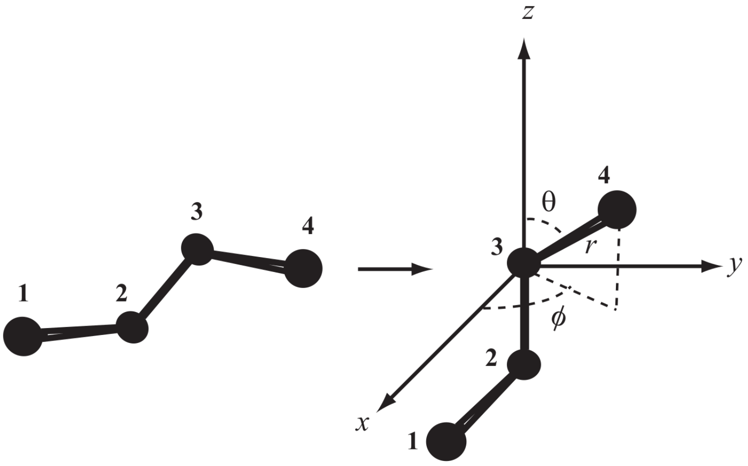


Figure 5.3: Visualization of 4 bounded atoms denoted by 1, 2, 3 and 4, and oriented such that the 1-2 axis lies along the $x$ axis of a coordinate system in which the atom 3 is the center. This orientation allows the definition of the dihedral angle $\phi$ based on the coordinate of the atom 4 in the spherical polar coordinate. Taken from (Tuckerman, 2010).

The dihedral angles can be computed from the positions of the four atoms as described in (Tuckerman, 2010).

Let four atoms be denoted 1, 2, 3 and 4 (in this methodology, atom 1 and 2 are the first and last C$_\alpha$ from the first helix while atoms 3 and 4 are the first and last C$_\alpha$ from the second helix, in the order of residue numberings), whose coordinates are denoted $\mathbf{r}_1$, $\mathbf{r}_2$, $\mathbf{r}_3$, $\mathbf{r}_4$.

To define the dihedral angle $\phi$, we write the coordinates of the vector, $\mathbf{r}_{3\to4} = \mathbf{r}_4 - \mathbf{r}_3$, connecting atom 3 to atom 4 into the base $\mathcal{B} = (\mathbf{x}, \mathbf{y}, \mathbf{z})$, where the unit vector $\mathbf{z}$ is aligned to the vector $\mathbf{r}_{2\to3} = \mathbf{r}_3 - \mathbf{r}_2$ and the vector $\mathbf{x}$ orients in the direction of the vector $\mathbf{r}_{21} = \mathbf{r}_1 - \mathbf{r}_2$ (see Fig. 5.3). With this convention the unit vectors of base $\mathcal{B}$ are given by

$$\mathbf{x} = \frac{\mathbf{r}_{2\to1} - \mathbf{r}_{2\to1} \cdot \mathbf{z}}{|\mathbf{r}_{2\to1} - \mathbf{r}_{2\to1} \cdot \mathbf{z}|} \tag{5.2}$$

$$\mathbf{y} = \mathbf{z} \times \mathbf{x} \tag{5.3}$$

$$\mathbf{z} = \frac{\mathbf{r}_{2\to3}}{|\mathbf{r}_{2\to3}|} \tag{5.4}$$

where $|\mathbf{u}| = \sqrt{\mathbf{u} \cdot \mathbf{u}}$ is the Euclidian norm of vector $\mathbf{u}$, $\cdot$ is the scalar product and $\times$ is the cross product.

It follows that the dihedral angle $\phi$ (see Fig. 5.3) is given by its cosine and sine

$$\cos\phi = \frac{\mathbf{r}_{3\to4}}{|\mathbf{r}_{3\to4}|} \cdot \mathbf{x} \tag{5.5}$$

$$\sin\phi = \frac{\mathbf{r}_{3\to4}}{|\mathbf{r}_{3\to4}|} \cdot \mathbf{y} \tag{5.6}$$

We chose to study the couple of helices containing at least one of helix 5, 7, 8, 10 or 12. These helices were selected because they are the longest with at least 18 residues. Then, it what follows we only show some specific couples that were identified as showing the largest difference between two states for the considered collective variables.

## 5.2 Study of distances between couple of helices

**Study of distances in crystallographic structures** Table 5.2 shows all the distances $L^{i-j}_{State}$ between the centers of mass of two helices that exhibited measurable differences during simulations. We also introduce the differences

$$\Delta_1^{i-j} = |L^{i-j}_{PPS} - L^{i-j}_{\mathrm{P_iR}}| \qquad \Delta_2^{i-j} = |L^{i-j}_{\mathrm{P_iR}} - L^{i-j}_{SA}| \qquad \Delta_3^{i-j} = |L^{i-j}_{PPS} - L^{i-j}_{SA}|, \tag{5.7}$$

which represent the pairwise differences between the values $L^{i-j}_{state}$ obtained for PPS, P$_\mathrm{i}$R and Strong-ADP, see Fig. 5.4. The state that is the farthest from the two others is indicated in the last column of Tab. 5.2. For instance is the case illustrated in Fig. 5.4, P$_\mathrm{i}$R is the farthest from the two others.

The distances that can be used to differentiate between two states are the ones associated with the largest $\Delta_i$. For instance, $\Delta_3^{i-j}$ as large as possible should be large to distinguish between PPS from Strong-ADP. According to table 5.2, the 5 largest values of $\Delta_3^{i-j}$ are for the helix couples 6-12, 9-12, 3-12, 14-12 and 13-14. The corresponding $L^{i-j}$ distances are therefore the most interesting for guiding the transformation from PPS to Strong-ADP.

**Study of distances along relaxations** Fig. 5.5 shows the boxplots corresponding to the distributions of the distances $L^{i-j}_{state}$ recorded along relaxation. for a selection of helices couples

| Helices i-j | $L_{PPS}^{i-j}$ | $L_{P_iR}^{i-j}$ | $L_{SA}^{i-j}$ | $\Delta_1^{i-j}$ | $\Delta_2^{i-j}$ | $\Delta_3^{i-j}$ | Max diff |
|---|---|---|---|---|---|---|---|
| 2-7 | 38.6 | 38.7 | 38.6 | 0.1 | 0.1 | 0.0 | $P_iR$ |
| 13-7 | 38.5 | 38.1 | 37.1 | 0.4 | 1.0 | 1.4 | SA |
| 4-8 | 33.3 | 34.0 | 34.0 | 0.7 | 0.0 | 0.7 | PPS |
| 5-8 | 33.8 | 34.4 | 34.0 | 0.6 | 0.4 | 0.2 | PPS |
| 5-10 | 37.0 | 36.3 | 35.6 | 0.7 | 0.7 | 1.4 | PPS & SA |
| 6-10 | 40.5 | 40.2 | 38.1 | 0.3 | 2.1 | 2.4 | SA |
| 9-10 | 35.3 | 36.5 | 38.0 | 1.2 | 1.5 | 2.7 | SA |
| 3-12 | 24.3 | 23.8 | 26.2 | 0.5 | 2.4 | 1.9 | SA |
| 4-12 | 32.3 | 31.7 | 32.6 | 0.6 | 0.9 | 0.3 | $P_iR$ |
| 5-12 | 31.9 | 32.1 | 31.0 | 0.2 | 1.1 | 0.9 | SA |
| 6-12 | 39.9 | 40.5 | 38.3 | 0.6 | 2.2 | 1.6 | SA |
| 7-12 | 22.8 | 23.2 | 20.8 | 0.4 | 2.4 | 2.0 | SA |
| 9-12 | 26.9 | 27.1 | 28.4 | 0.2 | 1.3 | 1.5 | SA |
| 14-12 | 43.8 | 44.2 | 49.7 | 0.4 | 5.5 | 5.9 | SA |
| 5-13 | 43.4 | 43.1 | 41.9 | 0.3 | 1.2 | 1.5 | SA |
| 13-14 | 41.6 | 43.4 | 38.6 | 1.8 | 4.8 | 3.0 | SA |

Table 5.2: Values $L_{state}^{i-j}$ of distances in Å between Center of Mass (COM) of helix i and COM of helix j (see Fig. 5.1) in the crystal structures of PPS, $P_iR$ and Strong-ADP. $\Delta_1^{i-j}$, $\Delta_2^{i-j}$ and $\Delta_3^{i-j}$ represent the difference of $L_{state}^{i-j}$ between PPS and $P_iR$, between $P_iR$ and Strong-ADP and between PPS and Strong-ADP, respectively. The last column gives the state giving the biggest difference with the two others, as described in Fig. 5.4, which is also the state not involved in the smallest $\Delta^{i-j}$.

among the ones shown in Tab. 5.2. Among the distances introduced in Tab. 5.2, we retained only the ones showing the maximum differences and the least overlap between the states to construct Fig. 5.5. These distances are thus the best candidates to discriminate between the states.

Based on these results, we obtain that the distances $L^{6-10}$, $L^{9-10}$, $L^{5-12}$, $L^{6-12}$, $L^{7-12}$, $L^{14-12}$ and $L^{13-14}$ are adequate to guide the transformation from PPS to Strong-ADP, see Fig. 5.5(a).

Similarly The distances $L^{4-8}$, $L^{5-8}$ and $L^{4-12}$ were found the best to distinguish the $P_iR$ state from the two others, see Fig. 5.5(b). However, the latter distances are less precise at isolating $P_iR$, as compared to the other distances hability to isolate PPS, due to the important

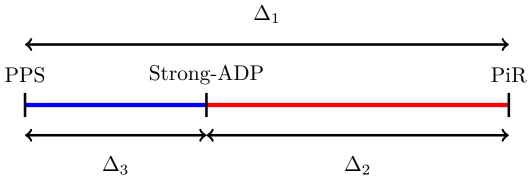


Figure 5.4: Graphical representation of $\Delta^{i-j}$ in the case the value in $P_iR$ is the most different compared to PPS and Strong-ADP. This means that $\Delta_1^{i-j}$ and $\Delta_2^{i-j}$ are both larger than $\Delta_3^{i-j}$. This is equivalent to say that $\Delta_3^{i-j}$ is lower than both $\Delta_1^{i-j}$ and $\Delta_2^{i-j}$.

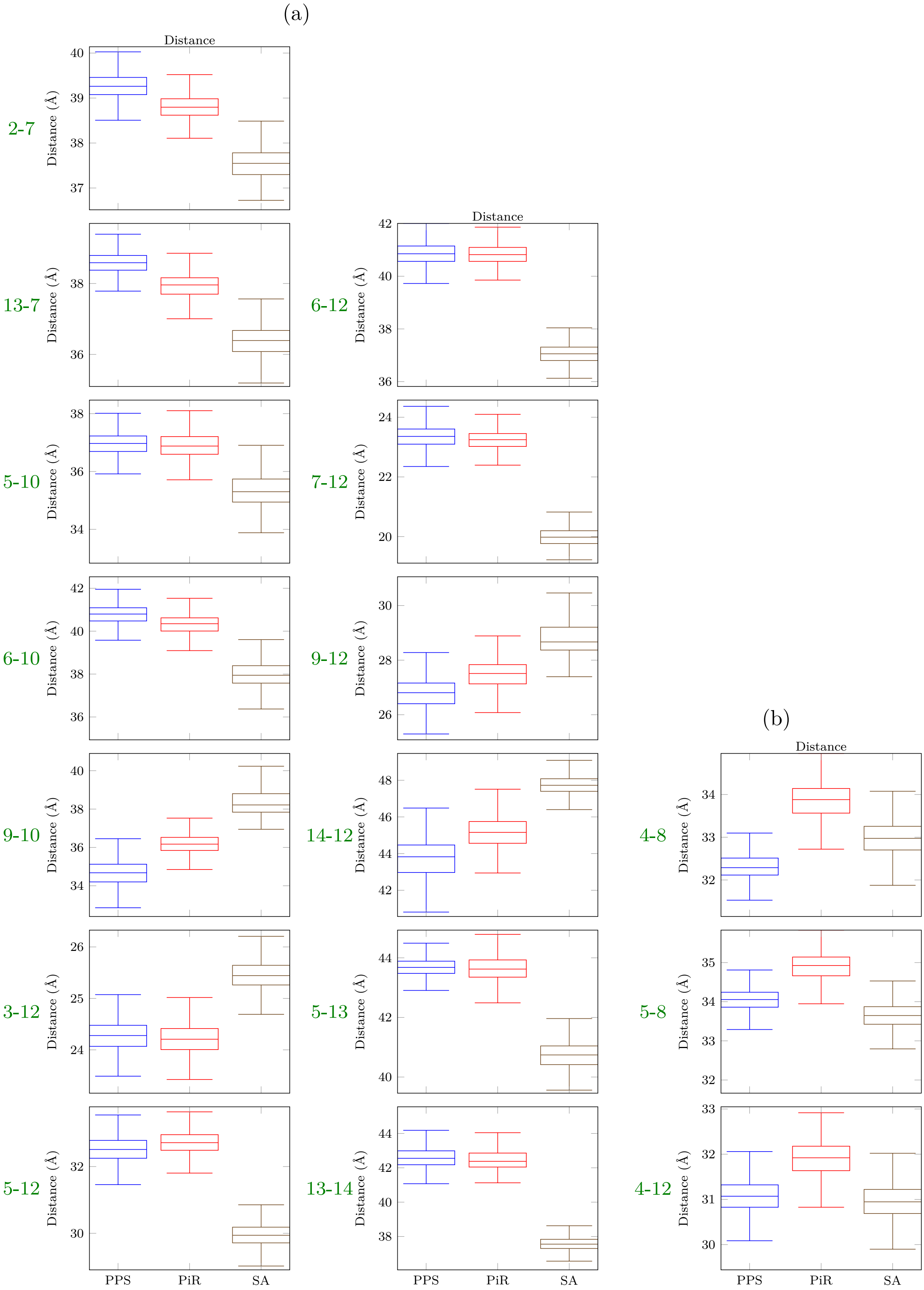


Figure 5.5: Representation with boxplots of distances $L^{i-j}$ between Center of mass (COM) of chosen helix's backbone (helix i) and COM of reference helix's backbone (helix j) along unconstrained dynamics of PPS (blue), $\mathrm{P_iR}$ (red) and Strong-ADP (brown). (a) Distances allowing discriminating PPS from Strong-ADP, (b) Distances allowing best discrimination of $\mathrm{P_iR}$.

overlap of the distributions.

## 5.3 Study of dihedrals between couple of helices

**Study in crystallographic structures** We denote as $D_S^{i-j}$ the dihedral angles between helices $i$ and $j$ in the crystal of state $S$. The values of a selection of such angles for PPS, P$_\mathrm{i}$R, and Strong-ADP structures are reported in Table 5.3. The selection was made based on the results of the relaxation molecular dynamics as described below. From these results, 14-10, 13-14 and 11-14 are found to be the best to discriminate Strong-ADP from PPS and P$_\mathrm{i}$R, see Fig. 5.6. However, it is not possible to discriminate PPS from P$_\mathrm{i}$R.

| Helices i-j | $D_{PPS}^{i-j}$ | $D_{P_\mathrm{i}R}^{i-j}$ | $D_{SA}^{i-j}$ |
|---|---|---|---|
| 2-7 | 115 | 121 | 101 |
| 13-7 | 81 | 80 | 61 |
| 5-10 | -113 | -106 | -120 |
| 14-10 | -49 | -56 | 9 |
| 5-12 | 11 | 18 | 8 |
| 6-12 | -134 | -117 | -147 |
| 9-12 | 168 | 177 | -176 |
| 14-12 | -80 | -82 | 87 |
| 5-13 | -85 | -86 | -103 |
| 13-14 | -1 | 10 | 98 |
| 11-14 | 64 | 66 | 161 |

Table 5.3: Values in degree of the dihedrals $D_{state}^{i-j}$ between the helices i and j in crystal states of PPS, P$_\mathrm{i}$R and Strong-ADP.

**Evolution of the dihedrals angles along relaxation** The boxplot of the selected dihedral angles between helices along unconstrained dynamics are shown on Fig. 5.6. As described above, this selection is based on the criterion that it exists sufficient difference between two states among PPS, P$_\mathrm{i}$R and Strong-ADP compared to fluctuations, allowing discrimination of at least one state. A direct observation of the boxplot shows that the biggest $\Delta_3^{i-j}$ are for helix couples 14-10, 5-13 and 11-14. This confirms some of the results from the crystal structures, as the dihedrals 14-10 and 11-14 were among the dihedrals with the most different values between PPS (or P$_\mathrm{i}$R) and Strong-ADP. However the dihedral 5-13 did not show large difference in crystal structures.

## 5.4 Comparison of consecutive helices

To characterize the relative position of consecutive helices, distances $L^{i-j}$, angles $A^{i-j}$ and dihedrals $D^{i-j}$ were computed for the 14 consecutive couples between the 15 helices.

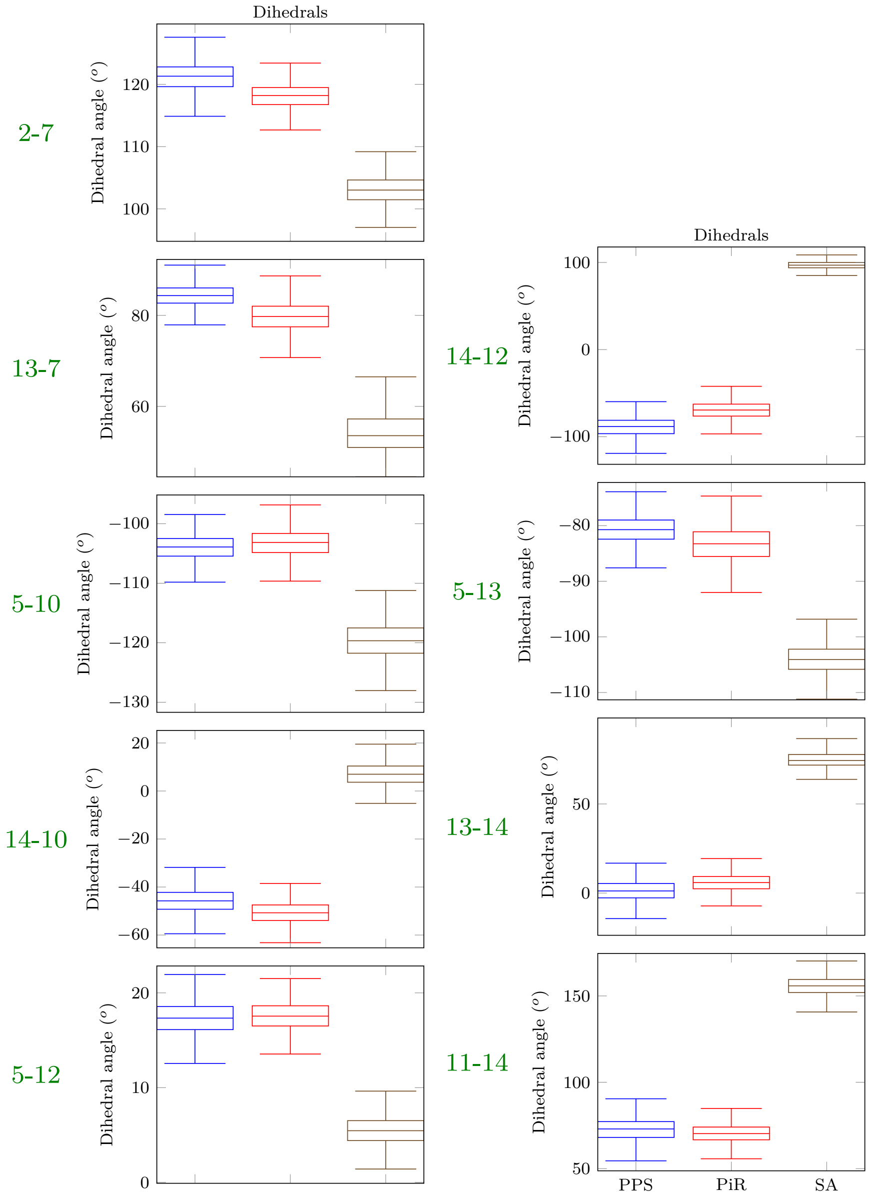


Figure 5.6: Representation with boxplots of selected dihedral angles $D^{i-j}$ between helices along relaxation of PPS (blue), P$_\mathrm{i}$R (red) and Strong-ADP (brown).

Figure 5.7: Parameters between consecutive helices in PPS (blue), P$_i$R (red) and Strong-ADP (brown). There are shown only the values for the one showing a difference between states and one example of each of the one said to be stables.

It should be taken into account that the scales are not uniform between the graphs, in order to more explicitly show the non-overlapping between boxplots (especially for the angles).

We consider that the helices i and j of the couples are stable relative to each other when the boxplots are completely overlapping. The values are said to be different when the boxplots are not overlapping at all between PPS (or $P_iR$) and Strong-ADP.

The distributions of the collective variables characterizing the relative positioning of the selected consecutive helices are shown on Fig. 5.7.

We observe that the geometric collective variables allowing to discriminate between the state are:

- angle and dihedral between helices 3 and 4,
- angle between helices 7 and 8 (8 deg),
- distance between helices 9 and 10,
- and the distance, angle and dihedral between helices 13 and 14.

Furthermore, some consecutive helices can be considered as part of the same "rigid body" as they show very good overlap between visited values along relaxation in the three states PPS, $P_iR$ and Strong-ADP. A typical example is between helices 1 and 2, see the first row of Fig. 5.7. The couples of helices that can be considered stables relative to each other are :

- 1-2, 2-3, 4-5, 10-11, 14-15 (very stable)
- quite stable but slight difference : 5-6, 6-7, 11-12 (stable with small fluctuations)
- 8-9, 12-13 (large fluctuations)

Unsurprisingly, we find among stable couples of consecutive helices the ones that are part of the same myosin subdomains, whereas the couples of helices 3-4 , 7-8 and 13-14 that exhibited differences between the simulations are part of different subdomains. The couple 12-13 also concern helices that are part of different subdomains, which could explain why we observed large fluctuations although we defined it stable. These observations thus lead us to conclude that the relative positions of helices can be used to define rigid domains. However, a limit to this claim has to be formulated as the couple 9-10 that is in the L50 subdomain exhibits discriminating behaviour between PPS and Strong-ADP.

## 5.5 Conclusion

In this chapter we defined a methodology based on the relative positions of the helices of myosin to describe the protein conformation. This description led us to identify collective variables that could be employed to identify whether the protein evolves close to some characteristics of PPS, $P_iR$ or Strong-ADP. The advantage of this description is that it could be employed on any system without a priori knowledge on its functioning, as it only necessitates to identify the longest helices, which identification could be performed by simple visualization of structures.

# Chapter 6

# Study of the dynamics of PPS, $P_i$R and Strong-ADP with rigid body and ramachandran angles

In this chapter we study the same simulations as in the chapter 4 and 5. Similarly to the chapter 5, in this chapter we develop here two methodologies to characterize the whole protein. The first methodology is based on the relative position of domains of the protein defined as rigid bodies. This relative position is defined using bases associated to each domain. Then, their relative position is described by the vector between their centers of mass and the rotation between their bases.

Another methodology is developed to study the behaviour of the myosin and characterize the difference between its different states. This methodology is based on the comparison of the distribution of the backbone dihedrals $\phi$ and $\psi$ on the Ramachandran space.

## Contents

## 6.1 Use of rigid bodies methodology to describe protein conformation

### 6.1.1 Presentation of the methodology for the study of the protein dynamics with rigid bodies

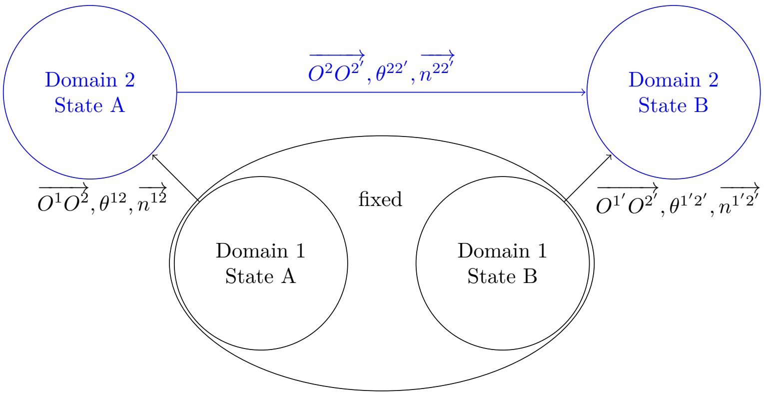


Figure 6.1: Representation of the methodology for the transformation from one state of the protein into another based on domains, which are assumed rigid. We consider one domain as a reference, thus fixed (domain 1 in this example), while the other (domain 2) is mobile. We start by calculating the transformation from Domain 1 to Domain 2 in both considered states. The transformation consists in a translation $\overrightarrow{O^1O^2}$ and a rotation $\theta^{12}$ around the axis $\overrightarrow{n^{12}}$. We can then compute a transformation on domain 2 from the state A to the state B. This transformation is composed of a translation $\overrightarrow{O^2O^2}$ and a rotation $\theta^{22'}$ around the axis $\overrightarrow{n^{22'}}$.

It is assumed that the large changes of protein structure between two states (e.g. PPS, P$_\mathrm{i}$R and Strong-ADP) can be formulated as rigid motions of protein domains. This methodology reduces the problem of positioning the $N_d$ domains from 3 $N_{atoms}$ degrees of freedom (3 coordinates of each atom) to a problem in which solids are positioned with respect to each other with only $6(N_d - 1)$ degrees of freedom.

The way these domains are defined is explained in the following subsection. As represented on Fig. 6.1, one domain (Domain 1) is assumed fixed while the other is considered mobile[1] (Domain 2). Domain 1 is then considered as the reference domain.

Domains 1 (resp. Domain 2) is associated to a centroid, also known as geometric center, noted $O^1$ (resp. $O^2$), and one orthonormal basis $B^1$ (resp. $B^2$) respectively, together forming

---

[1] This definition of a fixed domain and a mobile domain is due to the fact that the protein is moving and rotating along the simulations in the reference frame of the simulation box. Thus, we cannot define "absolute" transformation in the reference frame of the simulation box. This induces that we need a reference domain that we can consider "fixed", which is the reason we define everything in the reference frame of the fixed domain.

the reference frame $R^1 = (O^1, B^1)$ (resp. $R^2 = (O^2, B^2)$). We start by calculating the position of Domain 1 with respect the Domain 2 in both considered states. The transformation consists in one position vector $\overrightarrow{O^1O^2}$ and one rotation $\theta^{12}$ around the axis $(O^2, \overrightarrow{n^{12}})$. Then we can characterize the transformation that allows to transform Domain 2 from its state A into state B, with the translation vector $\overrightarrow{O^2O^{2'}}$ and one rotation $\theta^{22'}$ around the axis $(O^{2'}, \overrightarrow{n^{22'}})$. Here $O^{2'}$ is the centroid of Domain 2 in state B.

#### 6.1.1.1 Rigid body representation of the domains

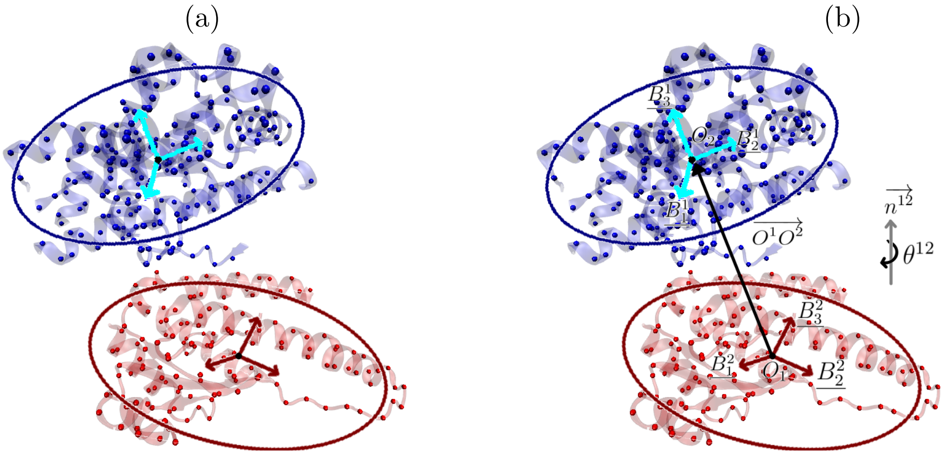


Figure 6.2: (a) Example of reference frame associated to U50 and L50 and the equivalent ellipsoid (b) Example of transformations with a translation and a rotation.

The Domain D position is defined by 6 degrees of freedom: 3 for position of centroid $O^D$ and 3 for the orientation of the associated orthonormal basis $B^D$). The centroid $O^D$ of the Domain $D$ is computed as follows:

$$\overrightarrow{OO^D} = \frac{1}{N_D} \sum_{(i) \in D} \overrightarrow{r}^{(i)} \tag{6.1}$$

where $O$ is the origin of simulation box, the term '$(i) \in D$' means the $i$th backbone atoms in Domain D and $N_D$ is the number of backbone atoms in Domain D.

The basis $B^D$ is defined from three non-aligned points of Domain D. Let name this three points $A^D$, $B^D$ and $C^D$. The vector $\overrightarrow{A^DB^D}$ is normalized, which is then noted $\overrightarrow{u}^{(1)}$. Let define $\overrightarrow{u}^{(2)}$ the normalized part of $\overrightarrow{A^DC^D}$ noncollinear to $\overrightarrow{u}^{(1)}$ as :

$$\overrightarrow{u}^{(2)} = \frac{\overrightarrow{A^DC^D} - \overrightarrow{A^DC^D} \cdot \overrightarrow{u}^{(1)}}{\left| \overrightarrow{A^DC^D} - \overrightarrow{A^DC^D} \cdot \overrightarrow{u}^{(1)} \right|} \tag{6.2}$$

where $\cdot$ denotes the dot (scalar) product, and the $|\overrightarrow{a}| = \sqrt{\overrightarrow{a} \cdot \overrightarrow{a}}$ is the Euclidean norm. The third unit vector $\overrightarrow{u}^{(3)}$ is defined as the cross product of $\overrightarrow{u}^{(1)}$ and $\overrightarrow{u}^{(2)}$, such as

$$\overrightarrow{u}^{(3)} = \overrightarrow{u}^{(1)} \times \overrightarrow{u}^{(2)} \ . \tag{6.3}$$

It is worth mentioning here that another method using the eigenvectors of the gyration matrix has been initially used. This approach was not satisfactory because of the non-uniqueness of the definition of eigenvectors, when there are two or three equal eigenvalues in the gyration matrix.

Then, we define the rotation matrix of domain $D$

$$\mathbf{R}^{(D)} = \begin{pmatrix} u_1^{(1)} & u_1^{(2} & u_1^{(3)} \\ u_2^{(1)} & u_2^{(2} & u_2^{(3)} \\ u_3^{(1)} & u_3^{(2} & u_3^{(3)} \end{pmatrix}_{,B^0} \tag{6.4}$$

where $\overrightarrow{u}_1, \overrightarrow{u}_2$ and $\overrightarrow{u}_3$ are expressed in simulation box base, $B^0$. This matrix provides the coordinates of vector $\overrightarrow{a}_{,B^0}$ expressed in simulation box base into the base $B^D$ associated to the Domain D, by the following expression

$$\overrightarrow{a}_{,B^D} = \mathbf{R}^{(D)} \overrightarrow{a}_{,B^0} \tag{6.5}$$

#### 6.1.1.2 Relative position of domains with the respect the reference Domain 1

Once the domain positions are defined, we can compute the relative positions of the domains in comparison to a reference frame. This reference frame $R^1$ is taken from the domain considered fixed (Domain 1), which follows the global movement of the protein in the global reference frame.

The relative position of Domain $D$ with respect to the reference Domain 1 is obtained by a change of frame. The centroid position of Domain $D$ in reference frame $R^1$ is given by

$$\overrightarrow{O^1O^D} = \overrightarrow{OO^D} - \overrightarrow{OO^1} \tag{6.6}$$

The rotation matrix of Domain D with respect the reference frame $R^1$ is

$$\mathbf{R}^{(1 \to D)} = \mathbf{R}^{(D)} . \left( \mathbf{R}^{(1)} \right)^T \tag{6.7}$$

where the superscript $^T$ denotes matrix transposition.

From rotation matrix $\mathbf{R}^{(1 \to D)}$, we define :

**Rotation axis** The unit vector representing the direction of the rotation axis $\overrightarrow{n^{1D}}$ is determined by taking the eigenvector associated to the purely real eigenvalue (equal to 1) from $\mathbf{R}^{(1 \to D)}$. Indeed, this eigenvector corresponds to the direction which is not impacted by the rotation, i.e. in direction of the rotation axis.

**Rotation angle** $\theta[D-1]$ of Domain D with the respect of reference Domain 1is computed as follows:

$$\theta[D-1] = \cos^{-1}\left(\frac{\mathrm{tr}\left[\mathbf{R}^{(1\to D)}\right]-1}{2}\right) \tag{6.8}$$

Here, the $\theta[D-1]$ angle is defined modulo $\pi$.

#### 6.1.1.3 Definition of Domain D motion between two different states

Here, the position of Domain D in state A (resp. in state B) is defined by the vector $\overrightarrow{O^1_{(A)}O^D_{(A)}}$ (resp. $\overrightarrow{O^1_{(B)}O^D_{(B)}}$) and the rotation matrix $\mathbf{R}^{(1\to D)}_{(A)}$ (resp. $\mathbf{R}^{(1\to D)}_{(B)}$). The key idea to describe the motion of Domain D from state A to state B is to impose a fixed reference Domain 1, which implies $O^1_{(A)} = O^1_{(B)}$ and $B^1_{(A)} = B^1_{(B)} = B^1$ . Therefore, the motion of Domain D from state A to state B is define by:

**the vector position** $\overrightarrow{O^D_{(A)}O^D_{(B)}}$ expressed in the reference base $B^1$, such as

$$\overrightarrow{O^D_{(A)}O^D_{(B)}} = \overrightarrow{O^1_{(B)}O^D_{(B)}} - \overrightarrow{O^1_{(A)}O^D_{(A)}} \ , \tag{6.9}$$

**the rotation matrix** of Domain D from state A to state B defined by

$$\mathbf{R}^D_{(A\to B)} = \mathbf{R}^{(1\to D)}_{(B)} . \left(\mathbf{R}^{(1\to D)}_{(A)}\right)^T \ . \tag{6.10}$$

The axis of rotation and the rotation angle are calculated following the same procedure as in previous paragraph.

### 6.1.2 Definition of the rigid Domain

Rigid Domain are defined by searching for the selection of residues in a part of the protein in order to minimize the RMSD of backbone atoms (C, N, $C_\alpha$ & O) when comparing crystal structures in different states. This definition is thus based on static information. Selected domains are represented in the whole protein on Fig. 6.3.a. $C_\alpha$ atoms part of U50 and L50 are represented on Fig. 6.3.b to illustrate the fact that we only take backbone atoms for the definition of domains and their transformations. The residue ranges and the number of residues in each of the selected domains are summarized in the table 6.1.

| Domain | Residue ranges | Number of residues |
|---|---|---|
| Nter | 57-172, 680-705 | 142 |
| U50 | 183-191, 228-305, 313-352, 365-392, 413-432, 602-620 | 194 |
| L50 | 460-530, 540-598, 643-660 | 148 |
| Conv | 706-770 | 65 |

Table 6.1: Definition of the Domain by residue ranges and number of residues.

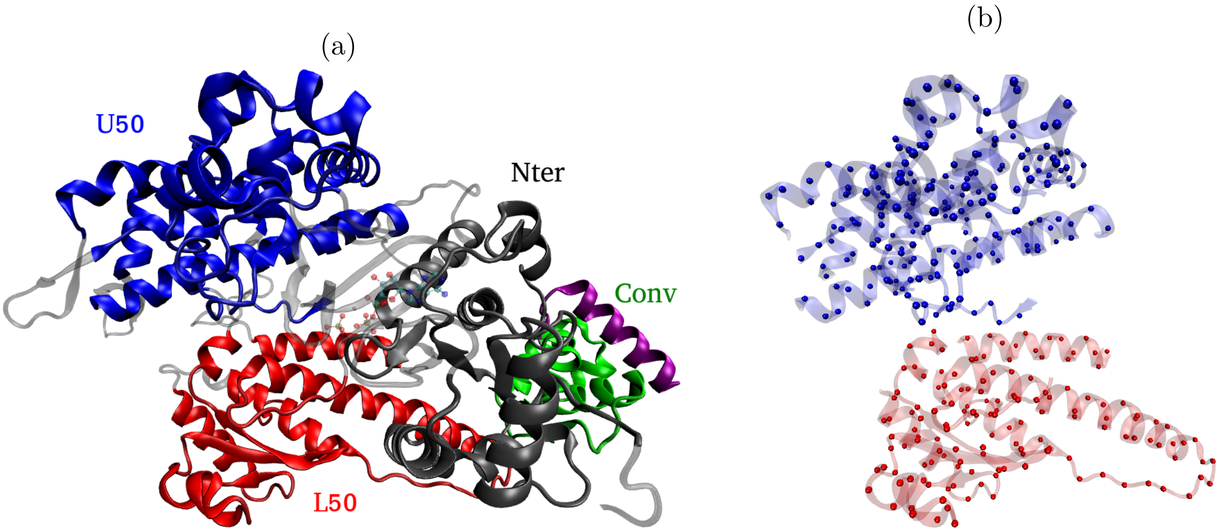


Figure 6.3: Representation of the four domains assumed as rigid. U50 in blue, L50 in red, Nter in black, Conv in green. The terminal helix in purple is the insert 2. The definition of these domains notably remove the actin binding loops, and most connectors between domains, which are too flexible to be considered rigid. (b) Representation of U50 and L50 with the $C_\alpha$ represented with balls.

### 6.1.3 Positions of rigid Domains for crystal structures (PPS, $P_i$R and Strong-ADP)

The distances between the centroids of the rigid Domain in the crystal structures are given in table 6.2, with the following notation D[Conv-Nter] to indicate the distance between the centroids of Domain Conv and Domain Nter. We observe that the distance D[Conv-Nter] is much smaller in Strong-ADP than in PPS or $P_i$R (see column light blue in the Table 6.2). This allows discriminating PPS and $P_i$R from Strong-ADP. The second largest difference is observed for D[U50-Conv], which also allows discriminating Strong-ADP from the others (see column light red in the Table 6.2).

| | D[U50-Conv] | D[L50-Conv] | D[L50-U50] | D[L50-Nter] | D[U50-Nter] | D[Conv-Nter] |
|---|---|---|---|---|---|---|
| PPS | 61.5 | 43.8 | 33.9 | 27.0 | 38.8 | 27.4 |
| $P_i$R | 61.9 | 44.5 | 34.3 | 28.4 | 38.6 | 27.8 |
| Strong-ADP | 64.3 | 44.2 | 33.1 | 27.4 | 40.1 | 24.4 |
| Max. Diff. | 2.4 | 0.7 | 1.2 | 1.4 | 1.5 | 3.4 |

Table 6.2: Distances in (Å) between centroid (COM) from the four domains : Upper 50kDa, Lower 50kDa, N-terminal and Converter in the crystal structures. Maximum difference and corresponding discriminating state are highlighted

Table 6.3 summarizes the rotation angle between the rigid Domain in the crystal structures of PPS, $P_i$R and Strong-ADP, with the following notation $\theta$[Conv-Nter] to indicate the rotation angle between the Domain Conv and the Domain Nter. The greatest differences of orientation exist between U50 and Conv Domains, L50 and U50 Domains, L50 and Conv Domains, and

U50 and Nter Domains. There is no change observed between L50 and Nter Domains, and Conv and Nter Domains.

| | $\theta$[U50-Conv] | $\theta$[L50-U50] | $\theta$[L50-Conv] | $\theta$[L50-Nter | $\theta$[U50-Nter] | $\theta$[Conv-Nter] |
|---|---|---|---|---|---|---|
| PPS | 179 | 62 | 154 | 175 | 122 | 91 |
| $P_iR$ | 172 | 73 | 171 | 179 | 123 | 95 |
| Strong-ADP | 149 | 55 | 158 | 174 | 140 | 94 |
| Max. Diff. | 30 | 18 | 17 | 5 | 18 | 4 |

Table 6.3: Angles (in degree) between the four subdomains : Upper 50kDa, Lower 50kDa, N-terminal and Converter in the crystal structures. Max. Diff. represents the maximum difference existing between two crystal structures modulo $\pi$.

### 6.1.4 Comparison along unconstrained dynamics

In this section, the relative positions of each Domain are monitored along the same unconstrained simulations of 300 ns as used in Chapters 4 and 5 (see Section §4.2 for details).

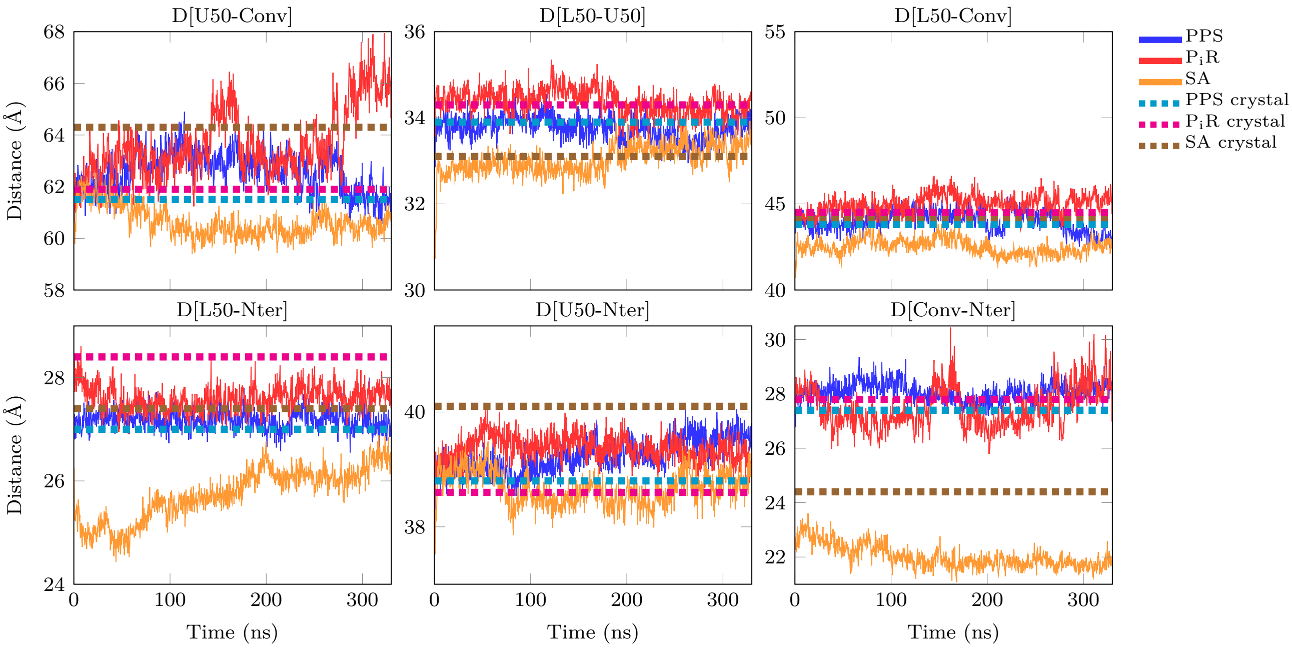


Figure 6.4: Evolution of the distance in Å between the domains Lower 50 kDa, Upper 50kDa, N-terminal and Converter during unconstrained dynamics.

**Position of Domain centroids** The evolution of distances between centroid of the four Domains along unconstrained simulations are represented on Fig. 6.4. Most distances do not show clearly separated values that would allow to differentiate PPS, $P_iR$ or Strong-ADP.

In particular, the distances D[L50-U50] and D[U50-Nter] are closer than the crystallographic structures, they could be considered equal within fluctuations. The distance D[U50-Conv] at the start of the simulations is similar in every states, however during simulation Strong-ADP explores values that are shorter than in PPS and $\mathrm{P_iR}$. This is surprising as in the crystal structure this distance was different (and larger) in Strong-ADP compared to the other states. Conversely, the distance D[L50-Nter] shows a difference between Strong-ADP and PPS/$\mathrm{P_iR}$, while tending to approach values obtained with PPS and $\mathrm{P_iR}$. Furthermore, we observe that none of these distances allow to differentiate PPS from $\mathrm{P_iR}$. However, the distances D[L50-Conv] and even more D[Conv-Nter] exhibits large differences between Strong-ADP and PPS/$\mathrm{P_iR}$, which could then help discriminate Strong-ADP. This last observation suggests that the distance D[Conv-Nter] can be used to discriminate Strong-ADP from others.

The relative position of a Domain D with respect to another reference Domain 1 is not only defined by a distance, but also by the direction of the vectors connecting the centroids $\overrightarrow{u}^{(D1)} = \overrightarrow{O^D O^1} / \left| \overrightarrow{O^D O^1} \right|$. Indeed, when the Domain D rotates around an axis passing through the Domain 1 centroid, $O^1$, the direction of the position vector, $\overrightarrow{O^D O^1}$ changes but not its norm.

To follow the evolution of the position vectors direction associated with each Domain with respect to the chosen reference Domain, we draw pole figures in Fig. 6.5. A reminder of the pole figures is given in the appendix C. Each point on the pole figure corresponds to position vector direction for a snapshot taken from the unconstrained simulations. The fact that the points remain concentrated in a small part of the pole figures means that the position vector directions remain the same during these unconstrained simulations. Moreover, when the color dots, relating to each structure, are superimposed, this means that the relative direction of the domains remains the same for the tree studied structures as for the position of Nter Domain with respect U50 Domain and L50 Domain with respect U50 Domain. In the middle pole figure, the yellow dots do not overlap with others, it illustrates that the vector position of the Domain Conv in reference frame of Domain U50 changes direction.

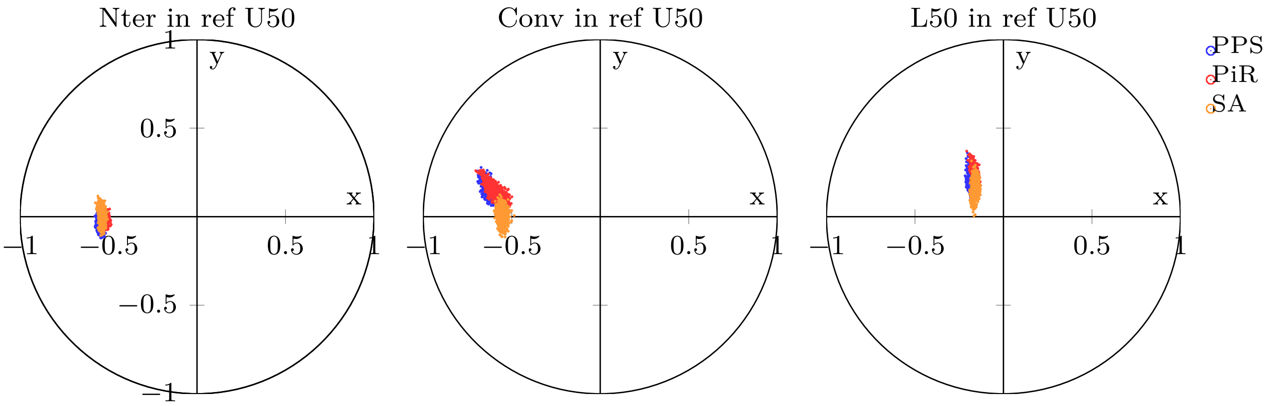


Figure 6.5: Pole figures show evolution of the direction of the position vector of centroid of Domains Lower 50 kDa, N-terminal and Converter in the frame reference of Upper 50 kDa Domain during unconstrained dynamics.

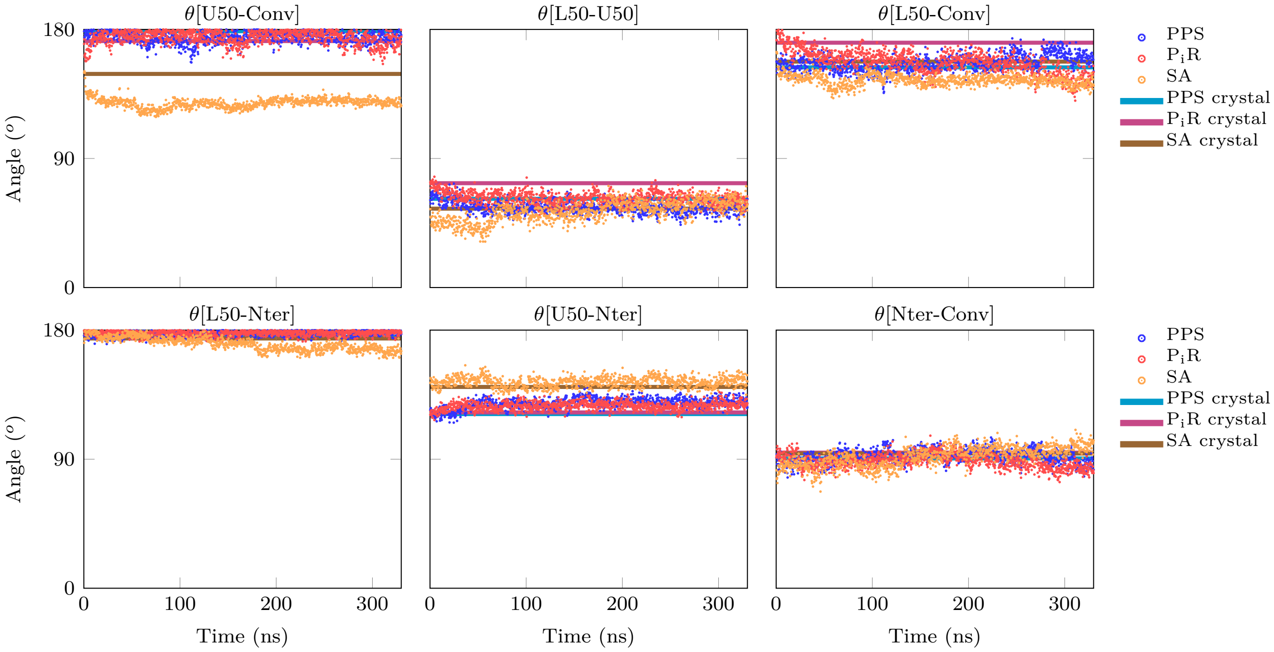


Figure 6.6: Evolution of the orientation angle $\theta[A-B]$ (define in Eq. 6.8) in degree between the domains Lower 50 kDa, Upper 50kDa, N-terminal and Converter during unconstrained dynamics. With the notation $\theta$[A-B], the rotation is defined by the orientation of domain A compared to the domain B. The values observed in the crystal structures of each states are represented with horizontal bars. (blue) PPS, (green) $\mathrm{P_iR}$, (magenta) Strong-ADP.

**Relative orientation of Domain** The evolution of angles between the four subdomains along unconstrained simulations are represented on Fig. 6.6. We observe a difference of angle $\theta$[U50-Conv] for Strong-ADP compare to its value in PPS and $\mathrm{P_iR}$. This angle $\theta$[U50-Conv] is also lower of ∼25° in the crystal structures.

The pole figures in Fig. 6.7 show the evolution of rotation axes direction in the frame of reference Domain U50. For the orientation of Nter Domain with respect the Domain U50, all the dots are superimposed in the small area of the left pole figure of Fig. 6.7, which implies that the direction of the axis of rotation is the same for the three structures during the unconstrained simulations. On the center and right pole figures in Fig. 6.7, the yellow dots are not in the same are of blue and red dots, which shows that the direction of rotation axe of Domain Conv and Domain L50 with respect the reference Domain U50 is different in the Strong-ADP unconstrained simulation. It is interesting to notice that although the distance of the yellow points from the others seems important on the middle pole figure, this is not really the case because of the periodicity of the direction space.

### 6.1.5 Interest of Domain position analysis

The relative positions and orientations of Domains being defined in the reference frame Domain, it is now possible to define a trajectory for each Domain to pass from one structure to another. Nevertheless, there remains an indeterminacy to remove because this trajectory is

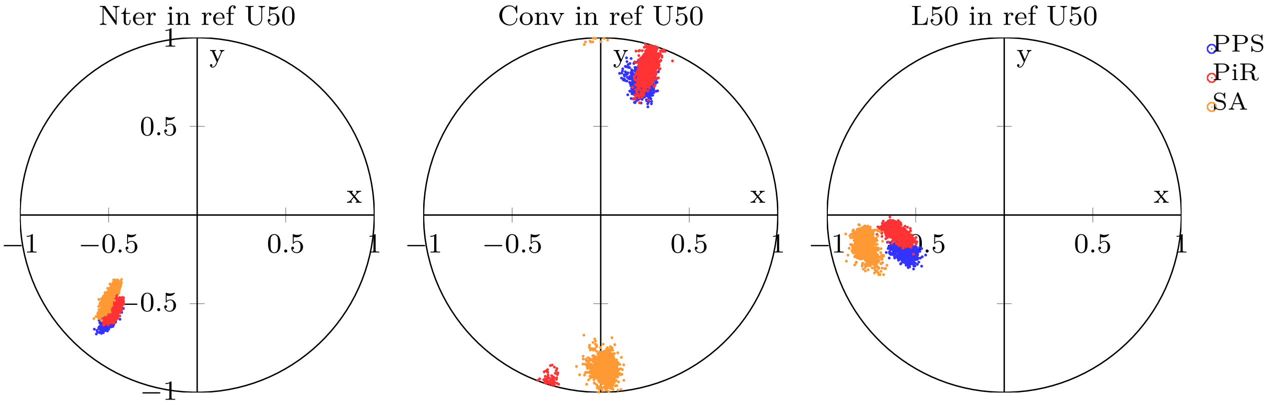


Figure 6.7: Evolution of the direction of the axis of rotation between the domains Lower 50 kDa, N-terminal and Converter in the frame reference of Upper 50 kDa during unconstrained dynamics.

not unique. A possible choice would be to decompose the Domain motions into a simultaneous translation and rotation to pass from one structure to another. To stay closer to the macromolecular architecture of myosin, we decided to study the conformational changes of the protein backbone, which could transform one structure into another and finding the relative positions and orientations identified in this study. This study of backbone conformation is the subject of the next section.

## 6.2 Characterisation of conformations by backbone dihedrals changes

In this methodology, we go one step farther in the lower scales by comparing individual residue dynamics and especially their backbone atoms between two time period from a single simulation or two independant simulations. Backbone conformation impacts the global conformation of the protein.

The spatial arrangement of the backbone atoms can be described partially by the two backbone dihedral angles, $\phi$ and $\psi$, see Fig. 6.8. These angles are defined for each amino acids [2]. The couple of value ($\phi$,$\psi$) can be represented in the Ramachandran plane (Ramachandran et al., 1963) (on the Ramachandran diagram), representing $\phi$ along the horizontal axis and $\psi$ along the vertical axis. We assumed that if a point ($\phi$,$\psi$) changes position between two simulation snapshots, the corresponding residue has changed conformation, and/or behaves differently. We thus compare the distribution of these points for a given time period for each residue individually. Actually, one more information can be inferred from the distributions. A

---

[2] An exception to the definition of $\phi$ and $\psi$ exists for the N-terminal and C-terminal parts that respectively are the first and last residue of a chain. Indeed, the N-terminal residue does not have a preceding residue and so the $\phi$ angle cannot be defined, while the C-terminal does not have a following residue, so the $\psi$ angle cannot be defined.

residue might have a large distribution in a certain time period and have a limited distribution in another, meaning that it was freely moving in the first and restricted in its torsion movement by some interactions in the other. The dihedral angles in general and then the backbone dihedrals are defined in following subsections.

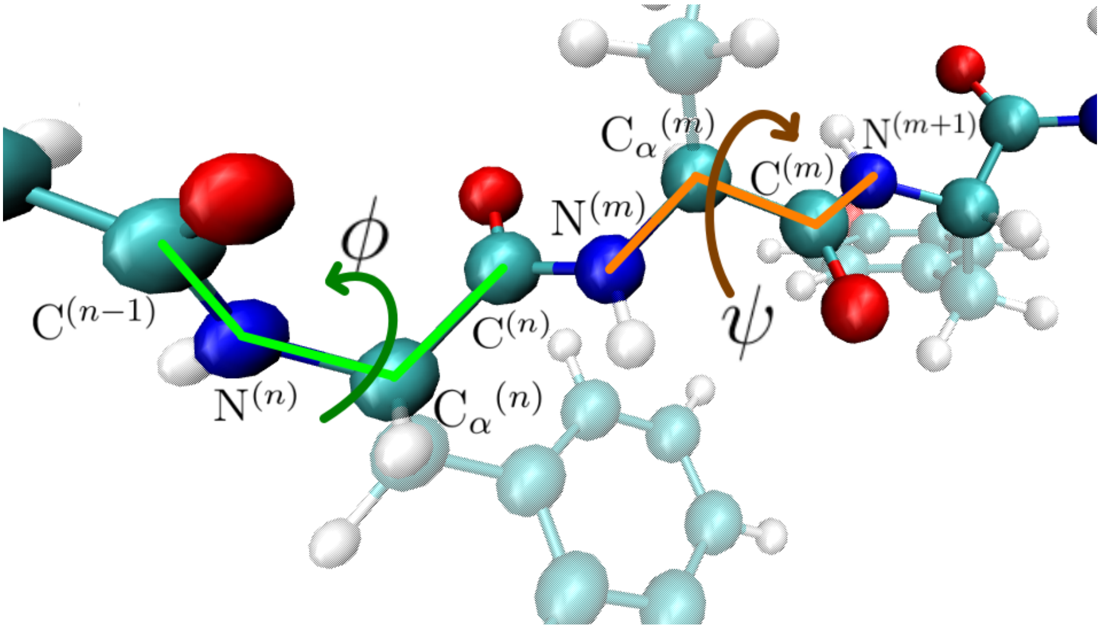


Figure 6.8: Representation of the backbone dihedral angles $\phi$ (green) and $\psi$ (orange). Carbon atoms are represented in cyan, oxygen atoms in red and nitrogen atoms in blue. These angles are represented on separate amino acids only to not superpose them.

### 6.2.1 Definition of dihedral angles

Dihedrals are angle in space based on 4 points, delimiting three segments. The way it is determined is represented on Fig. 6.9. A dihedral angle is the angle formed by the projection of the two end segments (blue and red) on a plane orthogonal to the central segment (segment in black), as represented on Fig. 6.9.b.

An alternative way to see it is the following: for someone looking from behind the first segment (1-2, segment in blue on the figure) in the direction of the central segment (2-3, in black), as represented with the eye on Fig. 6.9.a. What the observator would see is the figure on Fig. 6.9.b. The angle formed by the two segments is the dihedral angle. The convention for the value of the angle is to take as reference the first segment (1-2, in blue).

### 6.2.2 Definition of backbone dihedrals

The Figure 6.8 gives a representation of the backbone dihedral angles. The $\phi$ dihedral angle is formed by the sequence starting with the carboxylic carbon from the preceding residue, followed by the nitrogen, the $\alpha$-carbon and the carboxylic carbon from the current residue, and thus can be written:

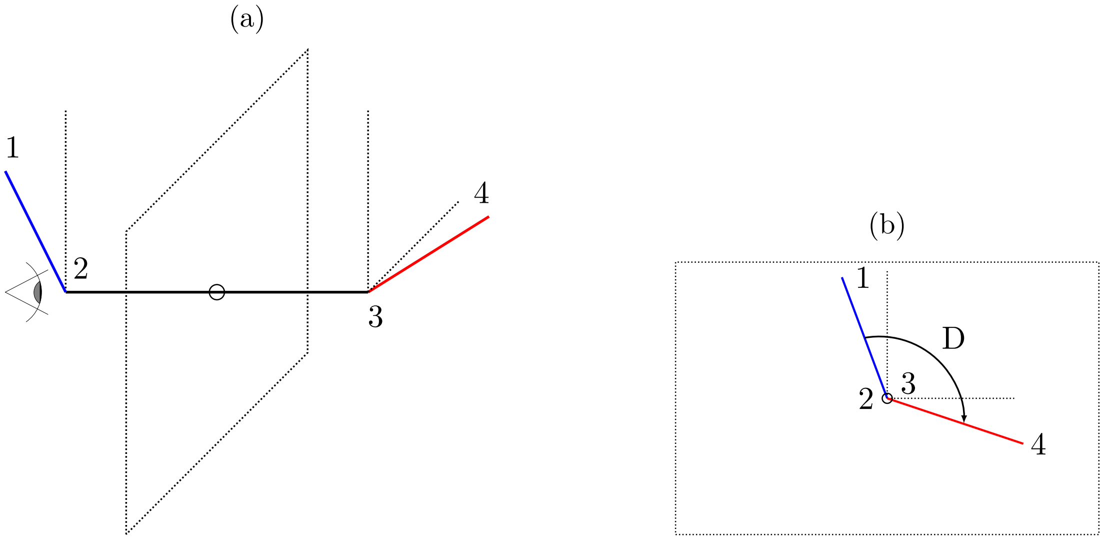


Figure 6.9: (a) Representation in a 3D space of a dihedral angle. (b) Representation of the dihedral angle with the projection of axes on the plane orthogonal to the central segment. Alternatively, view for an observator looking in direction of the segment 2-3 (in black). The dihedral angle corresponds to the angle formed between the segment blue and red.

$$C^{(n-1)}\text{-}N^{(n)}\text{-}C_{\alpha}{}^{(n)}\text{-}C^{(n)}$$

On the other hand, the $\psi$ angle is formed by the sequence starting with the nitrogen, followed by the $\alpha$-carbon and the carboxylic carbon all from the current residue and finishing with the nitrogen from the following residue, we can thus write it as follows:

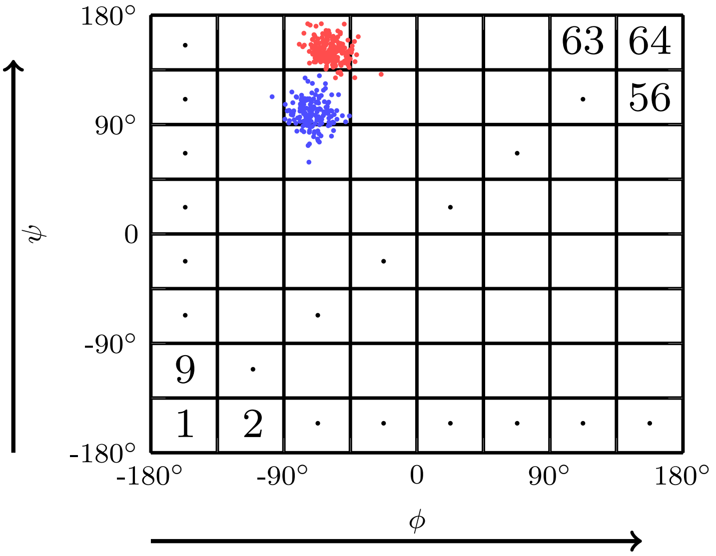
$$N^{(m)}\text{-}C_{\alpha}{}^{(m)}\text{-}C^{(m)}\text{-}N^{(m+1)}$$

As a matter of completeness, one could say that there exist a supplementary dihedral angle with the backbone atoms, this is the dihedral angle formed by $C_{\alpha}{}^{(n)}\text{-}C^{(n)}\text{-}N^{(m)}\text{-}C_{\alpha}{}^{(m)}$ on Fig. 6.8, usually noted $\omega$. However, this dihedral angle concerns the peptide bond, a central double covalent bond which does not permit much deviation from planarity, and leads to not having much fluctuations. The values and fluctuations were found to be around $179.5 \pm 3.8°$ (Karplus, 1996). This implies that there is not much degree of freedom around this angle, taking its slight fluctuations into account would not be useful to describe the backbone conformation.

### 6.2.3 Description of the Conformational Change (CC) analysis

The Conformation Change (CC) analysis aims at detecting conformational changes of an amino acid either along a simulation or between simulations of two different structures. Values for ($\phi$,$\psi$) are determined for each analysed snapshots of the trajectories obtained during unconstrained simulations (see Section §4.2 for details). For example, this represents one snapshot every 50 ps simulated time during relaxations of 300 ns, thus collecting 6000 points during the whole trajectory. The whole trajectory is divided into 20 time periods, $\Delta T_i$, thus representing 15 ns of simulated time each and 300 points.

(a)



(b)

$$\underline{V_n}[\Delta T_1, PPS] = \begin{bmatrix} V_n[\Delta T_1, PPS](1) \\ \vdots \\ V_n[\Delta T_1, PPS](43) \\ \vdots \\ V_n[\Delta T_1, PPS](50) \\ V_n[\Delta T_1, PPS](51) \\ V_n[\Delta T_1, PPS](52) \\ \vdots \\ V_n[\Delta T_1, PPS](59) \\ V_n[\Delta T_1, PPS](60) \\ \vdots \\ V_n[\Delta T_1, PPS](64) \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 27 \\ \vdots \\ 1 \\ 137 \\ 0 \\ \vdots \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

(c)

$$\underline{V_n}[\Delta T_2, PPS] = \begin{bmatrix} V_n[\Delta T_2, PPS](1) \\ \vdots \\ V_n[\Delta T_2, PPS](43) \\ \vdots \\ V_n[\Delta T_2, PPS](50) \\ V_n[\Delta T_2, PPS](51) \\ V_n[\Delta T_2, PPS](52) \\ \vdots \\ V_n[\Delta T_2, PPS](59) \\ V_n[\Delta T_2, PPS](60) \\ \vdots \\ V_n[\Delta T_2, PPS](64) \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ \vdots \\ 0 \\ 4 \\ 2 \\ \vdots \\ 151 \\ 8 \\ \vdots \\ 0 \end{bmatrix}$$

Figure 6.10: (a) Representation of the discretization of the Ramachandran plane in the methodology studying backbone conformation with $8^2 = 64$ equivalent boxes. Boxes are numbered following increasing values of $\phi$, followed by increasing values of $\psi$. (b) Example of vector $\underline{V_n}\,[\Delta T_1, PPS]$, for a $n$-th residue, containing 64 coordinates corresponding to the number of points for the blue distribution counted in each of the 64 boxes for PPS structure for the time period $\Delta T_1$ (c) Example of another conformation vector $\underline{V_n}\,[\Delta T_2, PPS]$, for the same residue and structure but another time period $\Delta T_2$ (red distribution). In this example, the conformation similarity indicator $S_n\,[\Delta T_1, PPS; \Delta T_2, PPS] = 0.0259$, which is low as expected in relation to the red and blue distributions pretty different.

In order to define the distribution of points, the Ramachandran plane is discretized in N$_{box}$ x N$_{box}$ square boxes of same dimensions which are represented on Fig. 6.10.(a) for an example of 8x8 boxes.

For a given time interval $\Delta T_i$ and a given structure (e.g. PPS), the distribution of the points is extracted by counting the number of points in each boxes of Ramachandran plane. The counts are assembled in the conformation vector of $n$-th amino acid in PPS structure during the $\Delta T_i$ time interval, $\underline{V_n}\left[\Delta T_i, PPS\right]$ of length N$_{box}$ x N$_{box}$.

With the conformation vector, it is possible to compare the conformation of a given amino acid at two different times of the same simulation by using the conformation similarity indicator $S_n\left[\Delta T_1, PPS; \Delta T_2, PPS\right]$S, given by:

$$S_n\left[\Delta T_1, PPS; \Delta T_2, PPS\right] = \frac{\underline{V_n}\left[\Delta T_1, PPS\right]}{\left|\underline{V_n}\left[\Delta T_1, PPS\right]\right|} \cdot \frac{\underline{V_n}\left[\Delta T_2, PPS\right]}{\left|\underline{V_n}\left[\Delta T_2, PPS\right]\right|} \ . \tag{6.11}$$

$S_n\left[\Delta T_1, PPS; \Delta T_2, PPS\right]$ would equal 1 for two absolutely identical conformational state of $n$-th residue, while being equal to 0 if the conformational state are completely different.

This indicator allow us to detected a conformation change of given residue during time period $\Delta T_i$ of two different structures. To do this, the conformation similarity indicators is defined by

$$S_n\left[\Delta T_i, PPS; \Delta T_i, P_{\mathrm{i}}R\right] = \frac{\underline{V_n}\left[\Delta T_i, PPS\right]}{\left|\underline{V_n}\left[\Delta T_i, PPS\right]\right|} \cdot \frac{\underline{V_n}\left[\Delta T_i, P_{\mathrm{i}}R\right]}{\left|\underline{V_n}\left[\Delta T_i, P_{\mathrm{i}}R\right]\right|} \ . \tag{6.12}$$

### 6.2.4 Study of backbone conformation differences

In this section, we apply the CC analysis to study the differences between backbone conformations of PPS, P$_{\mathrm{i}}$R and Strong-ADP unconstrained simulations. We chose the last time period, $\Delta T_f$, to compute the conformation similarity indicators for all residues, defined in Eq. (6.12).

The Figure 6.11 shows in red the residues, where there is a difference of conformation between the structure PPS, P$_{\mathrm{i}}$R and Strong-ADP.

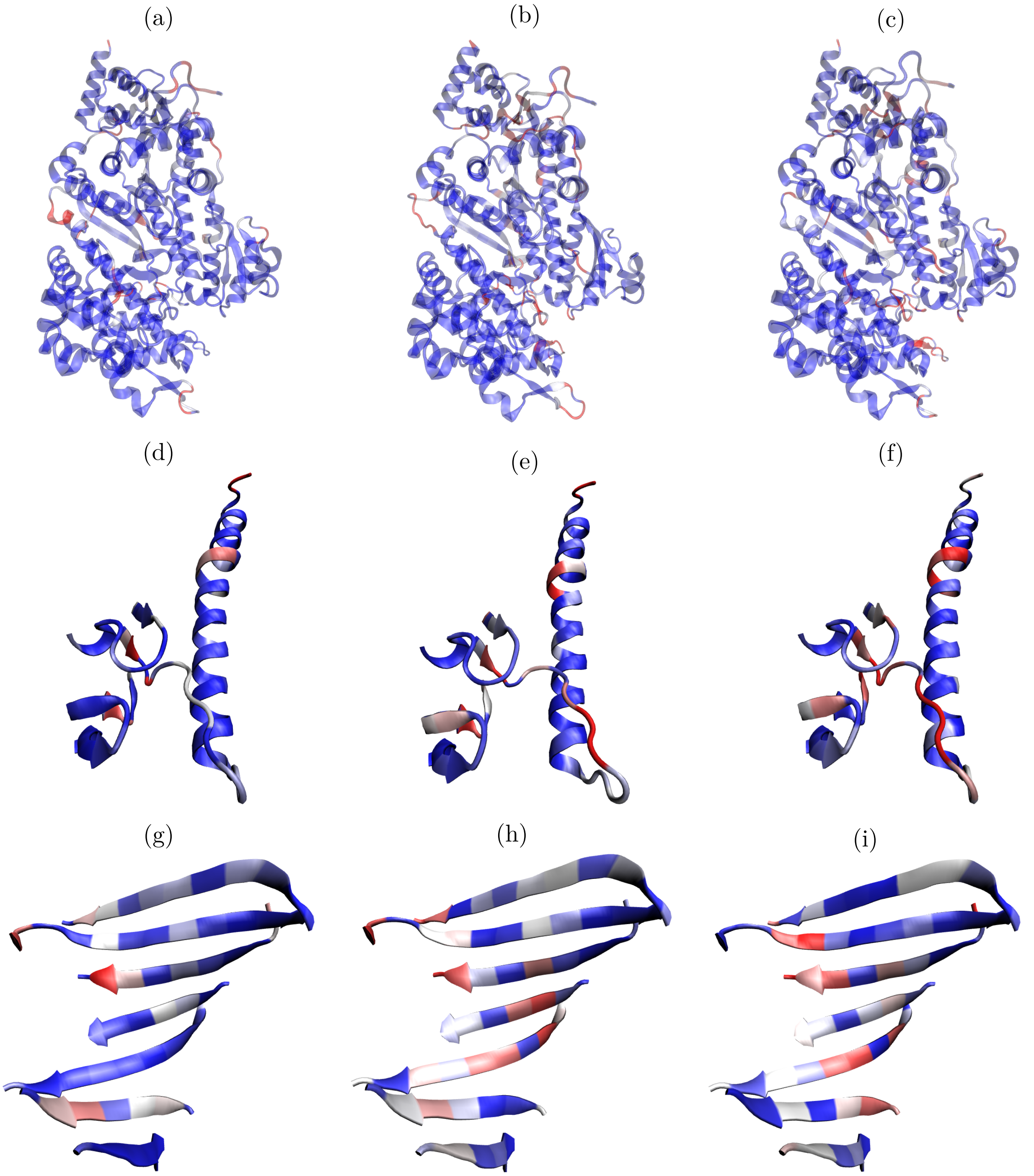


Figure 6.11: Representation of residues changing of backbone conformation between (a) PPS & P$_{\mathrm{i}}$R, (b) P$_{\mathrm{i}}$R & Strong-ADP and (c) PPS & Strong-ADP. The representation are made with the crystal structure of the first state declared. In this continuous representation, the most blue correspond to almost identical distributions, while the most red correspond to the most different distributions. (d), (f) & (e) Zoom on Nucleotide Binding Loops and Relay Helix. (g), (h) & (i) Zoom on Transducer Beta sheets.

The Table 6.4 lists all the residues, where the conformation changed between the structure PPS, P$_{\mathrm{i}}$R and Strong-ADP, i.e. the residues $k$ such as conformation similarity indicators is lower a threshold $\epsilon = 0.15$ (e.g. $S_k\left[\Delta T_f, PPS; \Delta T_f, P_iR\right] < \epsilon$ for the difference between PPS and P$_{\mathrm{i}}$R).

| States compared | Residues |
|---|---|
| PPS/P$_{\mathrm{i}}$R | 1, 3, 25, 26, 33, 34, 35, 36, 37, 38, 50, 51, 52, 54, 57, 118 119, 120, 175, 176, 177, 178, 179, 203, 305, 306, 357, 358, 359 442, 444, 446, 455, 457, 500, 501, 505, 506, 518, 551, 552, 620 621, 622, 623, 624, 626, 629, 631, 633 |
| P$_{\mathrm{i}}$R/SA | 1, 3, 25, 26, 33, 34, 35, 36, 37, 51, 52, 54, 119, 120, 145, 172 175, 176, 177, 178, 179, 203, 204, 228, 239, 240, 355, 356, 357, 359 394, 395, 396, 397, 398, 401, 403, 404, 406, 444, 455, 456, 461, 462 486, 487, 500, 501, 504, 505, 506, 508, 509, 535, 536, 598, 599, 620 621, 624, 626, 627, 628, 629, 630, 635, 636, 637, 641, 662, 703, 704 705, 706, 707, 748, 749, 750, 754, 756, 757, 759, 760, 761, 770, 771 |
| PPS/SA | 2, 3, 37, 38, 50, 51, 57, 67, 172, 173, 207, 239, 240, 305, 306, 307 355, 356, 358, 394, 395, 396, 397, 398, 401, 403, 404, 446, 447, 454 456, 457, 460, 461, 462, 486, 488, 509, 518, 535, 536, 551, 552, 598 599, 621, 622, 627, 628, 629, 630, 631, 632, 633, 635, 636, 637, 641 664, 680, 704, 705, 706, 707, 748, 749, 750, 755, 756, 757, 759, 760 761, 770, 771 |

Table 6.4: List of residues whose dihedral angle distribution changed such that their conformation similarity indicators is less than 0.15 between last ranges of snapshots taken along unconstrained simulations of PPS, P$_{\mathrm{i}}$R and Strong-ADP. Numbers are colored depending on their affiliation to specific domains of the protein. gray for N-terminal Beta-Barrel, dark gray for N-terminal, cyan for Loop 1, magenta for Switch I, pink for Loop 4, brown for HCM Loop, orange for Switch II, red for U50kDa, olive for HTH, yellow for Loop 2, purple for Strut, green for Converter, violet for Insert 2, teal for Transducer.

As shown in Fig. 6.11 and Table 6.4, the comparison of PPS and P$_{\mathrm{i}}$R shows less difference than between either of PPS and Strong-ADP, or P$_{\mathrm{i}}$R and Strong-ADP. This is expected as PPS and P$_{\mathrm{i}}$R are two close conformations of the protein. However, changes can be found with residues part of one $\beta$ strand from the transducer, which is just before Switch II, and on the Switch II loop itself. Indeed residues 455 and 457 in the comparison PPS/P$_{\mathrm{i}}$R, and residue 456 and 461 or 456 and 460 changed in the comparisons P$_{\mathrm{i}}$R/SA and PPS/SA respectively. This is probably related to change in Switch II conformation between PPS, P$_{\mathrm{i}}$R and Strong-ADP.

In the comparisons of PPS and P$_{\mathrm{i}}$R with Strong-ADP, several residues above 700 are considered different, which is not the case in the comparisons of PPS with P$_{\mathrm{i}}$R. As this part consists of the Converter, this observation is actually expected, as there is a change of orientation of the converter in Strong-ADP compared to PPS and P$_{\mathrm{i}}$R related (or leading) to the power-stroke.

In addition, in every comparisons, several residues part of the N-terminal beta-barrel (residues 1 to 55) are considered as having different distribution between compared ensemble of snapshots (see Table 6.4). However, we considered this part of the protein as flexible and thus we do not consider the difference in this part as relevant.

## 6.3 Conclusion

In this chapter we described two methodologies to describe the conformational changes of the myosin. These methodologies concern the domain relative positioning and the CC analysis of residues. The first methodology allows a coarse description of Domain motion, while the second methodology describes the local changes that could lead to the coarser changes. Thus, they describe the conformations at two opposite scales of the protein and are complementary. The results obtained here could be used in order to propose collective variables to drive the transformation of the system from PPS to Strong-ADP.

# Chapter 7

# Conclusion & perspectives

## 7.1 Conclusion

After presenting the molecular motor system composed of myosin and actin, we have focused our study on the motor domain of the myosin VI. The thesis try to give insights on the hypothesis consisting in the existence in the myosin-actin cycle, after myosin binds to actin, of an intermediate state that would allow the $\mathrm{P_i}$ departure from the active site before the powerstroke. This hypothesis is based on the existence of a so-called $\mathrm{P_i}$-Release state. This thesis is therefore based on the modeling (using molecular dynamics simulations presented in the Appendix A) and characterisation of three states of myosin VI: Pre-Powerstroke (PPS), $\mathrm{P_i}$-Release, and Strong-ADP.

In a first part, we have presented the $\mathrm{P_i}$-Release hypothesis in details together with some previous structural studies that have studied the departure of $\mathrm{P_i}$. Some challenges to this hypothesis have been also exposed.

We then have studied the departure of $\mathrm{P_i}$ from the active site starting from structures in Pre-Powerstroke and $\mathrm{P_i}$-Release states. We have given a detailed description of the interactions along the $\mathrm{P_i}$ way out from the active site, which allows us to propose key residues controlling its departure. We have also observed that the $\mathrm{P_i}$ left the active site using the backdoor and populated the position found in $\mathrm{P_i}$-Release crystals.

In a second part, we have developed several methodologies to characterize the dynamic behaviour of PPS, $\mathrm{P_iR}$ and Strong-ADP using both their crystallographic structures and unconstrained simulations in absence of actin.

The first methodology concerns the study of distances in the cleft and in the active site of myosin. This allowed us to get criteria to distinguish the $\mathrm{P_iR}$ state compared to PPS and Strong-ADP. Moreover some other distances allow to discriminate Strong-ADP from PPS.

The second methodology aims at describing the conformational changes of small domains of myosin, based on the positions of $\alpha$-helix considered as small rigid bodies. This led us to define some criteria to distinguish Strong-ADP from PPS. However, this methodology was not able to capture specific features of $P_iR$.

Finally, we presented two methodologies to describe the protein conformation at opposite scales. A coarse description of the conformation based on the position of the subdomains considered as rigid bodies, and a very fine description based on the Ramachandran angles of the amino acids. The positions of the rigid domains describes well the differences between PPS and Strong-ADP. The Ramachandran angles were not enough exploited to draw clear conclusions.

In conclusion the methodologies developed to characterize the PPS, $P_iR$ and Strong-ADP states allows us to propose some collective variables to define these states. The table 7.1 summarizes all the identified collective variables useful to discriminate PPS, $P_iR$ and Strong-ADP. These collective variable can then be used to drive the transformation from PPS to Strong-ADP.

| Parameter category | Parameter | Discriminated state |
|---|---|---|
| Distances cleft | 239-464, 598-602 | PPS and Strong-ADP |
| | 536-603, 199-461/197-461 | $P_iR$ |
| Switch II state | 199-461/205-461 | $P_iR$ |
| Helices | Distances 6-10, 5-12, 6-12, 7-12, 14-12, 5-13, 13-14 | PPS and Strong-ADP |
| | Distances 4-8, 5-8, 4-12 | $P_iR$ |
| | Dihedrals 14-10, 5-13, 11-14 | PPS and Strong-ADP |
| Rigid bodies | Distance Nter-Conv (L50-Nter + L50-Conv) | PPS and Strong-ADP |
| | Angle L50-Conv, U50-Conv, U50-Nter | PPS and Strong-ADP |

Table 7.1: Summarizing table of the collective variables identified to discriminate one state to the 2 others.

## 7.2 Perspectives

In order to test the impact of the hydration of the magnesium on the $P_i$ departure from the active site, it could be tried to sample the free energy landscape of the coordination of $Mg^{2+}$in parallel to the departure of $P_i$. A preliminary study of the impact of the coordination of $Mg^{2+}$was performed by forcing a third and fourth water molecules in the coordination of $Mg^{2+}$, described in appendix F.

Some complementary simulations could also be performed to study the impact of stabilizing the $P_iR$ state on the $P_i$ departure and test the consequence of actin binding on this state.

We could use the collective variable that we have identified in this thesis in order to drive the transformation from PPS to Strong-ADP. This could lead to test whether the $P_iR$ state is visited along the transformation and validate or not the $P_iR$ hypothesis. Preliminary simulations of umbrella sampling simulations using some identified collective variables were already tested and described in appendix D.

We could also take profit of these simulations to validate collective variables that we could incorporate in a model of collective motors (Caruel et al., 2019).

A limit for the interpretation of the present simulations is the absence of actin. In the perspective of this thesis, simulation including the presence of actin would give more reliable insight of what happen during the powerstroke.

# Bibliography

Alberts, B., Johnson, A., Lewis, J., Morgan, D., Raff, M., Roberts, K., and Walter, P. (2014). *Molecular Biology of the Cell, Sixth Edition.* Garland Science.

Anandakrishnan, R., Aguilar, B., and Onufriev, A. V. (2012). H++ 3.0: automating pK prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations. *Nucleic Acids Research*, 40(W1):W537–W541.

Andersen, H. C. (1983). Rattle: A "velocity" version of the shake algorithm for molecular dynamics calculations. *Journal of Computational Physics*, 52(1):24–34.

Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A., and Haak, J. R. (1984). Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics*, 81(8):3684–3690.

Berraud-Pache, R., Lindh, R., and Navizet, I. (2018). QM/MM Study of the Formation of the Dioxetanone Ring in Fireflies through a Superoxide Ion. *The Journal of Physical Chemistry B*, 122(20):5173–5182.

Blanc, F. (2018). *Exploring chemo-mechanical transduction in the myosin molecular motor through computer simulations.* PhD thesis, Strasbourg.

Blanc, F., Isabet, T., Benisty, H., Sweeney, H. L., Cecchini, M., and Houdusse, A. (2018). An intermediate along the recovery stroke of myosin VI revealed by x-ray crystallography and molecular dynamics. *Proceedings of the National Academy of Sciences*, 115(24):6213–6218.

Brooks, B. R., Brooks III, C. L., Mackerell Jr., A. D., Nilsson, L., Petrella, R. J., Roux, B., Won, Y., Archontis, G., Bartels, C., Boresch, S., Caflisch, A., Caves, L., Cui, Q., Dinner, A. R., Feig, M., Fischer, S., Gao, J., Hodoscek, M., Im, W., Kuczera, K., Lazaridis, T., Ma, J., Ovchinnikov, V., Paci, E., Pastor, R. W., Post, C. B., Pu, J. Z., Schaefer, M., Tidor, B., Venable, R. M., Woodcock, H. L., Wu, X., Yang, W., York, D. M., and Karplus, M. (2009). Charmm: The biomolecular simulation program. *Journal of Computational Chemistry*, 30(10):1545–1614.

Brünger, A., Brooks, C. L., and Karplus, M. (1984). Stochastic boundary conditions for molecular dynamics simulations of ST2 water. *Chemical Physics Letters*, 105(5):495–500.

Caremani, M., Melli, L., Dolfi, M., Lombardi, V., and Linari, M. (2013). The working stroke of the myosin II motor in muscle is not tightly coupled to release of orthophosphate from its active site. *The Journal of Physiology*, 591(20):5187–5205.

Caremani, M., Melli, L., Dolfi, M., Lombardi, V., and Linari, M. (2015). Force and number of myosin motors during muscle shortening and the coupling with the release of the ATP hydrolysis products: Chemo-mechanical coupling during muscle shortening. *The Journal of Physiology*, 593(15):3313–3332.

Caruel, M., Moireau, P., and Chapelle, D. (2019). Stochastic modeling of chemical–mechanical coupling in striated muscles. *Biomechanics and Modeling in Mechanobiology*, 18(3):563–587.

Case, D., Belfon, K., Ben-Shalom, I., Brozell, S., Cerutti, D., Cheatham, T., III, Cruzeiro, V., Darden, T., Duke, R., Giambasu, G., Gilson, M., Gohlke, H., Goetz, A., Harris, R., Izadi, S., Izmailov, S., Kasavajhala, K., Kovalenko, A., Krasny, R., Kurtzman, T., Lee, T., LeGrand, S., Li, P., Lin, C., Liu, J., Luchko, T., Luo, R., Man, V., Merz, K., Miao, Y., Mikhailovskii, O., Monard, G., Nguyen, H., Onufriev, A., F.Pan, Pantano, S., Qi, R., Roe, D., Roitberg, A., Sagui, C., Schott-Verdugo, S., Shen, J., Simmerling, C., N.R.Skrynnikov, Smith, J., Swails, J., Walker, R., Wang, J., Wilson, L., Wolf, R., Wu, X., Xiong, Y., Xue, Y., York, D., and Kollman, P. (2005). The amber biomolecular simulation programs. *Journal of Computational Chemistry*, 26(16):1668–1688.

Cecchini, M., Alexeev, Y., and Karplus, M. (2010). Pi release from myosin: A simulation analysis of possible pathways. *Structure*, 18(4):458 – 470.

Chen, V. B., Arendall, III, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., and Richardson, D. C. (2010). *MolProbity*: all-atom structure validation for macromolecular crystallography. *Acta Crystallographica Section D*, 66(1):12–21.

Chenoweth, K., Van Duin, A. C., and Goddard, W. A. (2008). Reaxff reactive force field for molecular dynamics simulations of hydrocarbon oxidation. *The Journal of Physical Chemistry A*, 112(5):1040–1053.

Chipot, C. and Pohorille, A. (2007). Free energy calculations. *Springer series in chemical physics*, 86:159–184.

Coureux, P.-D., Wells, A. L., Ménétrey, J., Yengo, C. M., Morris, C. A., Sweeney, H. L., and Houdusse, A. (2003). A structural state of the myosin V motor without bound nucleotide. *Nature*, 425(6956):419.

Debold, E. P. (2021). Recent insights into the relative timing of myosin's powerstroke and release of phosphate. *Cytoskeleton*, 78(9):448–458.

Elber, R. and Karplus, M. (1990). Enhanced sampling in molecular dynamics: use of the time-dependent hartree approximation for a simulation of carbon monoxide diffusion through myoglobin. *Journal of the American Chemical Society*, 112(25):9161–9175.

Fisher, A. J., Smith, C. A., Thoden, J., Smith, R., Sutoh, K., Holden, H. M., and Rayment, I. (1995). X-ray structures of the myosin motor domain of dictyostelium discoideum complexed with MgADP.BeFx and MgADP. AlF4. *Biochemistry*, 34(28):8960–8972.

Geeves, M. A. (2016). Review: The ATPase mechanism of myosin and actomyosin. *Biopolymers*, 105(8):483–491.

Geeves, M. A. and Holmes, K. C. (2005). The Molecular Mechanism of Muscle Contraction. 71:161–193.

Gergely, J., Warshaw, D. M., Walker, W. F., Davies, R. E., Curtin, N. A., Wood, B., Alexander, R. M., Alpert, N. R., Crompton, R. H., and Newsom-Davis, J. M. (2022). *Muscle*.

Gourinath, S., Himmel, D. M., Brown, J. H., Reshetnikova, L., Szent-Gyorgyi, A. G., and Cohen, C. (2003). Crystal structure of scallop myosin S1 in the pre-power stroke state to 2.6 Å resolution: Flexibility and function in the head. *Structure*, 11(12):1621 – 1627.

Gravett, M. S. C., Cocking, R. C., Curd, A. P., Harlen, O., Leng, J., Muench, S. P., Peckham, M., Read, D. J., Rogers, J. F., Welch, R. C., and Harris, S. A. (2022). Moving in the mesoscale: Understanding the mechanics of cytoskeletal molecular motors by combining mesoscale simulations with imaging. *WIREs Computational Molecular Science*, 12(3):e1570.

Grigorenko, B. L., Rogov, A. V., Topol, I. A., Burt, S. K., Martinez, H. M., and Nemukhin, A. V. (2007). Mechanism of the myosin catalyzed hydrolysis of ATP as rationalized by molecular modeling. *Proceedings of the National Academy of Sciences*, 104(17):7057–7061.

Grubmüller, H., Heller, H., Windemuth, A., and Schulten, K. (1991). Generalized verlet algorithm for efficient molecular dynamics simulations with long-range interactions. *Molecular Simulation*, 6(1-3):121–142.

Gupta, R. S. (1995). Evolution of the chaperonin families (HSP60, HSP 10 and TCP-1) of proteins and the origin of eukaryotic cells. *Molecular Microbiology*, 15(1):1–11.

Gurel, P. S., Kim, L. Y., Ruijgrok, P. V., Omabegho, T., Bryant, Z., and Alushin, G. M. (2017). Cryo-em structures reveal specialization at the myosin VI-actin interface and a mechanism of force sensitivity. *Elife*, 6.

Hamelberg, D., de Oliveira, C. A. F., and McCammon, J. A. (2007). Sampling of slow diffusive conformational transitions with accelerated molecular dynamics. *The Journal of Chemical Physics*, 127(15):155102.

Houdusse, A. and Sweeney, H. L. (2016). How myosin generates force on actin filaments. *Trends in Biochemical Sciences*, 41(12):989 – 997.

Humphrey, W., Dalke, A., and Schulten, K. (1996). VMD – Visual Molecular Dynamics. *Journal of Molecular Graphics*, 14:33–38.

Iannuzzi, M., Laio, A., and Parrinello, M. (2003). Efficient exploration of reactive potential energy surfaces using car-parrinello molecular dynamics. *Phys. Rev. Lett.*, 90:238302.

Ishii, N. (2017). Groel and the groel-groes complex. *Macromolecular Protein Complexes: Structure and Function*, pages 483–504.

Jarzynski, C. (1997). Nonequilibrium equality for free energy differences. *Physical Review Letters*, 78(14):2690.

Kaliman, I., Grigorenko, B., Shadrina, M., and Nemukhin, A. (2009). Opening the arg-glu salt bridge in myosin: computational study. *Phys. Chem. Chem. Phys.*, 11:4804–4807.

Karplus, P. A. (1996). Experimentally observed conformation-dependent geometry and hidden strain in proteins. *Protein Science*, 5(7):1406–1420.

Kashefolgheta, S. and Vila Verde, A. (2017). Developing force fields when experimental data is sparse: Amber/gaff-compatible parameters for inorganic and alkyl oxoanions. *Phys. Chem. Chem. Phys.*, 19:20593–20607.

Kiani, F. A. and Fischer, S. (2014). Catalytic strategy used by the myosin motor to hydrolyze ATP. *Proceedings of the National Academy of Sciences*, 111(29):E2947–E2956.

Krivov, S. V. (2013). On reaction coordinate optimality. *Journal of Chemical Theory and Computation*, 9(1):135–146.

Kumar, S., Rosenberg, J. M., Bouzida, D., Swendsen, R. H., and Kollman, P. A. (1992). The weighted histogram analysis method for free-energy calculations on biomolecules. i. the method. *Journal of Computational Chemistry*, 13(8):1011–1021.

Kunzmann, P., Anter, J. M., and Hamacher, K. (2022). Adding hydrogen atoms to molecular models via fragment superimposition. *Algorithms for Molecular Biology*, 17(1):7.

Kästner, J. (2011). Umbrella sampling. *WIREs Computational Molecular Science*, 1(6):932–942.

Lawson, J. D., Pate, E., Rayment, I., and Yount, R. G. (2004). Molecular dynamics analysis of structural factors influencing back door Pi release in myosin. *Biophysical Journal*, 86(6):3794–3803.

Llinas, P., Isabet, T., Song, L., Ropars, V., Zong, B., Benisty, H., Sirigu, S., Morris, C., Kikuti, C., Safer, D., Sweeney, H., and Houdusse, A. (2015). How actin initiates the motor activity of myosin. *Developmental Cell*, 33(4):401 – 412.

Lodish, H., Berk, A., Kaiser, C. A., Kaiser, C., Krieger, M., Scott, M. P., Bretscher, A., Ploegh, H., Matsudaira, P., et al. (2008). *Molecular cell biology*.

Lymn, R. and Taylor, E. W. (1971). Mechanism of adenosine triphosphate hydrolysis by actomyosin. *Biochemistry*, 10(25):4617–4624.

Mahmood, R. and Yount, R. G. (1984). Photochemical probes of the active site of myosin. irradiation of trapped 3'-o-(4-benzoyl)benzoyladenosine 5'-triphosphate labels the 50-kilodalton heavy chain tryptic peptide. *Journal of Biological Chemistry*, 259(21):12956–12959.

Maier, J. A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K. E., and Simmerling, C. (2015). ff14sb: Improving the accuracy of protein side chain and backbone parameters from ff99sb. *Journal of Chemical Theory and Computation*, 11(8):3696–3713.

Miyamoto, S. and Kollman, P. A. (1992). Settle: An analytical version of the shake and rattle algorithm for rigid water models. *Journal of Computational Chemistry*, 13(8):952–962.

Mornet, D., Bertrand, R., Pantel, P., Audemard, E., and Kassab, R. (1981). Proteolytic approach to structure and function of actin recognition site in myosin heads. *Biochemistry*, 20(8):2110–2120.

Mugnai, M. L. and Thirumalai, D. (2021). Step-wise hydration of magnesium by four water molecules precedes phosphate release in a myosin motor. *The Journal of Physical Chemistry B*, 125(4):1107–1117.

Ménétrey, J., Llinas, P., Mukherjea, M., Sweeney, H. L., and Houdusse, A. (2007). The structural basis for the large powerstroke of myosin VI. *Cell*, 131(2):300 – 308.

Okazaki, K.-i. and Hummer, G. (2013). Phosphate release coupled to rotary motion of f1-atpase. *Proceedings of the National Academy of Sciences*, 110(41):16468–16473.

Oliver, R. C., Read, D. J., Harlen, O. G., and Harris, S. A. (2013). A stochastic finite element model for the dynamics of globular macromolecules. *Journal of Computational Physics*, 239:147–165.

Onishi, H., ichiro Kojima, S., Katoh, K., Fujiwara, K., Martinez, H. M., and Morales, M. F. (1998). Functional transitions in myosin: Formation of a critical salt-bridge and transmission of effect to the sensitive tryptophan. *Proceedings of the National Academy of Sciences*, 95(12):6653–6658.

Ovchinnikov, V. and Karplus, M. (2012). Analysis and elimination of a bias in targeted molecular dynamics simulations of conformational transitions: Application to calmodulin. *The Journal of Physical Chemistry B*, 116(29):8584–8603.

Ozer, G., Valeev, E. F., Quirk, S., and Hernandez, R. (2010). Adaptive steered molecular dynamics of the long-distance unfolding of neuropeptide y. *Journal of Chemical Theory and Computation*, 6(10):3026–3038.

Park, H., Li, A., Chen, L.-Q., Houdusse, A., Selvin, P. R., and Sweeney, H. L. (2007). The unique insert at the end of the myosin VI motor is the sole determinant of directionality. *Proceedings of the National Academy of Sciences*, 104(3):778–783.

Park, S., Khalili-Araghi, F., Tajkhorshid, E., and Schulten, K. (2003). Free energy calculation from steered molecular dynamics simulations using jarzynski's equality. *The Journal of Chemical Physics*, 119(6):3559–3566.

Poole, K. and Elber, R. (2019). Atomically detailed simulation of the powerstroke in myosin II by milestoning. *Biophysical Journal*, 116(3):259a.

Powell, K. J., Brown, P. L., Byrne, R. H., Gajda, T., Hefter, G., Sjöberg, S., and Wanner, H. (2005). Chemical speciation of environmentally significant heavy metals with inorganic ligands. part 1: The Hg2+– Cl–, OH–, CO32–, SO42–, and PO43– aqueous systems (IUPAC technical report). *Pure and Applied Chemistry*, 77(4):739–800.

Preller, M. and Holmes, K. C. (2013). The myosin start-of-power stroke state and how actin binding drives the power stroke. *Cytoskeleton*, 70(10):651–660.

Ramachandran, G., Ramakrishnan, C., and Sasisekharan, V. (1963). Stereochemistry of polypeptide chain configurations. *Journal of Molecular Biology*, 7:95–99.

Rayment, I., Rypniewski, W., Schmidt-Base, K., Smith, R., Tomchick, D., Benning, M., Winkelmann, D., Wesenberg, G., and Holden, H. (1993). Three-dimensional structure of myosin subfragment-1: a molecular motor. *Science*, 261(5117):50–58.

Robert-Paganin, J., Pylypenko, O., Kikuti, C., Sweeney, H. L., and Houdusse, A. (2019). Force generation by myosin motors: A structural perspective. *Chemical Reviews*, 120(1):5–35.

Ruppel, K. M. and Spudich, J. A. (1996). Structure-function studies of the myosin motor domain: importance of the 50-kDa cleft. *Molecular Biology of the Cell*, 7(7):1123–1136.

Ryckaert, J.-P., Ciccotti, G., and Berendsen, H. J. (1977). Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of Computational Physics*, 23(3):327–341.

Schwarzl, S. M., Smith, J. C., and Fischer, S. (2006). Insights into the chemomechanical coupling of the myosin motor from simulation of its ATP hydrolysis mechanism. *Biochemistry*, 45(18):5830–5847.

Scott, B., Marang, C., Woodward, M., and Debold, E. P. (2021). Myosin's powerstroke occurs prior to the release of phosphate from the active site. *Cytoskeleton*, 78(5):185–198.

Scott, W. R. P., Hünenberger, P. H., Tironi, I. G., Mark, A. E., Billeter, S. R., Fennen, J., Torda, A. E., Huber, T., Krüger, P., and van Gunsteren, W. F. (1999). The gromos biomolecular simulation program package. *The Journal of Physical Chemistry A*, 103(19):3596–3607.

Senftle, T. P., Hong, S., Islam, M. M., Kylasa, S. B., Zheng, Y., Shin, Y. K., Junkermeier, C., Engel-Herbert, R., Janik, M. J., Aktulga, H. M., et al. (2016). The reaxff reactive force-field: development, applications and future directions. *npj Computational Materials*, 2(1):1–14.

Shen, M.-y. and Sali, A. (2006). Statistical potential for assessment and prediction of protein structures. *Protein Science*, 15(11):2507–2524.

Shin, Y. K., Ashraf, C. M., and van Duin, A. C. (2021). Development and applications of the reaxff reactive force field for biological systems. In *Computational Materials, Chemistry, and Biochemistry: From Bold Initiatives to the Last Mile*, pages 157–182.

Smith, C. A. and Rayment, I. (1996). X-ray structure of the magnesium(II) · adp · vanadate complex of the dictyostelium discoideum myosin motor domain to 1.9 Å resolution,. *Biochemistry*, 35(17):5404–5417.

Solernou, A., Hanson, B. S., Richardson, R. A., Welch, R., Read, D. J., Harlen, O. G., and Harris, S. A. (2018). Fluctuating finite element analysis (ffea): A continuum mechanics software tool for mesoscale simulation of biomolecules. *PLOS Computational Biology*, 14(3):1–29.

Song, Y., DiMaio, F., Wang, R.-R., Kim, D., Miles, C., Brunette, T., Thompson, J., and Baker, D. (2013). High-resolution comparative modeling with rosettaCM. *Structure*, 21(10):1735–1742.

Sugita, Y. and Okamoto, Y. (1999). Replica-exchange molecular dynamics method for protein folding. *Chemical Physics Letters*, 314(1):141–151.

Sun, R., Sode, O., Dama, J. F., and Voth, G. A. (2017). Simulating Protein Mediated Hydrolysis of ATP and Other Nucleoside Triphosphates by Combining QM/MM Molecular Dynamics with Advances in Metadynamics. *Journal of Chemical Theory and Computation*, 13(5):2332–2341.

Sweeney, H. L. and Houdusse, A. (2010). Structural and Functional Insights into the Myosin Motor Mechanism. *Annual Review of Biophysics*, 39(1):539–557.

Torrie, G. and Valleau, J. (1977). Nonphysical sampling distributions in monte carlo free-energy estimation: Umbrella sampling. *Journal of Computational Physics*, 23(2):187 – 199.

Tuckerman, M. (2010). *Statistical mechanics: theory and molecular simulation.* Oxford university press.

Várkuti, B. H., Yang, Z., and Malnasi-Csizmadia, A. (2015). Structural model of weak binding actomyosin in the prepowerstroke state. *Journal of Biological Chemistry*, 290(3):1679–1688.

Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A., and Case, D. A. (2004). Development and testing of a general amber force field. *Journal of Computational Chemistry*, 25(9):1157–1174.

Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F. T., de Beer, T. A., Rempfer, C., Bordoli, L., Lepore, R., Schwede, and Torsten (2018). SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Research*, 46(W1):W296–W303.

Webb, B. and Sali, A. (2016). Comparative protein structure modeling using modeller. *Current Protocols in Bioinformatics*, 54(1):5.6.1–5.6.37.

Webb, B. and Sali, A. (2017). *Protein Structure Modeling with MODELLER*, pages 39–54. Springer New York, New York, NY.

Wells, A. L., Lin, A. W., Chen, L.-Q., Safer, D., Cain, S. M., Hasson, T., Carragher, B. O., Milligan, R. A., and Sweeney, H. L. (1999). Myosin VI is an actin-based motor that moves backwards. *Nature*, 401(6752):505–508.

Woledge, R. C., Curtin, N. A., and Homsher, E. (1985). Energetic aspects of muscle contraction. 41:1–357.

Wriggers, W. and Schulten, K. (1999). Investigating a back door mechanism of actin phosphate release by steered molecular dynamics. *Proteins: Structure, Function, and Bioinformatics*, 35(2):262–273.

Yount, R. G., Lawson, D., and Rayment, I. (1995). Is myosin a ”back door” enzyme? *Biophysical journal*, 68(4 Suppl):44S–49S.

Zhou, Y., Ojeda-May, P., Nagaraju, M., Kim, B., and Pu, J. (2018). Mapping Free Energy Pathways for ATP Hydrolysis in the E. coli ABC Transporter HlyB by the String Method. *Molecules*, 23(10):2652.

# Appendix A

# Molecular Dynamics simulations

This thesis is mostly based on molecular dynamics (MD) simulations. We first introduce rapidly the principle of the MD. Then we describe how to perform a MD : how to prepare the molecular model and what are the choices for the numerical algorithm. We finally discuss about methods that are used in order to accelerate sampling of conformations and compute free energy associated with transformations.

## Contents

## A.1 Principle of Molecular Dynamics

Molecular Dynamics is based on the assumption that a molecular system can be described by the position of the nucleus of its constituting atoms with classical mechanics. The computational technique consists in solving the Newton equation for the N nucleus i of mass $m_i$ and coordinates $x_i$ submitted to potential function (or force field) $U$ :

$$m_i \frac{d^2 x_i}{dt^2} = -\nabla_{x_i} U \tag{A.1}$$

In the process of molecular dynamics simulations, a certain number of steps are necessary to be carried out.

The first ensemble of steps consist in constructing the molecular model for the system of interest. This molecular model can be decomposed in two parts.

**(i) Structural information :** is the coordinates and the nature of every atoms composing the system.

**(ii) Numerical information :** where we can find the force fields, which are functions of potentials, to which are subjected the atoms of the system, and their associated parameters.

Then, molecular dynamics relies on the integration of the classical equations of motions. The integration scheme to use might depend on the choice of the thermostat and the barostat. However, using most popular all-in-hands molecular dynamics package do not necessitate to choose which integration scheme to use as this choice is usually hidden under the choice of the parameters.

## A.2 Molecular model

### A.2.1 Source of proteins' molecular structures

The molecular structure may come from different sources. One of the most natural way to get an almost complete molecular structure of a protein is to use a molecular structure solved with X-ray crystallography. These structures are called crystallographic structures. These models have a high resolution and thus are convenient to use.

Another type of models comes from cryo-Elecron-Micrography (cryo-EM). These models were until quite recently at very low resolution and would only allow domain conformation visualization. However, the resolution recently rapidly increased. Indeed, new models determined from cryo-EM now reach a resolution comparable or even better than their crystallographic counterpart. This increase in resolution arises partially from an important work on cryo-EM machines, but even more on progress in data processing, reconstruction and refinement softwares.

One great advantage of the cryo-EM method is that it allows to solve molecular structure of a greater number of systems of proteins, as it does not necessitate crystallization. For example, it is very difficult to crystallize filament of actin. Thus models containing filament of actin are still currently only obtained with Cryo-EM.

In some cases, no molecular structure of the protein of interest is available. In these cases we have to build a new model with all the available information. Several solutions exist, such as *ab initio* reconstruction, template-based modelling, and recently developed machine learning approaches. Machine learning approaches recently became a very powerful tool, which is perceptible as regard to the obtained results during the Critical Assessment of protein Structure Prediction (CASP) competition. This competition on the problem of protein structure prediction is recently always won by machine learning teams as AlphaFold.

These steps can be carried out with the use of multiple tools such as SwissMODEL (Waterhouse et al., 2018), MODELLER (Webb and Sali, 2016), ROSETTA (Song et al., 2013). These tools are inherently different by their usage and their functioning, however they mostly rely on homology modelling. The homology modelling consists in comparing the sequence to find similarities. Based on similarity, the programs determine homolog proteins (sharing large sequence similarity) which can be used as reference in order to predict the form of the wanted protein.

### A.2.2 Completion of the structure

Most often, the molecular structures obtained from X-ray crystallography or cryo-EM do not contain every atoms because some positions could not be determined. This is due to an high flexibility of the corresponding elements causing fast movement of the atoms. However, it is important to have a complete structure in order to simulate the system correctly. Several methods exist to complete molecular structures.

The most efficient way is to use homology or comparative modelling. This can be performed with the use of MODELLER (Webb and Sali, 2017) or other homology modelling servers. In this thesis, we have used MODELLER.

MODELLER first randomly set positions of missing atoms and then use functions carrying weights depending on the correspondence between loop conformation and an optimal conformation. This generation of loop is repeated a certain number of times. In the frame of this thesis, ten models were generated. The user should then choose between the generated models with the help of scores such as DOPE (Discrete Optimized Protein Energy) (Shen and Sali, 2006). The user must be careful with the generated models as MODELLER can produce knots when attempting to reconstruct a loop (Blanc, 2018).

It is then important to check the chosen structures. In addition to the DOPE scores, it is a good practice to submit the structures to some server like Molprobity (Chen et al., 2010), that allows to check several parameters (like conformations and bad contacts among others), that can help to decide which model is the best. Then the model is ready for the next preparation step.

### A.2.3 Protonation state determination

Structures obtained from neutron crystallography contain hydrogen atoms. However, structures solved with neutron crystallography are rare. The molecular structures obtained with X-ray crystallography and cryo-EM lack information on the position of the hydrogen atoms. Therefore, to complete the model, the user have to add the hydrogen atoms.

For acido/basic groups, the protonation of the group depends on the local pH. This pH depends on the conditions of the environment of the protein. Particularly, the conditions impacting the protonation state are the global pH, the salinity of the solution and the local environment, i.e. which other atoms are in the vicinity of the acido-basic group. Protonation must be checked carefully. In particular the Histidine residues have a pKa close to neutral pH. This induces that their protonation state depends importantly on their environment. The most probable states are predicted based on the standard continuum solvent methodology, within the framework of the Poisson Boltzmann (PB) model. In this thesis we have used the H++ web server (Anandakrishnan et al., 2012).

New methods for the protonation of proteins are recently developed, such as (Kunzmann et al., 2022).

### A.2.4 Potential energy functional

The force field is a combination of a potential energy in a functional form and the associated parameters. There exist multiple kinds of force fields. Some force fields are full atom, but there also exist coarse grained force fields. Most popular force fields are AMBER (Maier et al., 2015), GROMOS (Scott et al., 1999) and CHARMM (Brooks et al., 2009). These force fields are pretty well designed for the proteins and nucleic acids. In addition they usually also have generic parameter values for the other types of molecules and ions that we can find in biomolecular systems. However, some particular molecules necessitate more specialised parameter values in order to simulate correctly the behaviour. In our system, it was particularly difficult to find appropriate force field parameters for the $\mathrm{P_i}$, which is not found satisfactory and might be a limitation of the study.

While some terms in the potential energy functional are well defined theoretically, force fields are also designed for the simplicity of integration and time consumption. Then the values of the parameters are determined both empirically and with the use of *ab initio* computations.

In the frame of this thesis we have used the force field AMBER ff14SB (Case et al., 2005), which functional form is described in eq.A.2.

$$
\begin{aligned}
V_{AMBER} = & \sum_{bonds} k_B \left(X_i - X_{eq}\right)^2 + \sum_{angles} k_A \left(\theta_i - \theta_{eq}\right)^2 + \sum_{dihedrals} \sum_{n} \frac{V_n}{2} \left[1 + cos\left(n\phi - \gamma\right)\right] \\
& + \sum_{i<j} \left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}} + \frac{q_i q_j}{\epsilon R_{ij}}\right]
\end{aligned}
\tag{A.2}
$$

The potential energy functional is composed of two parts. One part consists in bonded terms (in blue in eq.(A.2)), while the other parts concerns non bonded interactions (in red in eq.(A.2)).

Bonded interactions are composed of three terms. A first term represents the bond stretching, concerning each pair of bound atoms. A second term represents the angle stretching, concerning each group of three consecutively bound atoms. Both the first and the second term take the form of an harmonic potential. The third term represents the dihedral angles, concerning each group of four consecutively bound atoms. This term is also similar to one used to ensure the planar structure of double bonds, which are usually called improper torsions.

The non-bonded interactions are composed of a Lennard-Jones potential taking into account notably the van der Waals interactions and a Coulomb's Law potential taking into account electrostatics components. The non-bonded interactions are actually calculated between every pair of atoms that are separated by at least 3 bonds or not part of the same molecule. In the AMBER force field a special treatment is performed with non-bonded interactions inside the same molecule between atoms separated by three consecutive bonds. These are the "1-4 interactions", and are composed of the "1-4 vdW" and "1-4 electrostatics", which are divided by a factor of 2 and 1.2, respectively.

## A.3 Numerical conditions

### A.3.1 Temperature and pressure control

Multiple ways to control the temperature in molecular dynamics have been developed along its history. It ranges from simple thermostat consisting in rescaling all the velocities to correspond to the actual wanted temperature, up to more sophisticated thermostat based for example on Langevin Dynamics.

The Langevin Dynamics relies on an equation of motion of the form :

$$
m_i \frac{d^2 x_i}{dt^2} = -\nabla_{x_i} U - \gamma_i \frac{dx}{dt} + L(t) \tag{A.3}
$$

Where $m_i$ is the mass of the atom i, $x_i$ is one of its coordinate, $\gamma_i$ is a friction coefficient applied to i and L(t) is a Langevin random force.

In the preceding equation L(t) must satisfy some conditions.

Notably, it must be a white noise, which average is equal to 0 as follows :

$$\langle L(t) \rangle = 0 \tag{A.4}$$

And the variance must satisfy the condition of the fluctuation-dissipation theorem in order to have a canonical ensemble :

$$\langle L(t)L(t') \rangle = 2m_i\gamma_i k_B T\delta(t-t') \tag{A.5}$$

In the presented work of this thesis the temperature was controlled with the Langevin thermostat. (Tuckerman, 2010)

**Barostat** In this thesis we have used the Berendsen barostat to control the pressure. The Berendsen algorithm regulate the pressure by changing the volume by an increment proportional to the difference between the internal pressure and pressure in a weakly coupled bath (Berendsen et al., 1984).

### A.3.2 Integration scheme

The integration scheme used classically in molecular dynamics codes is the Verlet Algorithm (Grubmüller et al., 1991). This algorithm consists in separating the calculation for the next time step in two substeps. A first step consists in computing the forces to apply on each atom of the system based on their positions in the actual time step. The second step consists in updating the position of every atoms considering the calculated forces.

The Verlet algorithm is based on the Taylor expansion of the position.

$$x(t+\Delta t) = x(t) + \Delta t\dot{x}(t) + \frac{1}{2}\Delta t^2\ddot{x}(t) + O(\Delta t^3) \tag{A.6}$$

$$x(t-\Delta t) = x(t) - \Delta t\dot{x}(t) + \frac{1}{2}\Delta t^2\ddot{x}(t) + O(-\Delta t^3) \tag{A.7}$$

Combining eq.A.6 and A.7 leads to the expression of the new coordinates in which the velocities do not appear :

$$x(t+\Delta t) = 2x(t) - x(t-\Delta t) + \Delta t^2\ddot{x}(t) + O(\Delta t^4) \tag{A.8}$$

This scheme must however be adapted in some cases. For example with the use of a Langevin Thermostat, which modifies the deterministic dynamics into a Langevin dynamics. In this case, the integration is performed using the BBK integrator (Brünger et al., 1984), as follows :

$$x(t+\Delta t) = \frac{1}{1+\frac{\gamma\Delta t}{2m}} \cdot \left[2x(t) - x(t-\Delta t)\left[1-\frac{\gamma\Delta t}{2m}\right] - \frac{\partial U}{\partial x}\frac{\Delta t^2}{m} + L(t)\frac{\Delta t^2}{m}\right] \tag{A.9}$$

The extension of Verlet Algorithm has a large part of the barostat and thermostat can be found in the comprehensive textbook of (Tuckerman, 2010).

### A.3.3 Time Step

The choice of the time step size depends on the fastest degree of freedom of the system. In biomolecular system, the fastest degree of freedom are the bonds stretching, particularly those involving hydrogen atoms. The frequency of these bonds at 300K would constrain to use a maximum time step of 1 fs. However, constraining these bonds with the use of algorithms such as RATTLE (Andersen, 1983), SETTLE (Miyamoto and Kollman, 1992) or SHAKE (Ryckaert et al., 1977) allows to increase the time step to 2 fs.

## A.4 Accelerated Sampling

The number of degrees of freedom of usual biomolecular systems is pretty important. This limits the exploration of the phase space in the reachable timescales. Multiple approaches have thus been developed to accelerate the exploration of the phase space.

The number of degree of freedom can be reduced to accelerate the simulations with in return the limitation of some of the details. For example, faster and larger exploration of the phase space can be achieved by making an aggregation of several atoms into one bead. Other approaches completely change the formulation of the problem as for example the fluctuating finite element method (Oliver et al., 2013) (Solernou et al., 2018) (Gravett et al., 2022).

Some methods are focused on the exploration of the phase space, which can be wanted in order to get a sampling of possible conformations. Such techniques are called conformational sampling, and are specifically designed to avoid being stuck in a local minimum. They might be based on an energetic boost on the system, exchange between parallel simulations. The conformational sampling methods however might perturb the dynamics which is no longer realistic.

Another class of molecular dynamics techniques focus on determining the free energy landscape of the protein. This free energy landscape might be estimated around a particular point or for a specific transformation. In the case of the estimation of the free energy landscape for a transition, we use a specific control parameter called collective variable. The free energy landscape along a transformation depending on this collective variable is often called Potential of Mean Force (PMF). Determining the PMF is a powerful way to study how transitions occurs during the functioning of the protein.

### A.4.1 Free Energy methods

The free energy sampling methods are based on statistical physics to evaluate the free energy (Tuckerman, 2010). For a useful compendium of free energy calculation methods for other thermodynamic ensembles, I suggest the textbook of (Chipot and Pohorille, 2007).

A common point inside this class is that they use a constraint on some "reaction coordinate" (or collective variable). A reaction coordinate is a function of parameters from the system. This function of parameters is assumed to drive the transformation. Actually, the choice of the reaction coordinate means that we assume that the free energy barrier along this reaction coordinate is the most important. Alternatively, the reaction coordinate is the longest relaxing parameter, while all other parameters of the system relax and reach their equilibrium value faster. The choice of the reaction coordinate is crucial to get an optimal estimation, yet the choice can be difficult (Krivov, 2013).

**Umbrella Sampling** Umbrella sampling (Kästner, 2011) is a method used in classical molecular dynamics to precisely sample the energy surface of a protein as a function of a reaction coordinate. The principle is to separate the reaction path into several windows that we will explore separately during molecular dynamics. To do this, a potential is introduced for each window in order to force the reaction coordinate to remain close to the value of the imposed reaction coordinate. This potential is most often represented by a fully parameterizable quadratic potential (see Figure 8.2). The advantage of this method is to be able to sample a large number of possible conformations, some of which are not very or not accessible during molecular dynamics. The sampled values are then retrieved and, assuming overlap between multiple windows, a Weighted Histogram Analysis Method (WHAM) is used to plot the free energy potential (Kumar et al., 1992). WHAM consist in unbiasing the distribution and estimating the complete unbiased distribution.

The umbrella sampling method is often used to compare the stability of two different conformations, both when rotating through an angle, or when moving a loop.

**Others methods** Some others methods can be used to force the system to explore part of the phase space and estimate the free energy landscape. In the following, we list these methods and give their main characteristics.

- **Steering MD**. A force (represented by a spring with finite velocity) is apply on a ligand or ensemble of molecule to force it to move. This method is particularly useful to treat docking of a ligand to a substrate or to move something from a site and get the free energy landscape. However, the simulation is performed out of equilibrium so the Jarzinsky equality (Jarzynski, 1997) must be used to recover the PMF (Park et al., 2003).

- **Adaptive Steering MD**. This method is similar to Steering MD, but the reaction pathway is fragmented into sections ran in serial. Each sections starts from the final structure of the simulation that exhibits the closest free energy landscape estimation compared to the average of the pool of parallel simulations simulations (Ozer et al., 2010). It should help get more consistent results compared to classical Steering MD, as convergence might be obtained in each sections, but there might be limitations in the approach (might suffer from theoretical weaknesses or some specific framework of pertinence).

- **Targeted Molecular Dynamics (TMD)**. The targeted molecular dynamics aims at imposing a bias on a collective variable in order to produce a conformational transition path in a very short time. TMD is a kind a steering molecular dynamics in which the collective variable is generally the RMSD of (or part of) the system compared to a reference structure. This type of biased simulations tends to force large scale changes first while small-scale changes are slower (Ovchinnikov and Karplus, 2012). In order to remove this bias, (Ovchinnikov and Karplus, 2012) developed a methodology. Yet, even with this "correction", TMD still shows limitations in estimating a correct path.
- **Milestoning**. The reaction pathway is divided into windows separated by milestones. The system is simulated in each window a single time in order to get a canonical distribution or multiple short free simulations for each milestone until it reaches the next milestone. This method can be used to get free energy landscapes and mean first passage time. Milestoning is used on myosin II in the Thesis of Katelyn Elizabeth Poole (Poole and Elber, 2019) to drive the powerstroke.
- **Metadynamics**. This consists in filling up the potential energy wells by adding a bias in the free energy of the system to prevent it from being blocked in a region of space. This method might be quite costly computationally if the energy landscape is full of local minimums. This method was introduced in its modern form by (Iannuzzi et al., 2003).
- **Adaptive Biasing Force (ABF)**. This method is based on the exploration of the free energy landscape. It starts with a first estimation of the free energy landscape. Then it uses an adaptive biasing force that is the opposite of the average force experienced by the collective variable. This locally flattens the free energy landscape and allow diffusive dynamics at convergence. ABF is notabily used on myosin VI in (Blanc et al., 2018) to sample the free energy surface around Post-Rigor, Pre-Powerstroke and Pre-Transition State, the new structure the authors present.

### A.4.2 Conformational sampling

While the methods presented in the previous section are well designed to get a good approximation of the free energy landscape, some methodologies were particularly developed to explore a wider range of possible conformation in a limited amount of time without being stuck in a local potential energy minima. These are the conformational sampling techniques.

- **Locally Enhanced Sampling (LES)** This method is based on the simulation of several copies of *only* the fraction of interest from the molecule. This means that the less interesting parts of the system exist in only one examplary in the simulation, thus saving computing time. It can be particularly useful if what is of interest during a transformation is located in a small portion of a molecule. This method was first used by (Elber and Karplus, 1990) on myoglobin.

- **Accelerated Molecular Dynamics (aMD)** In this method we add a bias in the torsion term of the potential energy to lower the well while the free energy of the system is below a given limit. It can be a particularly powerful method while not having any clue neither on the reaction coordinate nor on the path it takes. This method was developed by (Hamelberg et al., 2007).

- **Replica Exchange Molecular Dynamics (REMD)** Run multiple parallel simulations with different initial conditions such as temperature. The trajectories are regularly exchanged between couples of replicas. It is powerful to get the free energy landscape while the reaction coordinate is not known although it is quite computationally expensive. This method was introduced by (Sugita and Okamoto, 1999).

# Appendix B

# Supplementary informations on $\mathrm{P_i}$ departure

## Contents

## B.1 Comparison of crystal states of PPS (2V26) and $\mathrm{P_iR}$ (4PJM)

The Figure B.1 gives the superposition of the crystal structures of PPS and $\mathrm{P_iR}$.

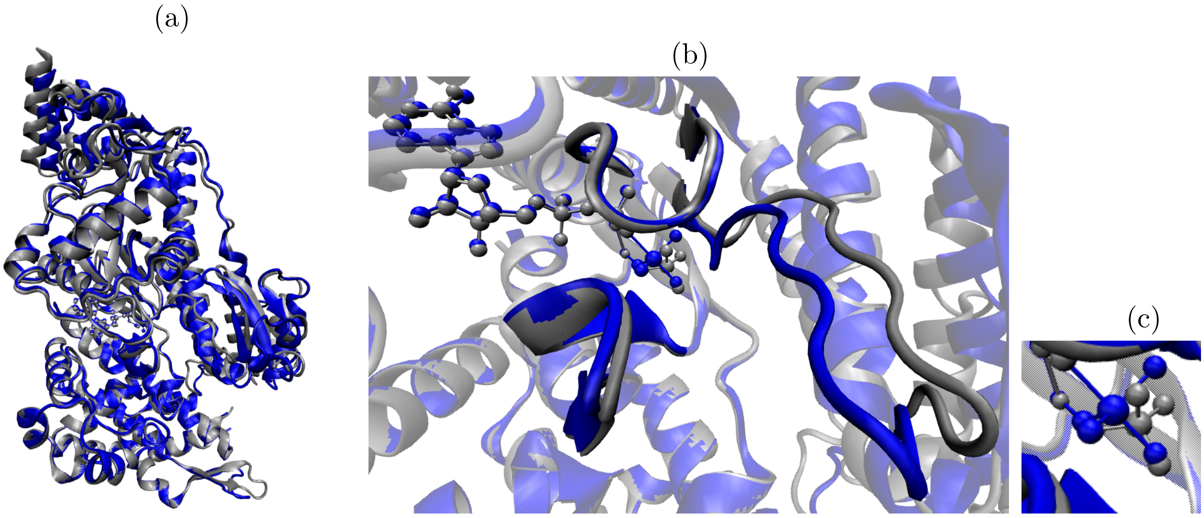


Figure B.1: Superposition of 2V26 (PPS, blue) and 4PJM ($\mathrm{P_iR}$, gray) crystallographic states superimposed on U50. (a) Whole protein view, (b) close up view on the active site with the nucleid binding loop (NBL) and the ligands in solid, while the rest of the protein structures is transparent. (c) Close up view on Pi.

## B.2 Relaxation of PPS structures

The Fig. B.2 and B.3 shows the superposition of the relaxed structures PPS-D, PPS-E and PPS-F which have the configuration (1,2).

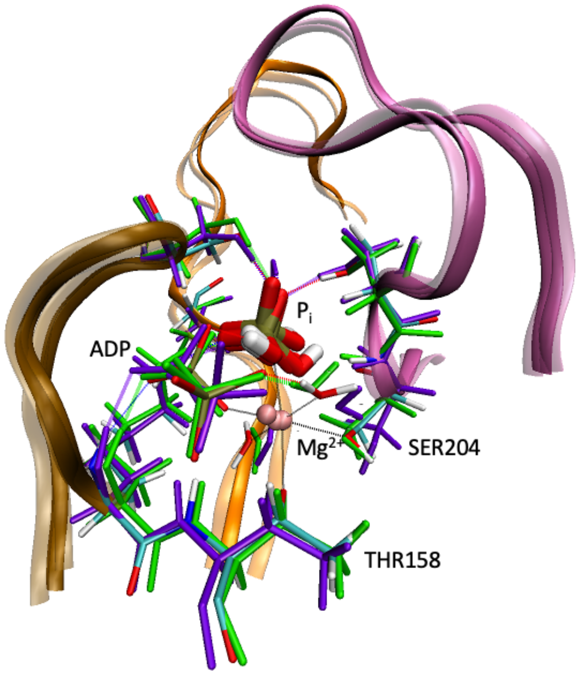


Figure B.2: Superposition of the nucleotide binding loop of relaxed structures PPS-D (green), PPS-E (purple) and PPS-F (all atoms). The H-bonds are shows in the color of the molecule. The interactions with $Mg^{2+}$ are shown in black dashed lines for PPS-F. In the PPS-E, SER204 is not longer in interaction with $Mg^{2+}$. The ADP Phosphate is in interaction with $Mg^{2+}$.

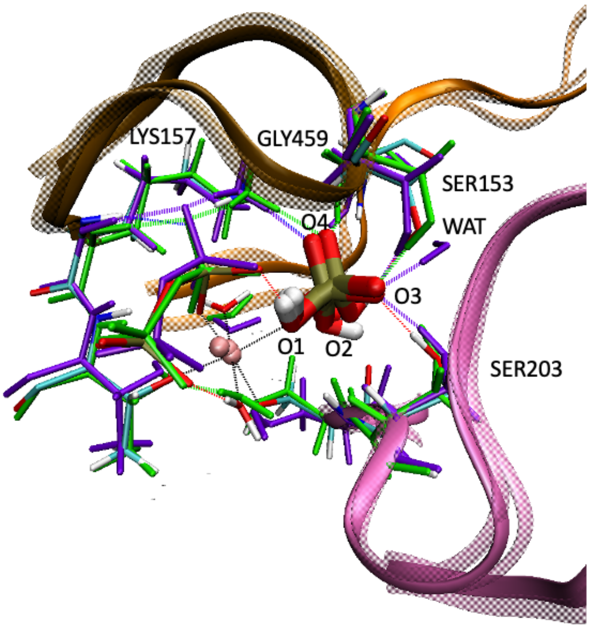
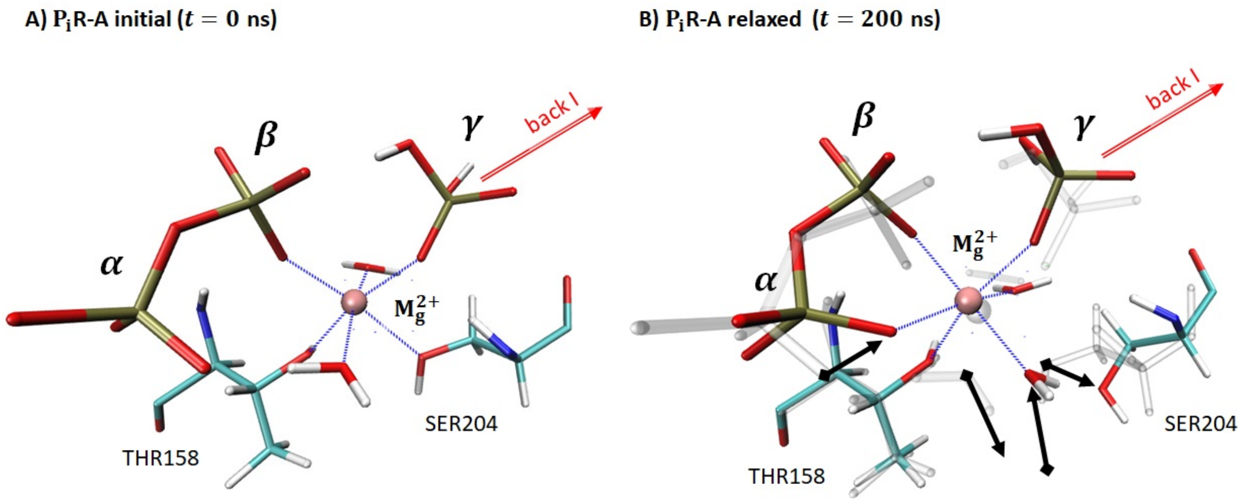


Figure B.3: Superposition of the nucleotide binding loop of relaxed structures PPS-D (green), PPS-E (purple) and PPS-F (all atoms). The H-bonds are shows in the color of the molecule. The interactions with $Mg^{2+}$ are shown in black dashed lines for PPS-F. View of the Pipart. In addition to the other the PPS-E shows an interaction of the O3 with a water molecule.

## B.3 Coordination of $Mg^{2+}$

The Fig. B.4 shows the coordination of $Mg^{2+}$ for the structure $P_iR$-A at the start and the end of the relaxation simulation.



A) $P_iR$-A initial ($t = 0$ ns)

B) $P_iR$-A relaxed ($t = 200$ ns)

Figure B.4: Representation of the coordination of $Mg^{2+}$ for starting and relaxed structure of $P_iR$-A. ADP without the nucleoside part, $Mg^{2+}$, THR158 of P-Loop, SER204 of Switch I and water molecules coordinated with $Mg^{2+}$ are represented in licorice with color code C cyan, H white, O red, N blue, P brown, and $Mg^{2+}$ light pink. The interactions with $Mg^{2+}$ are shown in blue dashed lines. The superposed starting structure of $P_iR$-A is represented in ghost on the relaxed structure of $P_iR$-A to show the motion of SER204 residue, oxygen of Pi$\alpha$ and water molecules, which are emphasized by black arrows.

## B.4 Color chart along relaxation simulation

The Fig. B.5, B.6, B.7 and B.8 give the interaction along the relaxation dynamics. For the interaction analysis of the 200 ns relaxation simulations, 100 snapshots were taken every 1 000 000 simulated snapshots, which corresponds to 1 snapshot every 2 ns simulated time. For each couple of residues (i,j), more than one heteroatom can be considered for each residue leading to formation of consecutive and/or simultaneous different H-bonds. For example, arginine residue has two N-H that can interact with the two O and two OH of Pi, leading to 16 different combinations of formation of H-bond. The concerned heteroatoms X, Y or Z of involved residue i are explicit in the colour charts as RESIDUEi@X,Y,Z. For each residue couples the minimum distance among all H-bond distances is collected. A coloured rectangle is plotted for each snapshot depending of this minimum distance : the darker the shorter is the distance, the more likely an H-bond is formed.

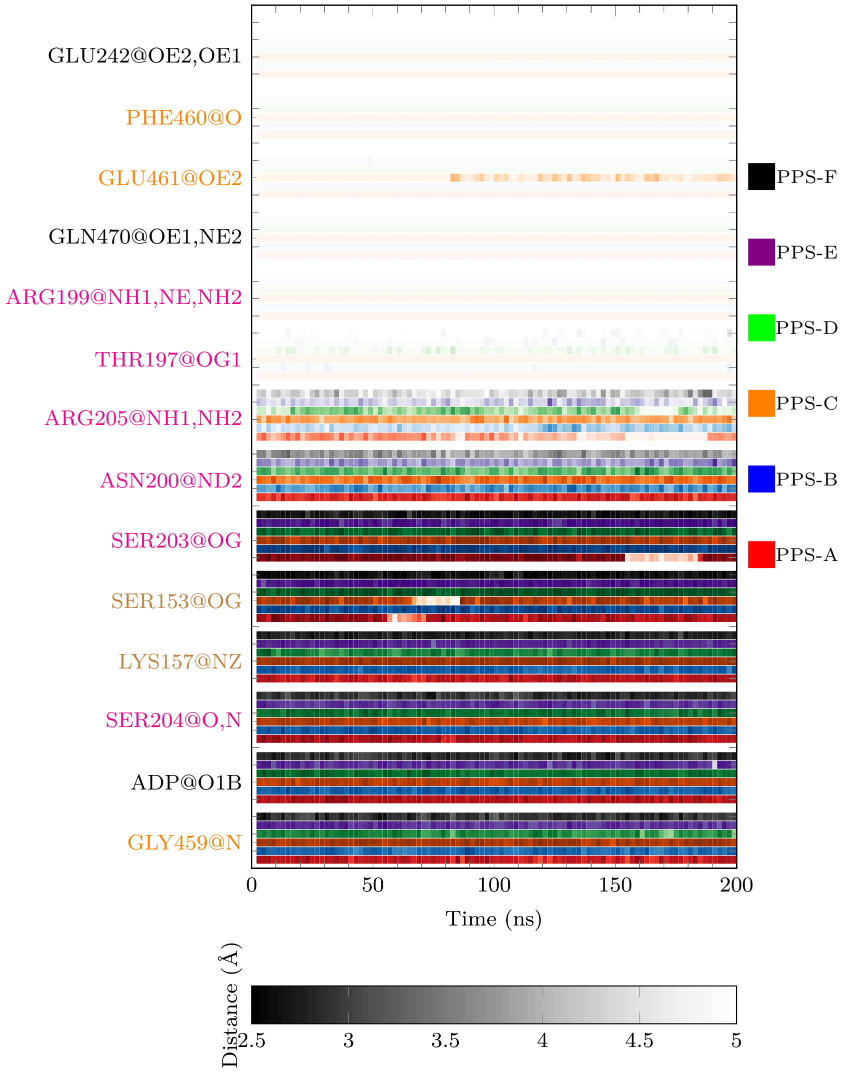


Figure B.5: Color chart of interactions of residues from Nucleotide Binding Loops with Piduring the relaxation molecular dynamics starting with PPS structures. The PPS-(A to F) correspond to different protonation states of the phosphate as described in the main text. The shortest distance of the atoms Xi or Xj of residue RESY (named RESY@Xi,Xj) with the phosphor of the phosphate is given for each simulation by a color box. The darker is the color the smaller the distance. Therefore, for each interaction, six lines from bottom to up for PPS-A to PPS-F are given.

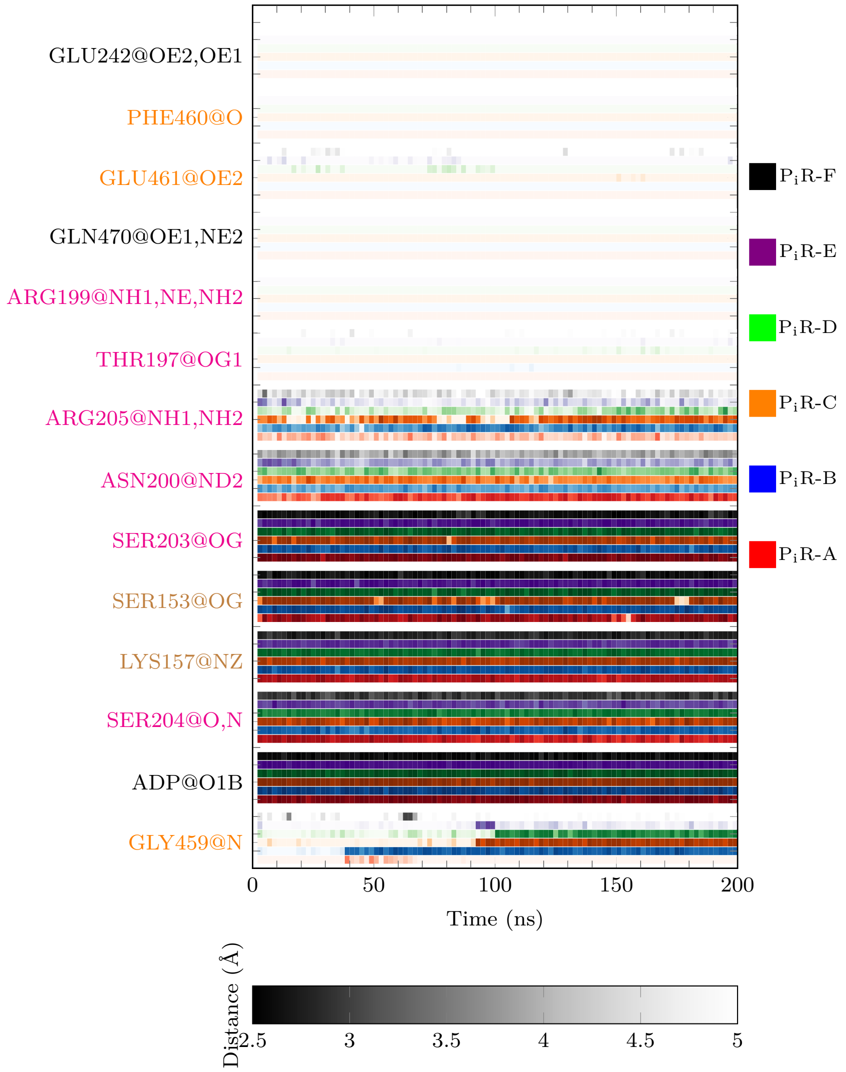


Figure B.6: Color chart of interactions of residues from Nucleotide Binding Loops with Pi during the relaxation molecular dynamics starting with $P_iR$ structures. The $P_iR$-(A to F) correspond to different protonation states of the phosphate as described in the main text. The shortest distance of the atoms Xi or Xj of residue RESY (named RESY@Xi,Xj) with the phosphor of the phosphate is given for each simulation by a color box. The darker is the color the smaller the distance. Therefore, for each interaction, six lines from bottom to up for $P_iR$-A to $P_iR$-F are given.

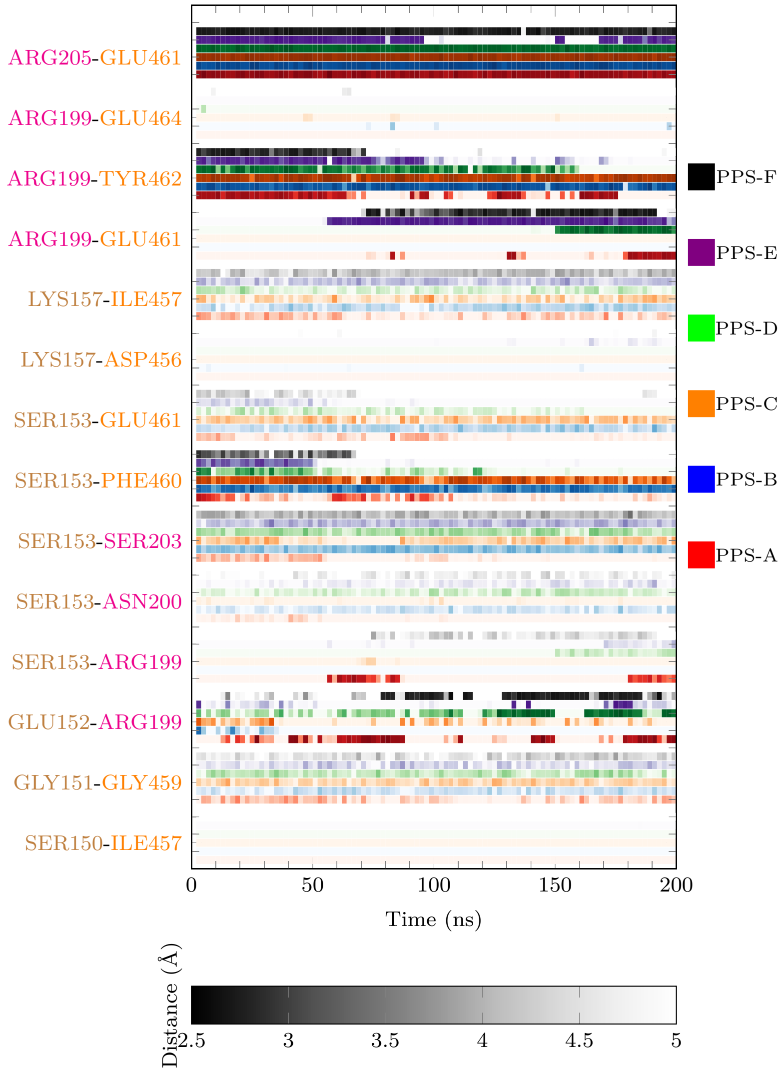


Figure B.7: Color chart of interactions between residues from Nucleotide Binding Loops during the relaxation molecular dynamics starting with PPS structures. The PPS-(A to F) correspond to different protonation states of the phosphate as described in the main text. The shortest distance between residue RESX and RESY are given for each simulation by a color boxes. The darker is the color of the box the smaller the distance. Therefore, for each interaction, six lines from bottom to up for PPS-A to PPS-F are given.

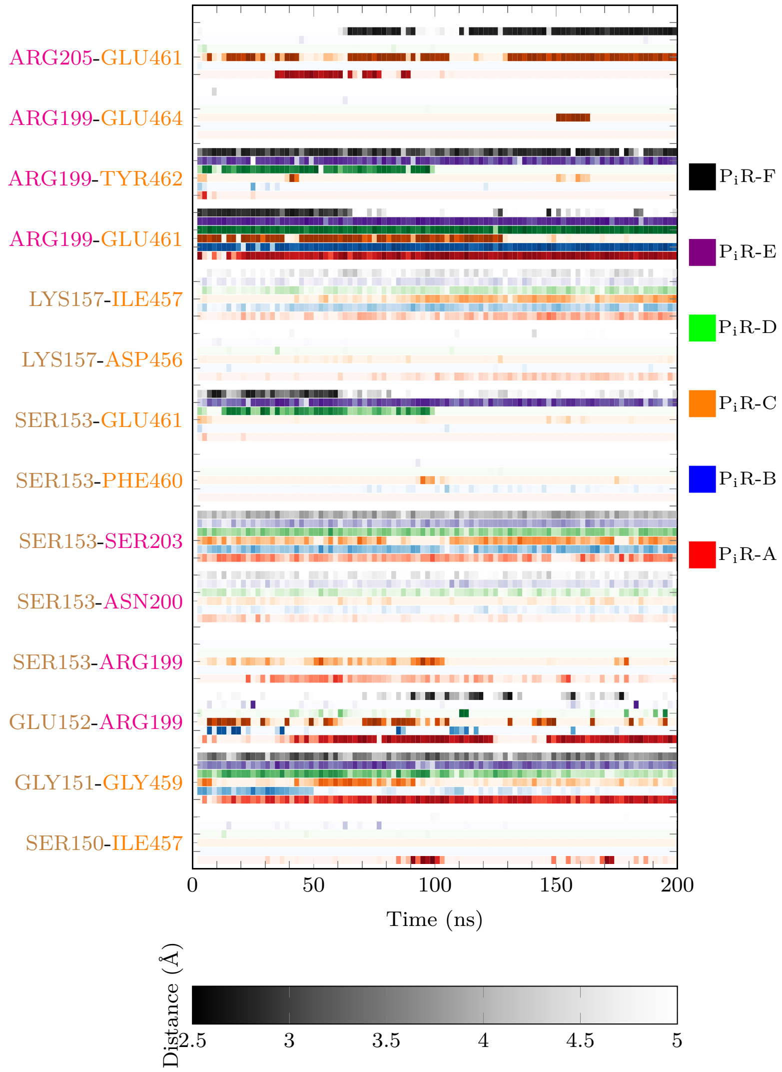


Figure B.8: Color chart of interactions between residues from Nucleotide Binding Loops during the relaxation molecular dynamics starting with $P_iR$ structures. The $P_iR$-(A to F) correspond to different protonation states of the phosphate as described in the main text. The shortest distance between residue RESX and RESY are given for each simulation by a color boxes. The darker is the color of the box the smaller the distance. Therefore, for each interaction, six lines from bottom to up for $P_iR$-A to $P_iR$-F are given.

## B.5 Distance between 460@O and 153@N, $d_3$

The Fig. B.9 shows the distance (460@O-153@N) $d_3$ for the structure PPS-E at the start and the end of the relaxation simulation.

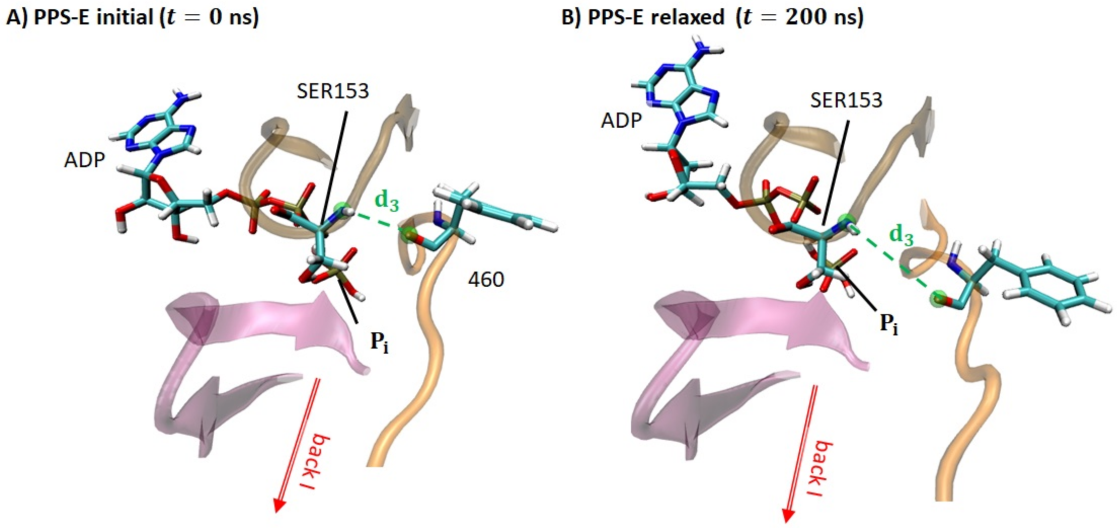


Figure B.9: Distance distance (460@O-153@N) $d_3$ is shown in green dashed line for starting (Switch II closed) and relaxed (Switch II open) structure of PPS-E. ADP without the nucleoside part, $Mg^{2+}$, SER153 of P-Loop, PHE460 of Switch II are represented in licorice with color code C cyan, H white, O red, N blue, P brown, and Mg2+ light pink. The interactions with $Mg^{2+}$ are shown in blue dashed lines.

## B.6 PMF for PPS showing the PPS-A simulation

The Fig.B.10 give a complement figure of PMF of PPS-(A to F) along the Umbrella Sampling.

## B.7 Color chart along Umbrella Sampling simulations

The Fig. B.11, B.12, B.13 and B.14 give the interaction along the Umbrella Sampling simulations. For the interaction analysis of US simulations colour chart analysis, 112 snapshots were taken every 1 500 000 simulated time steps, which corresponds to 2 snapshots per US windows/values of the constraint. For each residue couples the minimum distance among all H-bond distances is collected. A coloured rectangle is plotted for each snapshot depending of this minimum distance : the darker the shorter is the distance, the more likely an H-bond is formed (see §B.4).

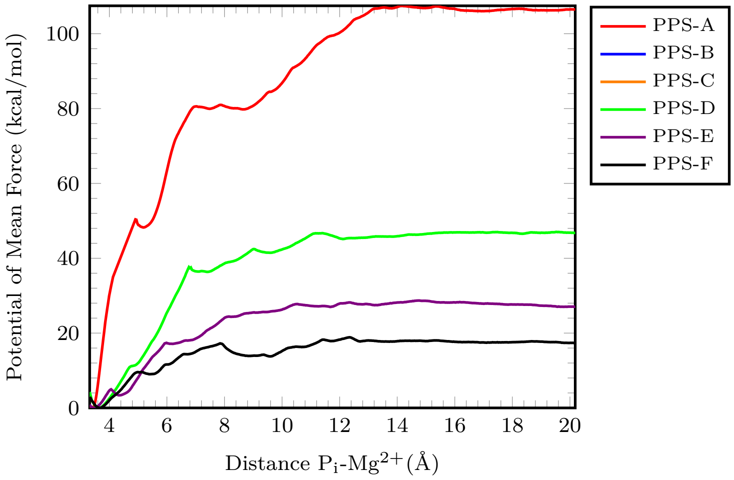


Figure B.10: Potential of mean force (PMF) for Pi departure for each obtained structure after relaxation of different initial orientations starting from PPS crystallographic structure PPS-(A to F). The PMF of structures PPS-B and PPS-C are not shown as the Piis not released. This figure is a complement of the figure of main text showing the total PMF of structure PPS-A.

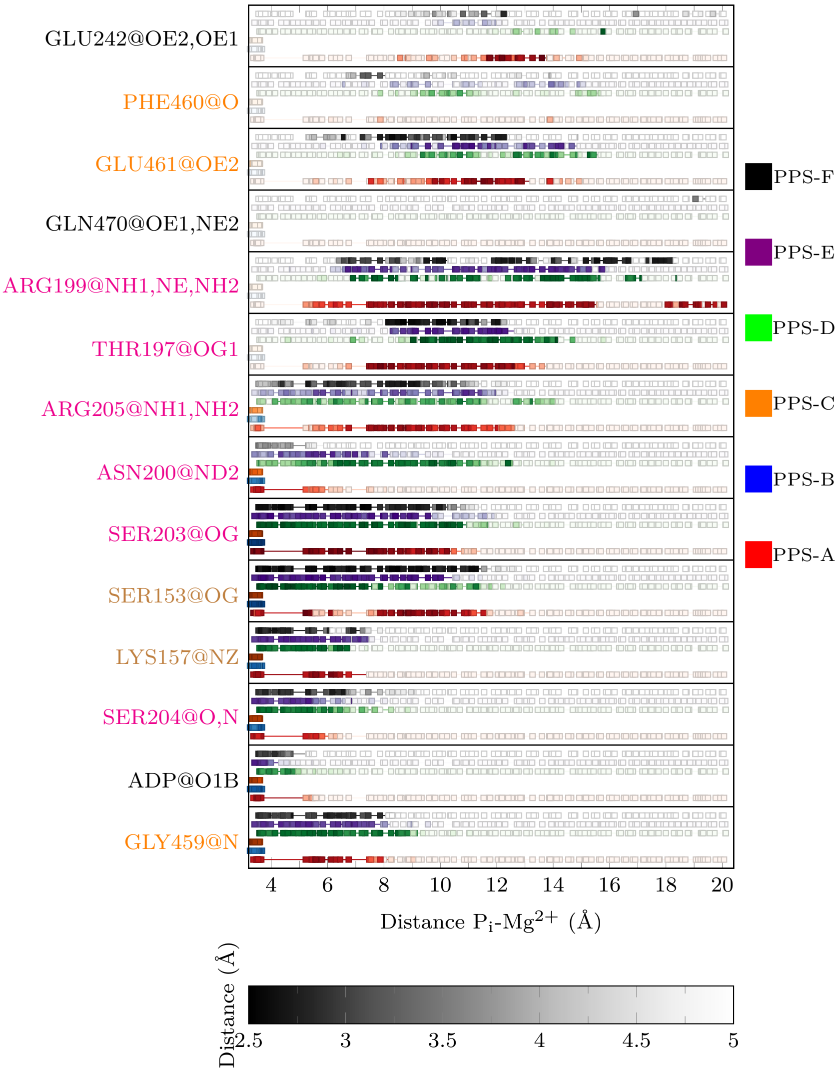


Figure B.11: Color chart of interactions of residues from Nucleotide Binding Loops with Pialong the Umbrella Sampling of departure of Pi starting with PPS structures. The PPS-(A to F) correspond to different protonation states of the phosphate as described in the main text. The shortest distance of the atoms Xi or Xj of residue RESY (named RESY@Xi,Xj) with the phosphor of the phosphate is given for each simulation by a color box. The darker is the color the smaller the distance. Therefore, for each interaction, six lines from bottom to up for PPS-A to PPS-F are given.

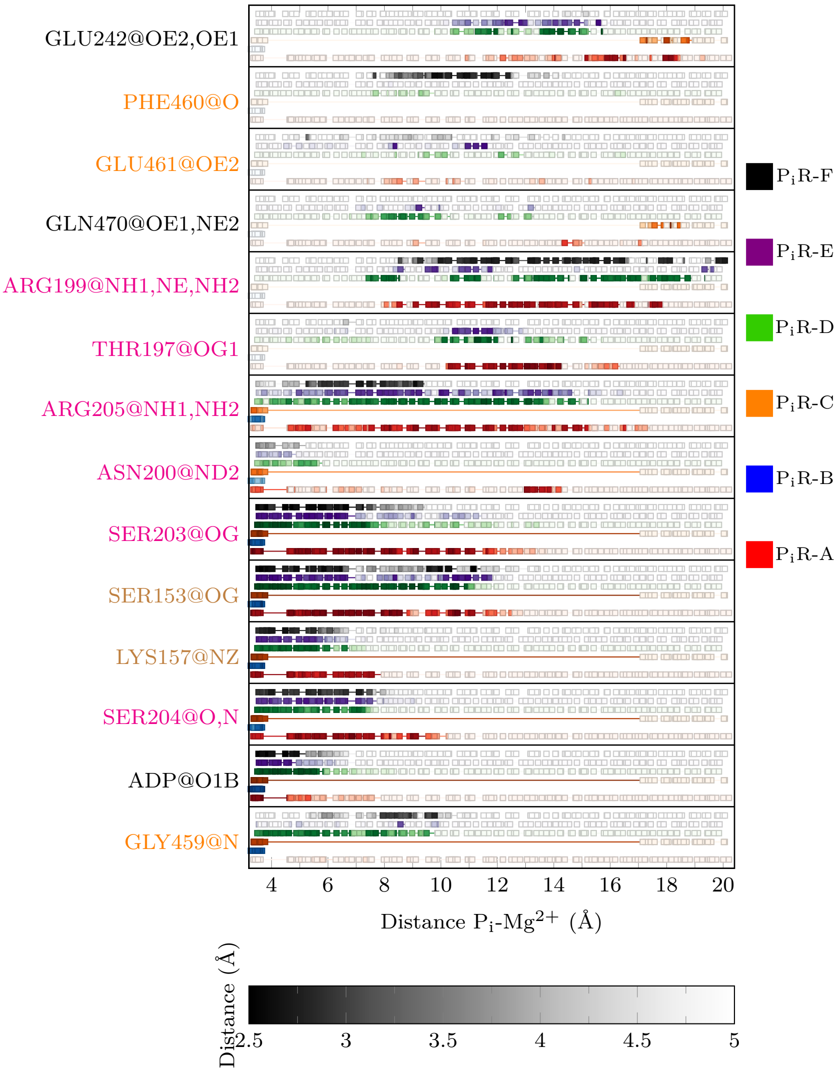


Figure B.12: Color chart of interactions of residues from Nucleotide Binding Loops with Pialong the Umbrella Sampling of departure of Pi starting with P$_i$R structures. The P$_i$R-(A to F) correspond to different protonation states of the phosphate as described in the main text. The shortest distance of the atoms Xi or Xj of residue RESY (named RESY@Xi,Xj) with the phosphor of the phosphate is given for each simulation by a color box. The darker is the color the smaller the distance. Therefore, for each interaction, six lines from bottom to up for P$_i$R-A to P$_i$R-F are given.

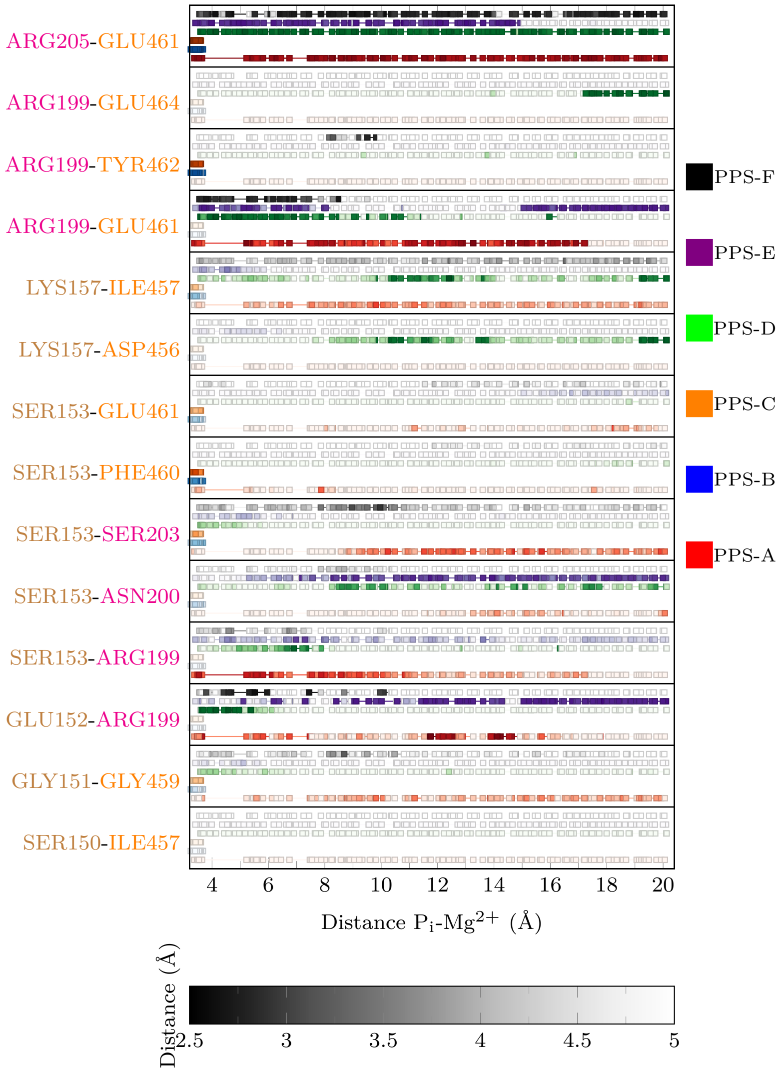


Figure B.13: Color chart of interactions between residues from Nucleotide Binding Loops along the Umbrella Sampling of departure of Pi starting with PPS structures. The PPS-(A to F) correspond to different protonation states of the phosphate as described in the main text. The shortest distance between residue RESX and RESY are given for each simulation by a color boxes. The darker is the color of the box the smaller the distance. Therefore, for each interaction, six lines from bottom to up for PPS-A to PPS-F are given.

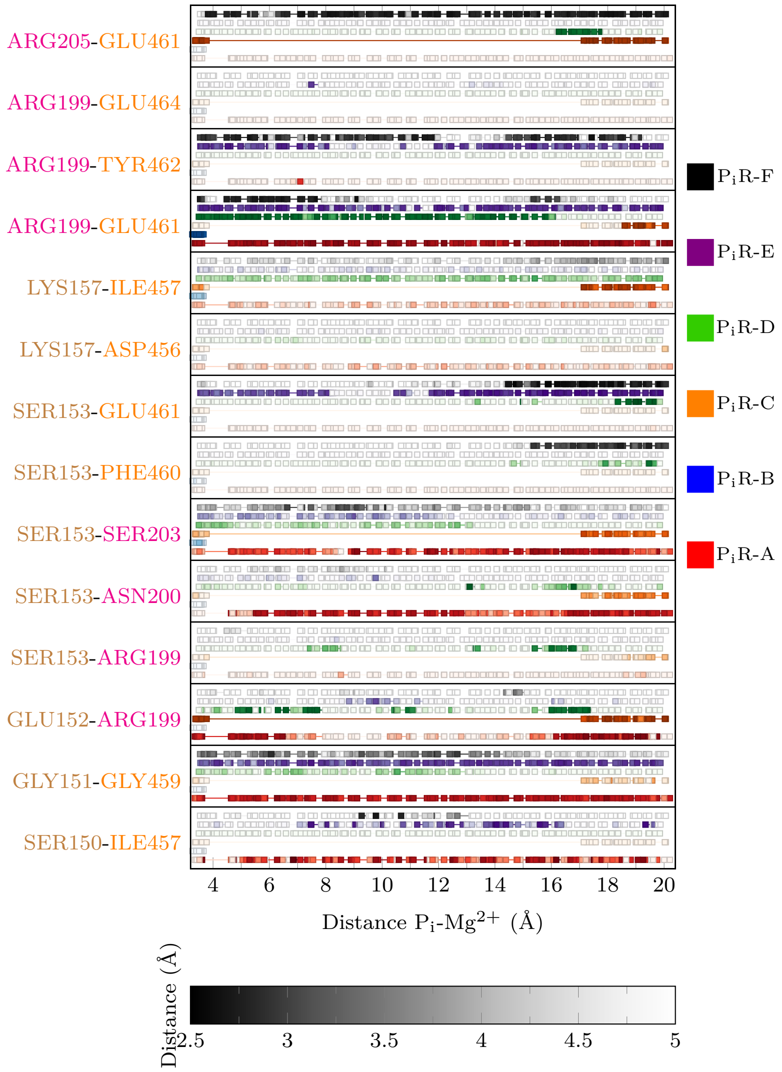


Figure B.14: Color chart of interactions between residues from Nucleotide Binding Loops along the Umbrella Sampling of departure of Pi  starting with $P_iR$ structures. The $P_iR$-(A to F) correspond to different protonation states of the phosphate as described in the main text. The shortest distance between residue RESX and RESY are given for each simulation by a color boxes. The darker is the color of the box the smaller the distance. Therefore, for each interaction, six lines from bottom to up for $P_iR$-A to $P_iR$-F are given.

# Appendix C

# Pole representation of vector directions

## Contents

## C.1 Stereographic projection

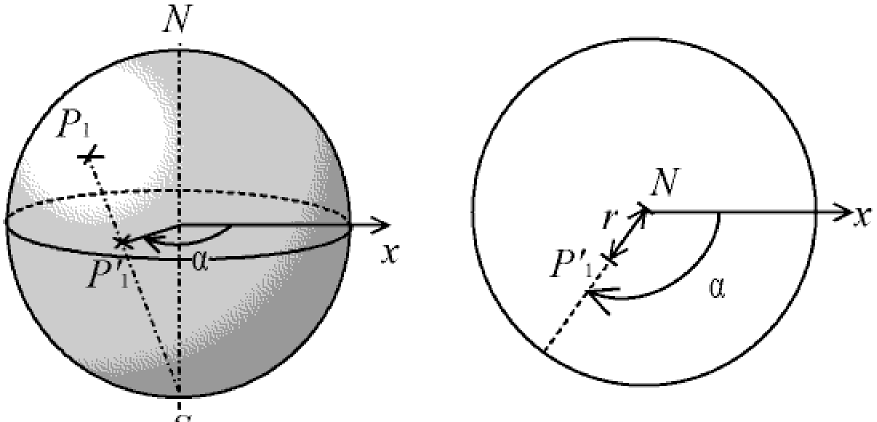


Figure C.1: Explication of the stereographic projection. Figure adapted from Christophe Dang Ngoc Chan (Wikimedia).

For the purpose of representing 3D direction on a 2D plane, we use the stereographic projection. The stereographic projection consists in a transformation of projection from 3D coordinates into a plane. This projection is represented on Fig. C.1.

Let's consider the 3D unit vector $\vec{V}$ with coordinates ($V_1$,$V_2$,$V_3$) in an orthogonal basis. By definition, this vector points on the surface of a sphere of radius 1 centered on (0,0,0).

The south pole of the sphere is defined as the point of coordinate (0,0,-1). Let's consider a straight line linking the south pole of the circle with the point of coordinates ($V_1$,$V_2$,$V_3$). The stereographic projection corresponds to the intersection of this straight line with the equatorial plane defined by z=0, as shown on Fig. C.1. The coordinates (x,y) of this vector after stereographic projection are calculated as follows:

$$\begin{bmatrix} x = \lambda \times V_1 \\ y = \lambda \times V_2 \end{bmatrix} \tag{C.1}$$

with

$$\lambda = \frac{1}{1+V_3} \tag{C.2}$$

## C.2 Representation of set of directions in a pole figure

Consider a straight line, $\Delta_i$, passing through the origin of the sphere, with direction vector $\overrightarrow{V}_{\Delta_i}$. To define the unit direction vector $\overrightarrow{V}_{\Delta_i}$, we have two possible choices, we take by convention the directing vector which points towards the North pole ($\overrightarrow{V}_{\Delta_i} \cdot \overrightarrow{z} > 0$).

The representation of the direction $\Delta_i$ in the pole figure corresponds to the points $P_i'$ which is the stereographic projection of the point $P_i$ such that $\overrightarrow{OP_i} = \overrightarrow{V}_{\Delta_i}$ . It is interesting to note that the convention ($\overrightarrow{V}_{\Delta_i} \cdot \overrightarrow{z} > 0$) guarantees that P' remains in the unit circle.

By extension, a set of directions $\{\Delta_i\}$ is represented by a set of points $P_i'$ which are the stereographic projection of the points $P_i'$ such that $\overrightarrow{OP_i'} = \overrightarrow{V}_{\Delta_i}$ with $\overrightarrow{V}_{\Delta_i}$ the unit direction vector of $\Delta_i$ the pointing towards the north pole $N$.

# Appendix D

# Transformation from PPS to Strong-ADP

## Contents

In this chapter we tested some constraints to force the system to transform from PPS to Strong-ADP. The idea was to give some insights to the question of whether the $\mathrm{P_iR}$ state was favorable or not during this transformation. If this state is ”mandatorily” (favorably) visited during this transformation, this would confirm its existence as hypothesized (between PPS and Strong-ADP). The manuscript in this form use some constraints that were used at main intend only as tests in the middle of the doctoral course. It appears that they do not allow complete transformation from PPS to Strong-ADP, although some change are observed. Although the parameters used were observed as having discriminating value between PPS and Strong-ADP, they have not been specifically chosen after an extensive search of optimal parameter for the exploration of these states. These parameters have rather been chosen as they were the first seen as possibly and readily usable.

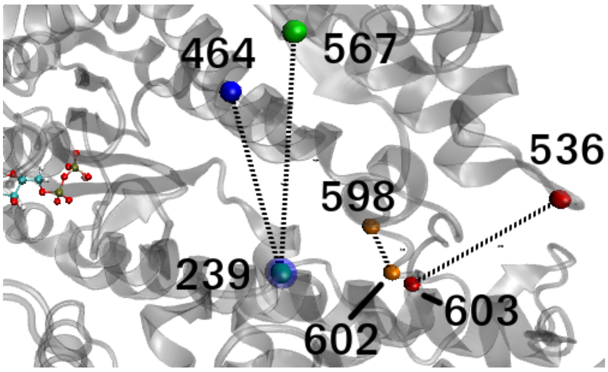


Figure D.1: Reminder of Fig. 4.2(a) to show the distances involved in the collective variable used for the constraint.

## D.1 Constraints used

### D.1.1 Sum of distances between backbone of 239-464 & 598-602 (1+4)

The distances 239-464 & 598-602 were found to be different in PPS and Strong-ADP (see 4), they were chosen to guide the transformation. In order to let some freedom on the system to evolve, the collective variable used is a sum of these distances.

Fig. D.2 shows the state of the protein at the start and at the end of the constrained molecular dynamics simulations performed with a constraint on the sum of the distances in the cleft 239-464 & 598-602. We can see a large rotation of the Converter in the sense PPS → Strong-ADP, which is wanted. However in the sense Strong-ADP → PPS, there is a huge rotation of the N-terminal Beta-Barrel. This rotation can be defined as a ”detachment” of this domain from its position which is not expected and is not wanted.

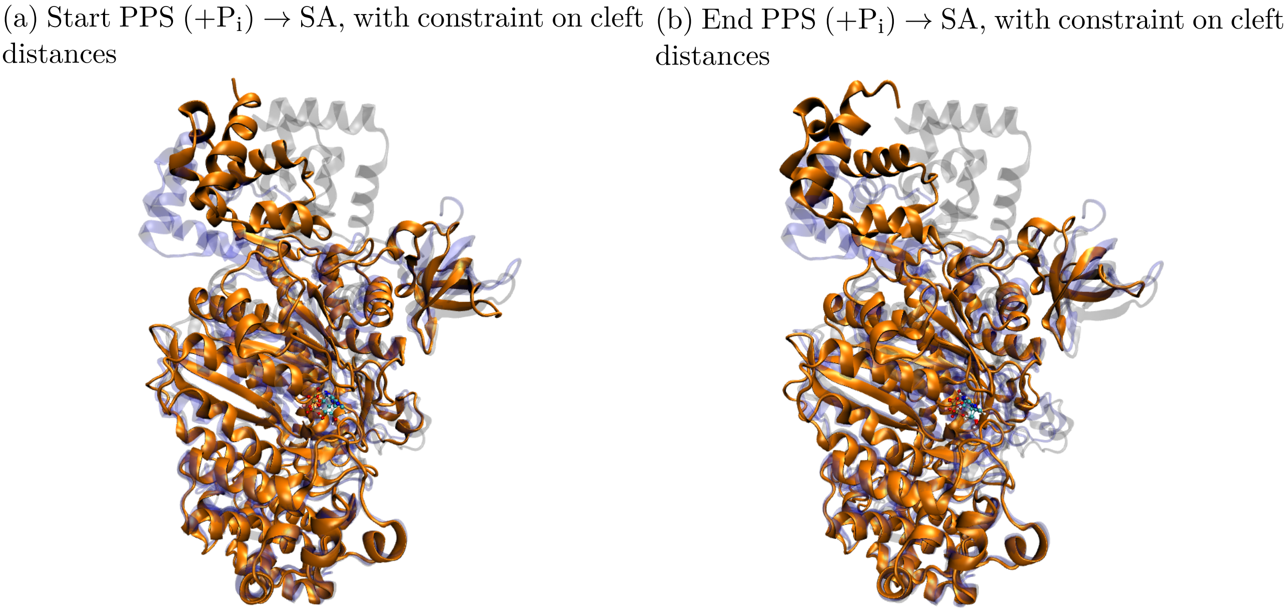
### D.1.2 Distance between Center of Mass of helices from the cleft (6 and 8)

The distance between COM of helices 6 and 8 was found to be different between PPS and Strong-ADP, along relaxations. It was then attempted a transformation using this only distance. This might have been ambitious as the difference was only very light (1.5 Åbetween their median).

Fig. D.6 shows the state of the protein at the start and at the end of the constrained molecular dynamics simulations performed with a constraint on the distance between COM of backbone atoms from helices 6 and 8. We cannot see large changes between the two structures.

### D.1.3 Distance between Center of Mass of helices 1 and 14

Fig. D.7 represent the state of the protein at the start and at the end of the constrained molecular dynamics simulations performed with a constraint on the distance between COM



(a) Start PPS (+$P_i$) → SA, with constraint on cleft distances

(b) End PPS (+$P_i$) → SA, with constraint on cleft distances

Figure D.2: Representation of the global state of the protein for the constrained dynamics with distances in the cleft 239-464 & 598-602. For PPS (+ $P_i$) → Strong-ADP (forward):(a) At the start (b) At the end. The shadows of crystal states of PPS (blue) and Strong-ADP (gray) are also shown.

of backbone atoms from helices 1 and 14. We can see an important rotation of the converter subdomain from its state in PPS to something closer from its state observed in Strong-ADP.

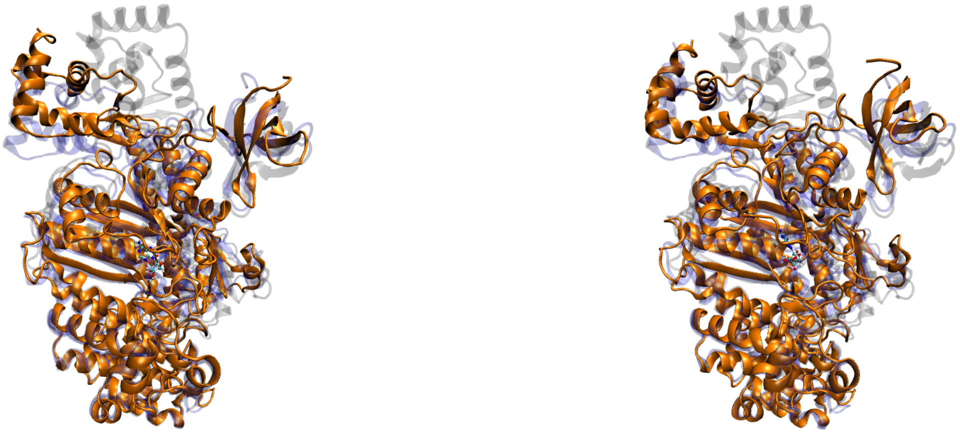
## D.2 Potential of Mean Force obtained

Fig. D.8 shows the PMF obtained with the attempts to transform the system from a PPS state into a Strong-ADP state.

The PMF obtained with the constraint on $L^{6-8}$ (see Fig. D.8.a) is of parabollic form centered on the average value observed in PPS relaxation. This form is due to the non dissociated range of values visited in PPS and Strong-ADP.

The most interesting study might be with the sum of distances from the cleft (1+4) (see Fig. D.8.b). This constraints was used for the transformation from PPS (with and without $P_i$) to Strong-ADP. What can be observed is that when $P_i$ is present in the active site, the PMF exhibits a large energy barrier toward the value in Strong-ADP, while it seems almost free to fluctuate toward PPS or Storng-ADP values when the $P_i$ is absent.

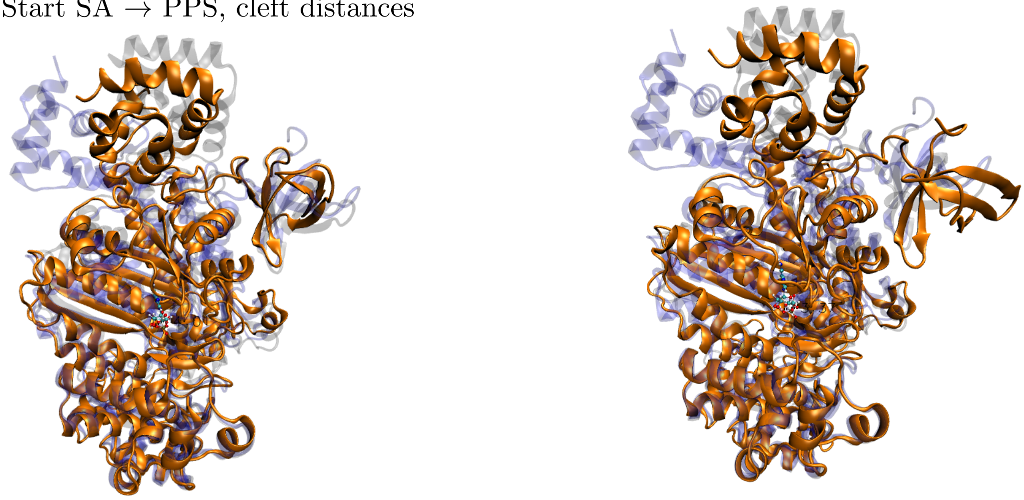
## D.3 Identification of structures



(a) Start PPS (no $P_i$) → SA, with constraint on cleft distances

(b) End PPS (no $P_i$) → SA, with constraint on cleft distances

Figure D.3: Representation of the global state of the protein for the constrained dynamics with distances in the cleft 239-464 & 598-602. For PPS (no $P_i$) → Strong-ADP (forward):(a) At the start (b) At the end. The shadows of crystal states of PPS (blue) and Strong-ADP (gray) are also shown.



(a) Start SA → PPS, cleft distances

(b) End SA → PPS, cleft distances

Figure D.4: Similar to D.2 for Strong-ADP → PPS (backward): (a) At the start (b) At the end.

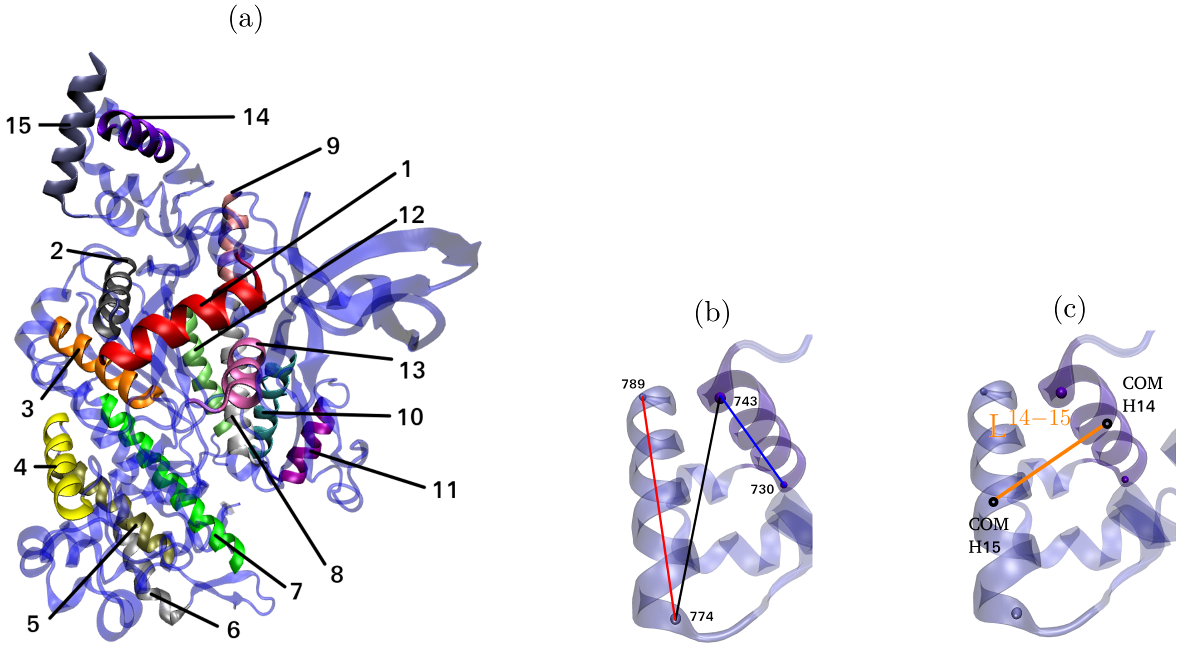
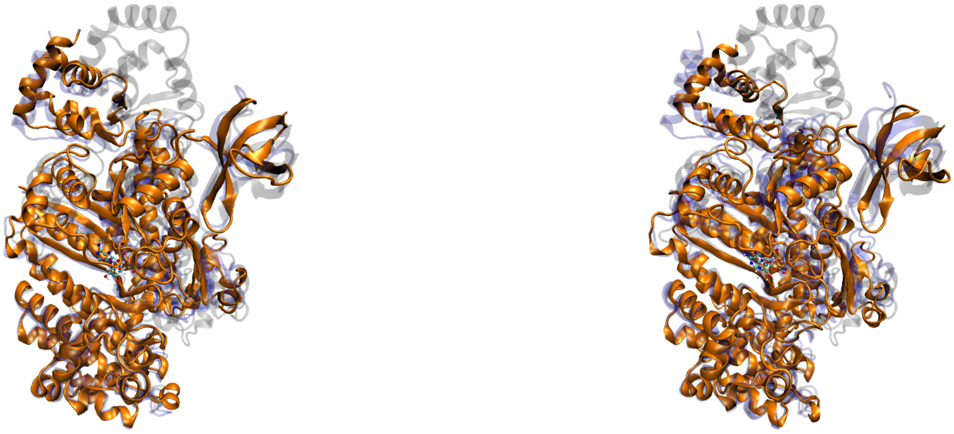


Figure D.5: Reminder of Fig. 5.1 to show the helices and how the distance between two helices is defined.



(a) Start PPS (+$P_i$) → SA, distance helices cleft (b) End PPS (+$P_i$) → SA, distance helices cleft

Figure D.6: Representation of the global state of the protein for the constrained dynamics with distances between COM of helices backbone 6 and 8. For PPS → Strong-ADP (forward):(a) At the start (b) At the end. The shadows of crystal states of PPS (blue) and Strong-ADP (gray) are also shown.

(a) Start PPS (+$P_i$) → SA, distance helices Converter (b) End PPS (+$P_i$) → SA, distance helices Converter

Figure D.7: Representation of the global state of the protein for the constrained dynamics with distances between COM of backbone atoms from helices 1 and 14. For PPS → Strong-ADP (forward):(a) At the start (b) At the end. The shadows of crystal states of PPS (blue) and Strong-ADP (gray) are also shown.

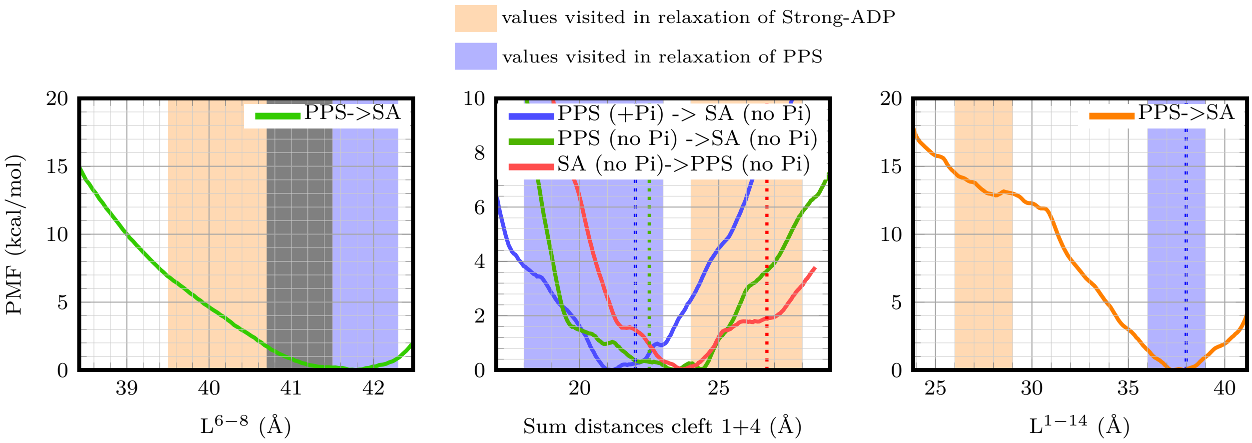


Figure D.8: Potential of Mean force for the transformation from PPS to Strong-ADP. (a) Using cleft distances. (b) Using helices distances from the cleft. (c) Using sum of 2 distances from the cleft, both from PPS to Strong-ADP and from Strong-ADP to PPS. (d) Using distance between Center of Mass from helices 6 and 8. Blue area corresponds to values visited along PPS unconstrained dynamics. Area correspond to values visited along unconstrained dynamics for (blue) PPS, (orange) Strong-ADP, (gray) superposition of area for PPS and Strong-ADP. Dotted lines corresponds to value of the initial snapshot used for the US simulation (blue) PPS, (red) Storng-ADP.

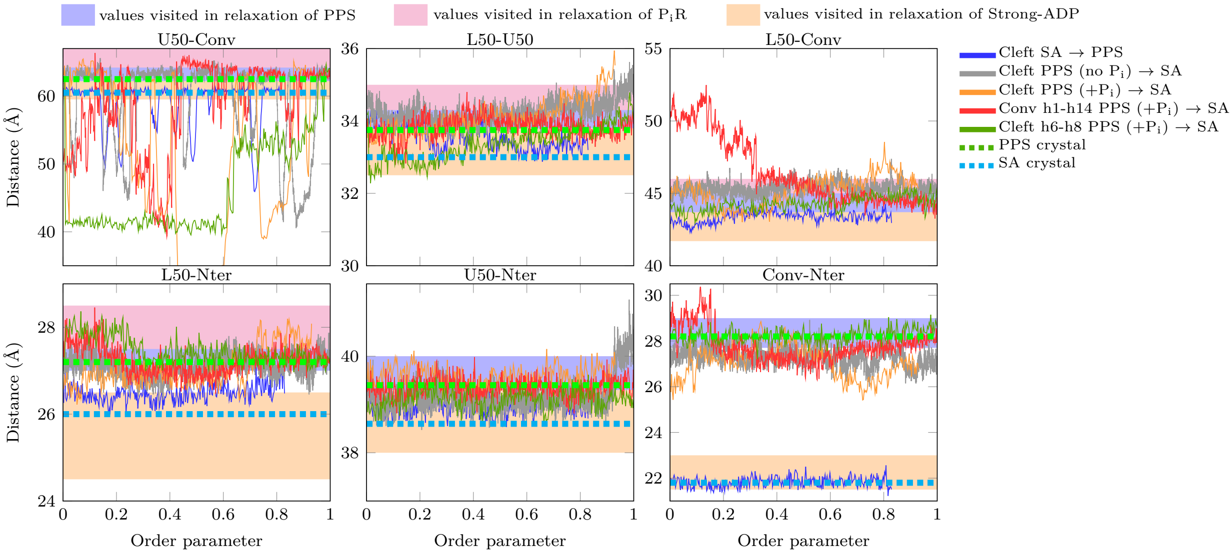


Figure D.9: Potential of Mean force for the transformation from PPS to Strong-ADP. (a) Using cleft distances. (b) Using helices distances from the cleft. (c) Using sum of 2 distances from the cleft, both from PPS to Strong-ADP and from Strong-ADP to PPS. (d) Using distance between Center of Mass from helices 6 and 8. Blue area corresponds to values visited along PPS unconstrained dynamics. Area correspond to values visited along unconstrained dynamics for (blue) PPS, (orange) Strong-ADP, (magenta) both PPS and Strong-ADP. Dotted lines corresponds to value of the initial snapshot used for the US simulation (blue) PPS, (red) Storng-ADP.

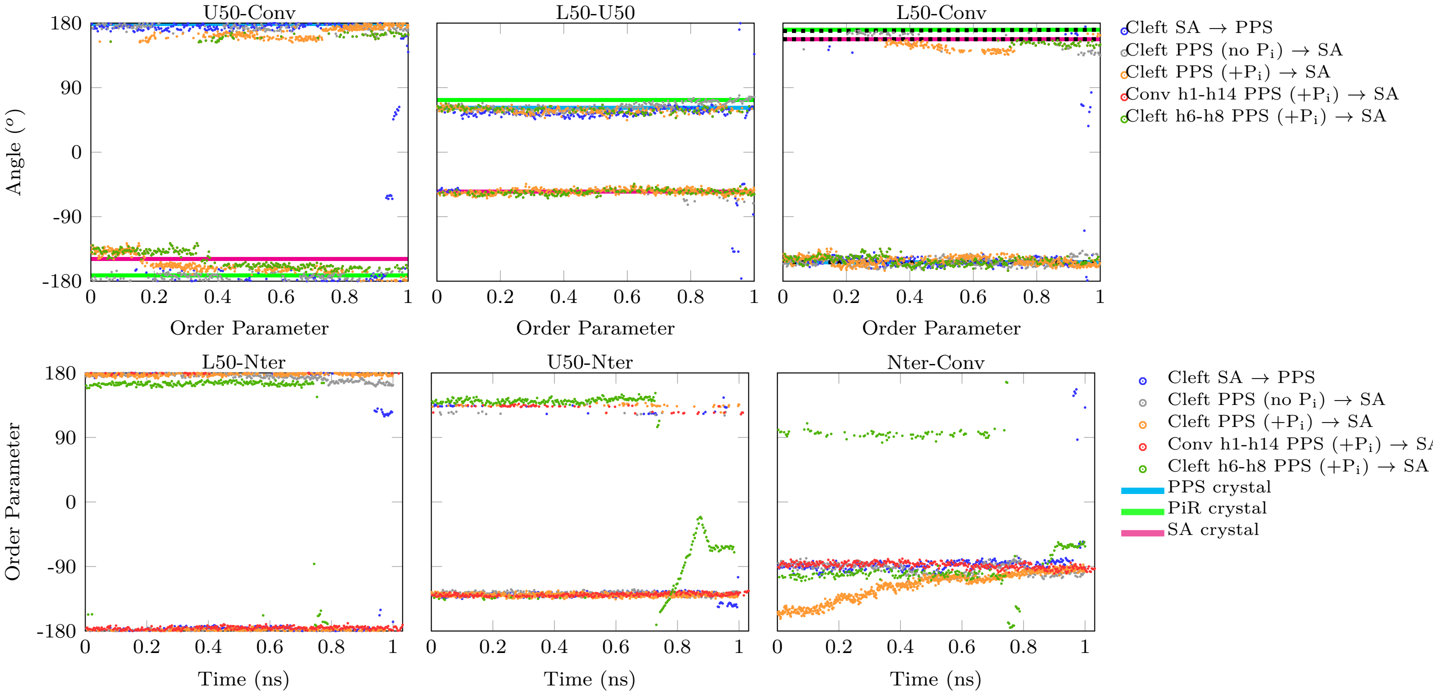


Figure D.10: Evolution of the angle in degree between the domains Lower 50 kDa, Upper 50kDa, N-terminal and Converter during Constrained dynamics. Is compared the first state to the second written above each graphs.

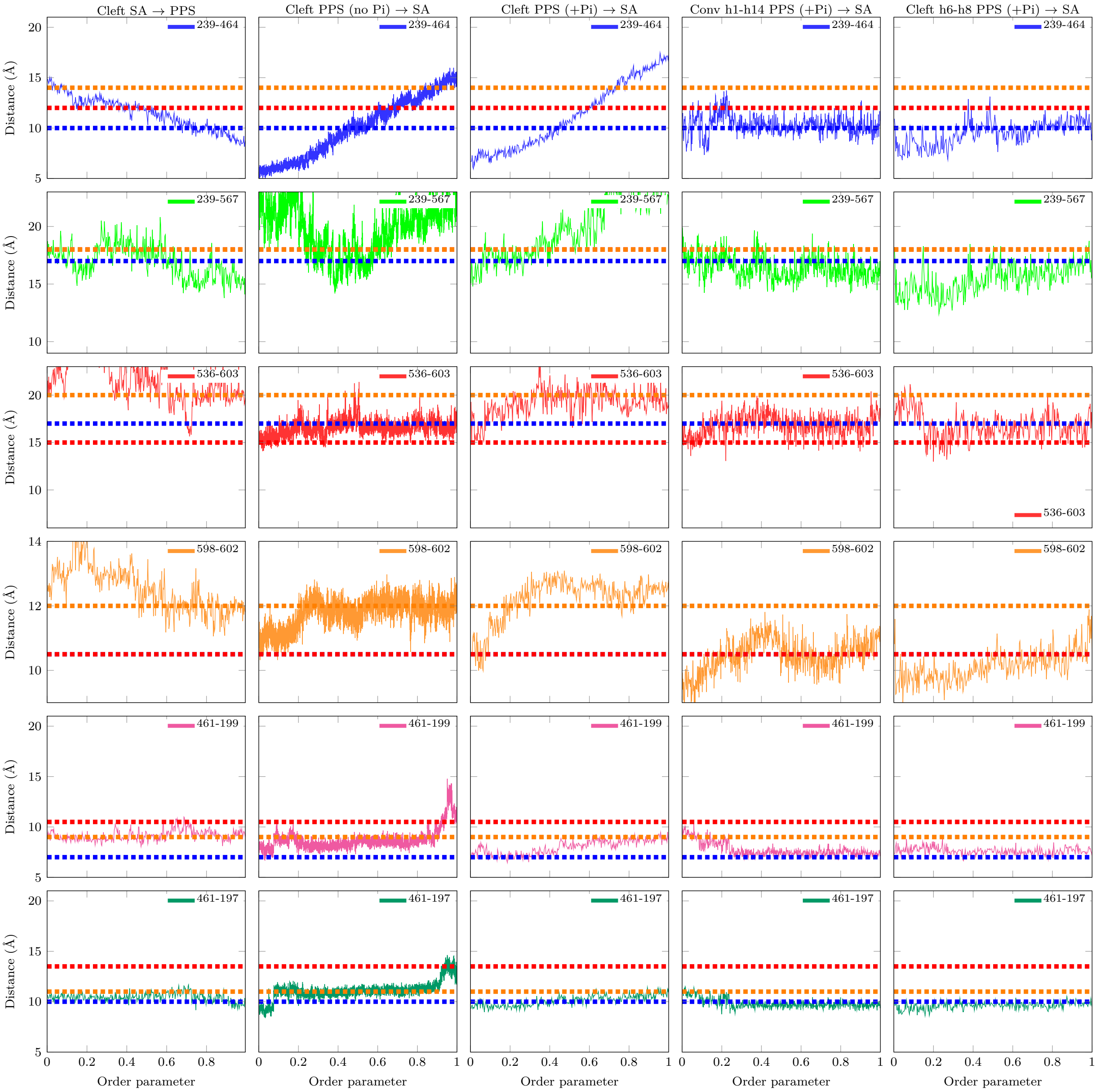


Figure D.11: Distances between $C_\alpha$ of residues representing the cleft closure along constrained dynamics of transformation from PPS to Strong-ADP. (blue) 239 and 464, (light green) 239 and 567, (red) 536 and 603, (orange) 598 and 602, (pink) 461-199, (dark green) 461-197.

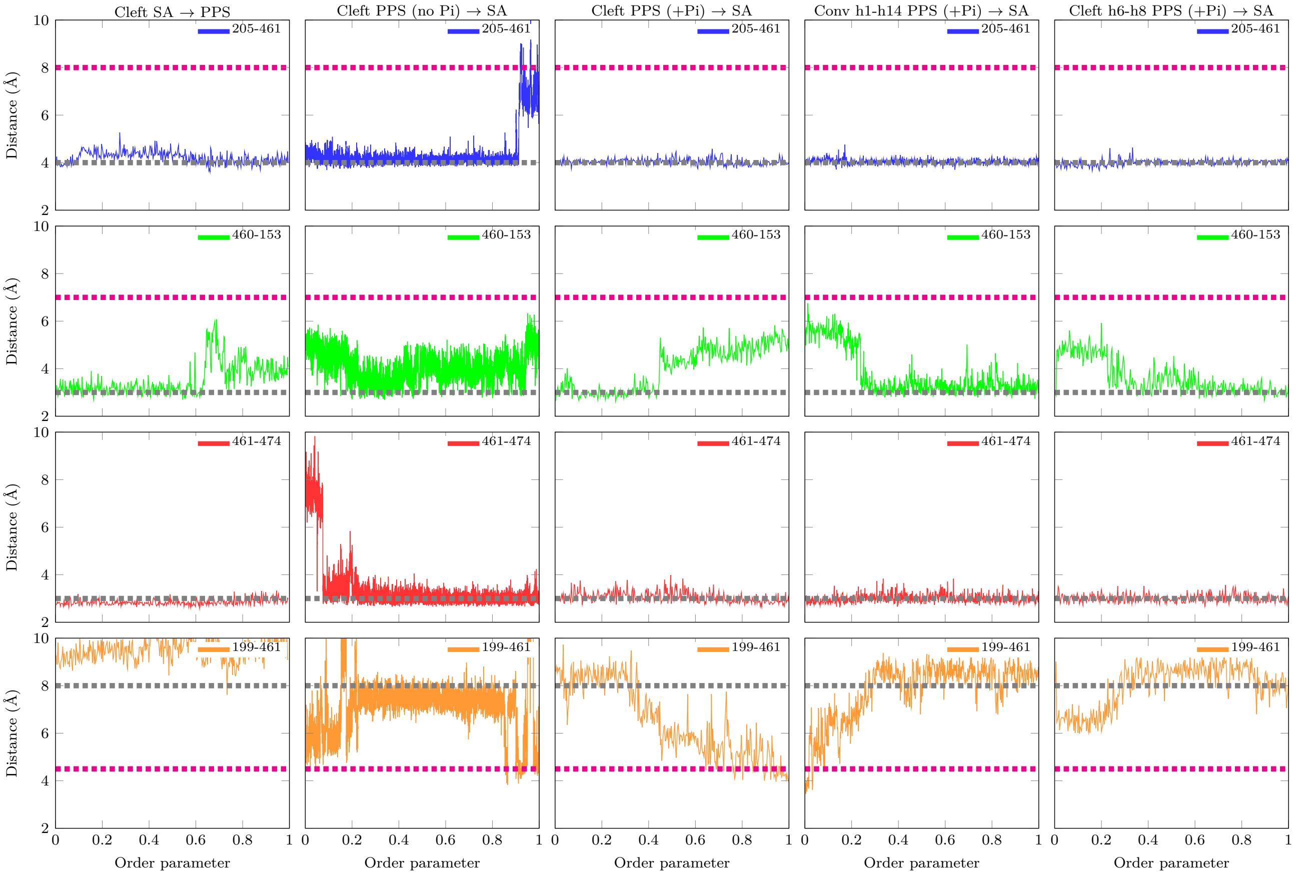


Figure D.12: Distances between $C_\alpha$ of residues representing the state of SwitchII along constrained dynamics of transformation from PPS to Strong-ADP. (blue) 205 and 461, (light green) 460 and 153, (red) 461 and 174, (orange) 199 and 461.

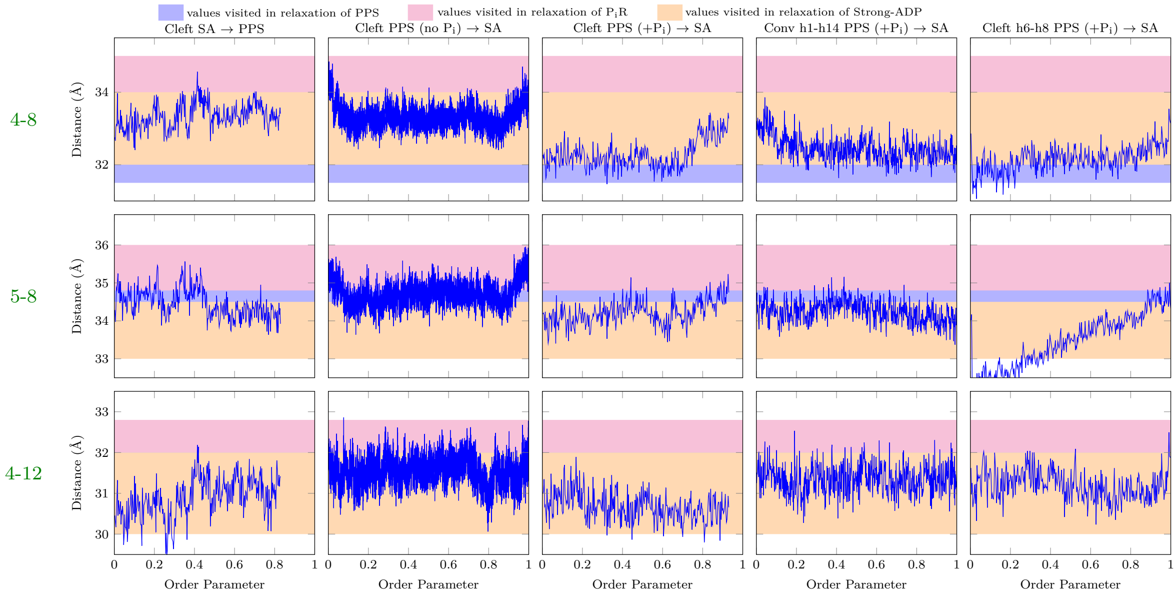


Figure D.13

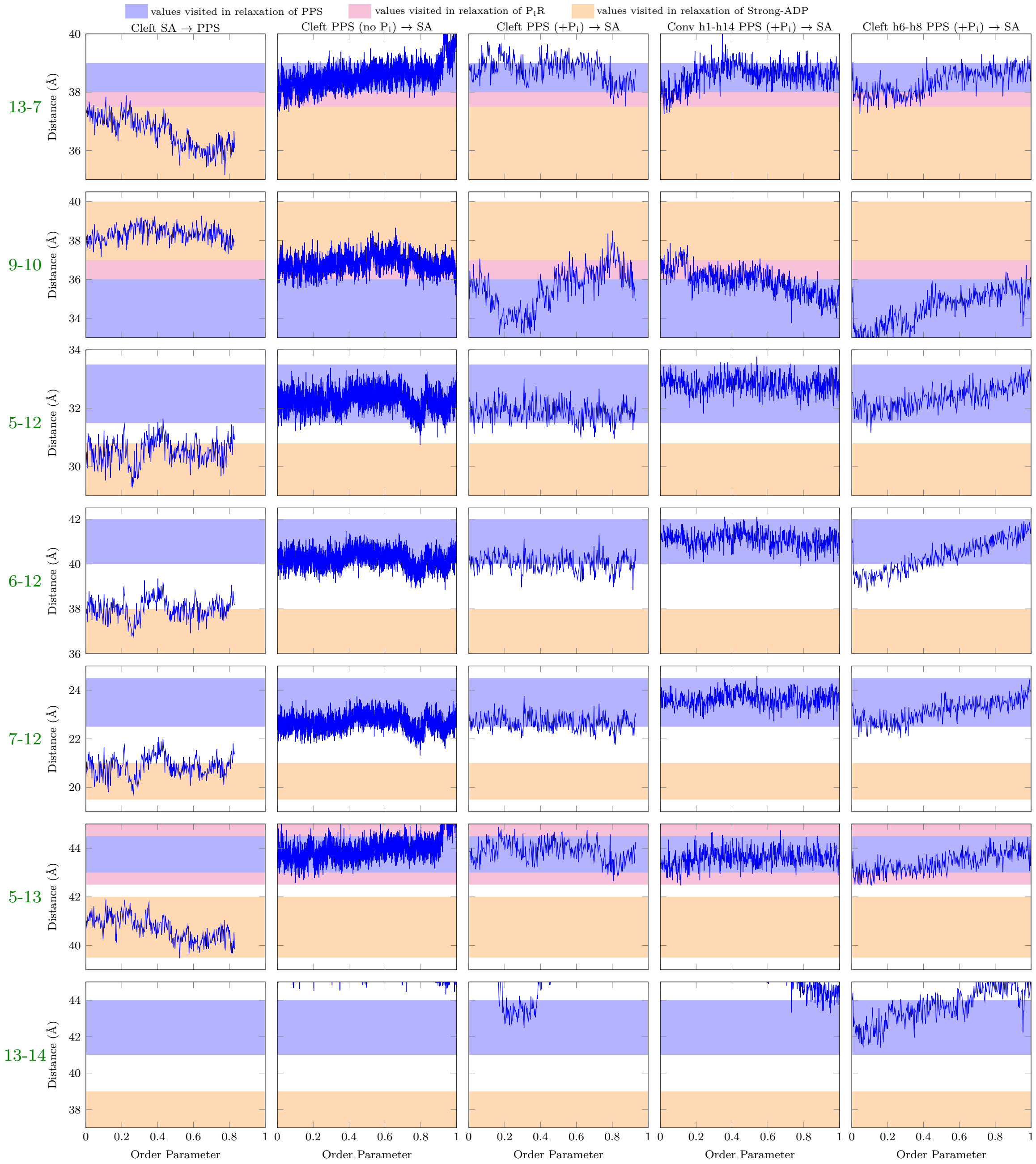

values visited in relaxation of PPS
values visited in relaxation of P_iR
values visited in relaxation of Strong-ADP
Cleft SA → PPS
Cleft PPS (no P_i) → SA
Cleft PPS (+P_i) → SA
Conv h1-h14 PPS (+P_i) → SA
Cleft h6-h8 PPS (+P_i) → SA
13-7
9-10
5-12
6-12
7-12
5-13
13-14
Distance (Å)
40
38
36
34
32
30
42
24
22
20
44
0
0.2
0.4
0.6
0.8
1
Order Parameter

Figure D.14

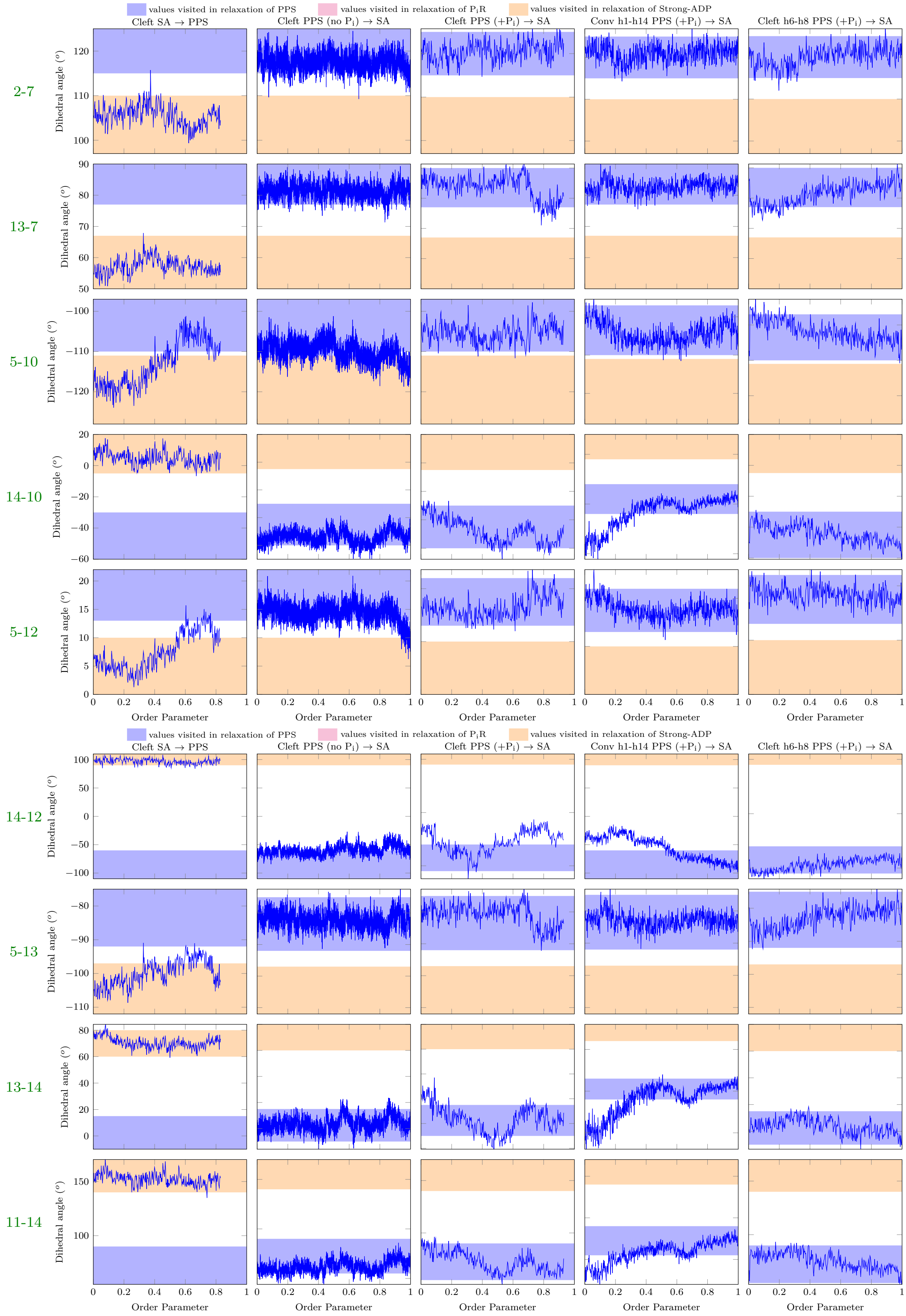

values visited in relaxation of PPS
values visited in relaxation of $P_iR$
values visited in relaxation of Strong-ADP
Cleft SA → PPS
Cleft PPS (no $P_i$) → SA
Cleft PPS (+$P_i$) → SA
Conv h1-h14 PPS (+$P_i$) → SA
Cleft h6-h8 PPS (+$P_i$) → SA
2-7
13-7
5-10
14-10
5-12
14-12
5-13
13-14
11-14
Dihedral angle (°)
Order Parameter
0
0.2
0.4
0.6
0.8
1
120
110
100
90
80
70
60
50
−100
−110
−120
20
0
−20
−40
−60
20
15
10
5
0
100
50
0
−50
−100
−80
−90
−100
−110
80
60
40
20
0
150
100

Figure D.15

## D.4 Conclusion

In conclusion we did not observe relevant transformation from PPS to Strong-ADP in our constrained simulations. Moreover, the $\mathrm{P_iR}$ state was not observed during these transformations. The choice of collective variable might not be good enough to describe the transformation. This tells that complex transformations like the one we wanted to perform cannot be performed with our choice of collective variables.

# Appendix E

# Influence of the presence of $P_i$ on the myosin active site

We wanted to know whether the presence of the $P_i$ had an impact on the myosin conformation and particularly the nucleotide binding loops and the backdoor opening. This questions originates from the fact that a $P_iR$ structure subjected to a high $P_i$ concentration tends towards a PPS conformation. Indeed, this phenomenon was observed with $P_iR$ crystals soaked enough time in a $P_i$ bath (Llinas et al., 2015).

The Fig. E.1 shows the evolution of distances between Nucleotide Binding Loops along PPS (a) and $P_iR$ relaxation in presence of $P_i$ in their active site. This $P_i$ was protonated in the six possible ways with two protonated oxygen atoms. By comparing the results obtained with PPS and $P_iR$, we can deduce a great global difference. The PPS simulations exhibit a maintained critical salt bridge (ARG205-GLU461) while the $P_iR$ simulations exhibit the presence of the secondary salt bridge. While these observations stand in average, we can still observe some partial formation of the critical salt bridge in some simulations from $P_iR$ and some opening with PPS.

The observation of the same distances with the Fig. E.2 shows that the global behaviour is mostly respected in the simulations of PPS and $P_iR$ in the absence of $P_i$. However there also exists one simulations from both PPS and $P_iR$ exhibiting the opposite behaviour. However this is also partially observed when $P_iR$ is simulated without $P_i$, although this closure might be less stable in these conditions. If this observation was to be confirmed, this would then suggest an impact of the presence of $P_i$ in the active site on the closure of the backdoor provoking a return to a state close to PPS.

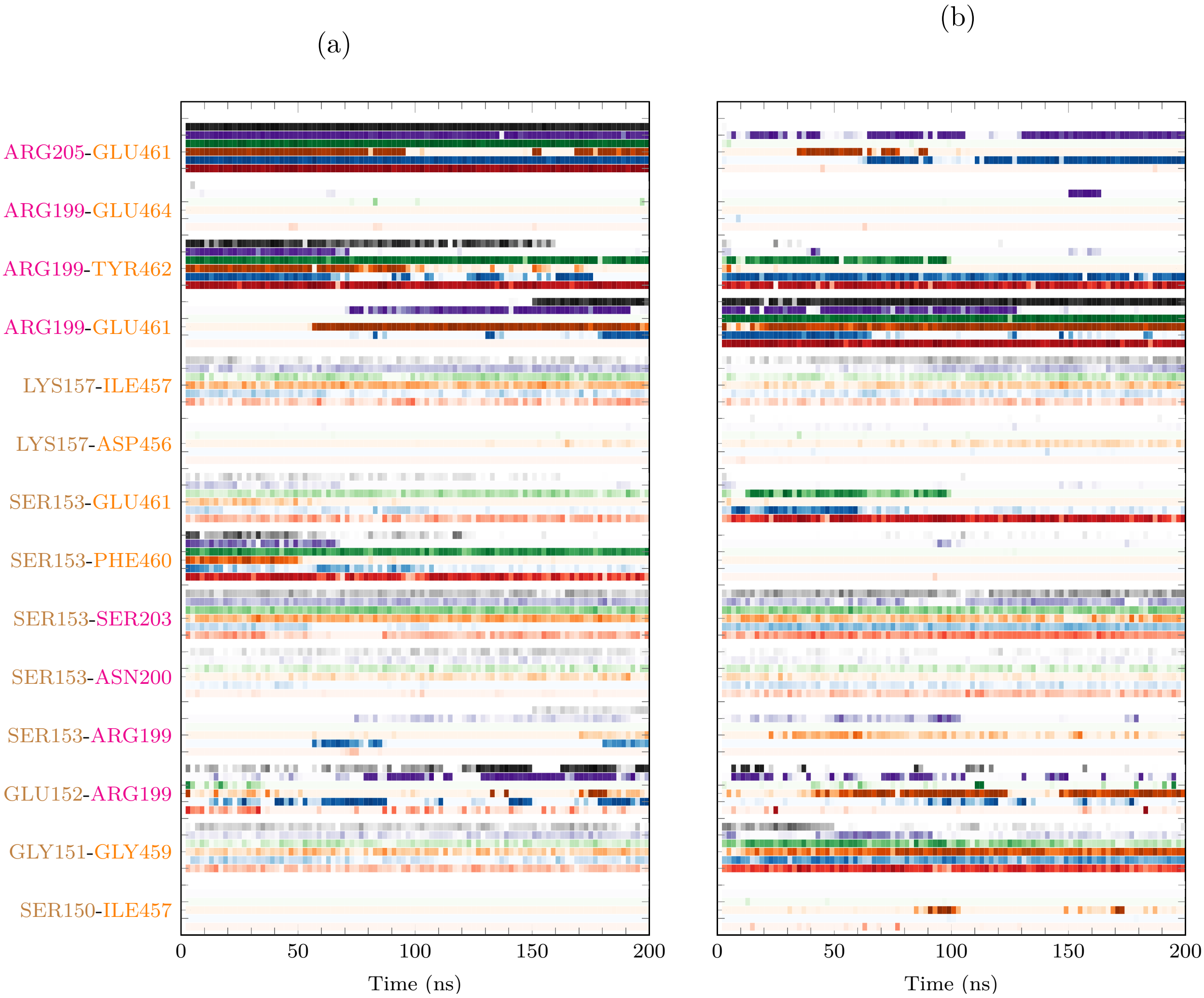


Figure E.1: Interactions between Nucleotide Binding Loops for the different simulations of PPS (a) and $P_iR$ (b) starting with different orientation of $P_i$.

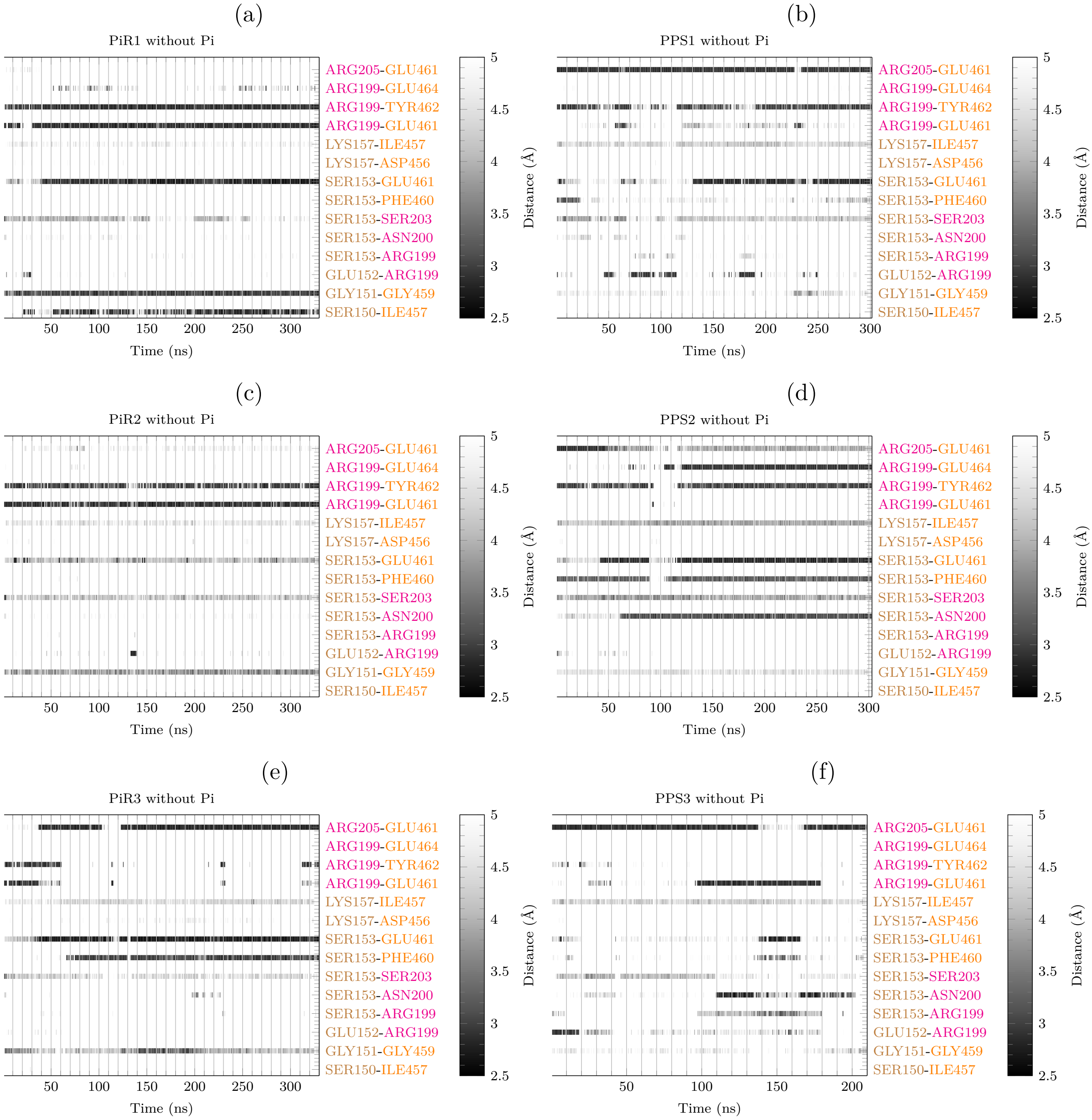


Figure E.2: Interactions between Nucleotide Binding Loops for the different simulations of $P_iR$ and PPS without $P_i$. (a,c,e) $P_iR$ without $P_i$ 1, 2 & 3 (b,d,f) PPS without $P_i$ 1, 2 & 3

# Appendix F

# Complements on the study of the forced departure of $P_i$

## Contents

In this chapter we describe the approaches used to try and give some insights into the mechanisms associated to the $P_i$ departure from the active site. First we present the final results concerning the $P_i$ departure with a constraint on the distance between $P_i$ and the $Mg^{2+}$. Then we present the attempts made previously to force the $P_i$ departure using Adaptive Steering Molecular Dynamics, Steering Molecular Dynamics and Umbrella Sampling. These attempts were made using a constraint on the distance between $P_i$ and a chosen helix next to the active site.

## F.1 Previous strategies developed to study $P_i$ departure

In addition to the presented results, several attempts have also been made with different couples of parameters of Umbrella Sampling, namely changing the "rigidity" of the constraint by changing the constant of the potential, the duration of sampling and the distance between centers of windows. In the case of ASMD it was tested several combinations of number of sections, velocity of pulling and number of trajectories.

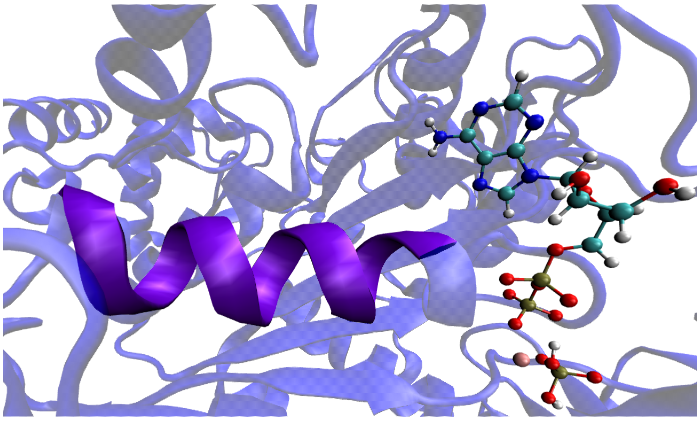


Figure F.1: Representation of the helix chosen as the first constraint. The helix is represented in violet. This helix corresponds to the helix 3 truncated.

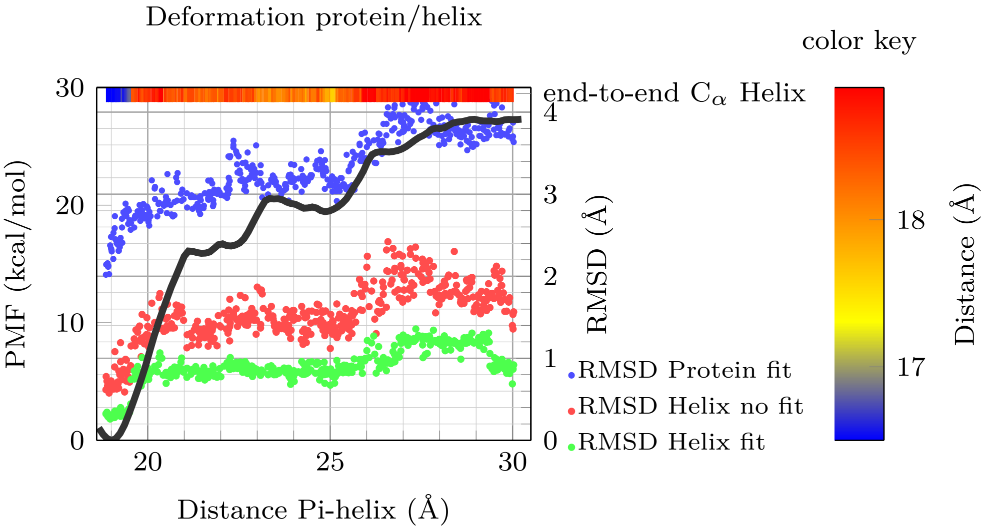


Figure F.2: Potential of Mean force of one simulation with Umbrella Sampling using distance with COM of the helix. Colored dots represent RMSD compared with the first the first snapshot of (blue) the whole protein after fitting the protein in each snapshot to minimize RMSD, (red) helix, without fitting, (green) constrained helix after fitting to minimize RMSD.

### F.1.1 First constraint : Distance with Helix

The first constraint used was composed of the distance between the center of mass of the $P_i$ and the center of mass of the $C_\alpha$ from the residues part of the portion of helix 3 comprised of residues 160 to 170. The chosen residues are represented on Fig. F.1.

### F.1.2 Pi departure with Umbrella Sampling using first constraint

We tested the computation of the free energy barrier with Umbrella Sampling, which we identified as a more reliable method, in the case of this constraint. However it can be seen with the greenish points on Fig. F.3 that the distance $P_i$-Mg is constant in the first part of the simulation. On contrary, the end-to-end distance between $C_\alpha$ of the helix is rapidly increasing at the

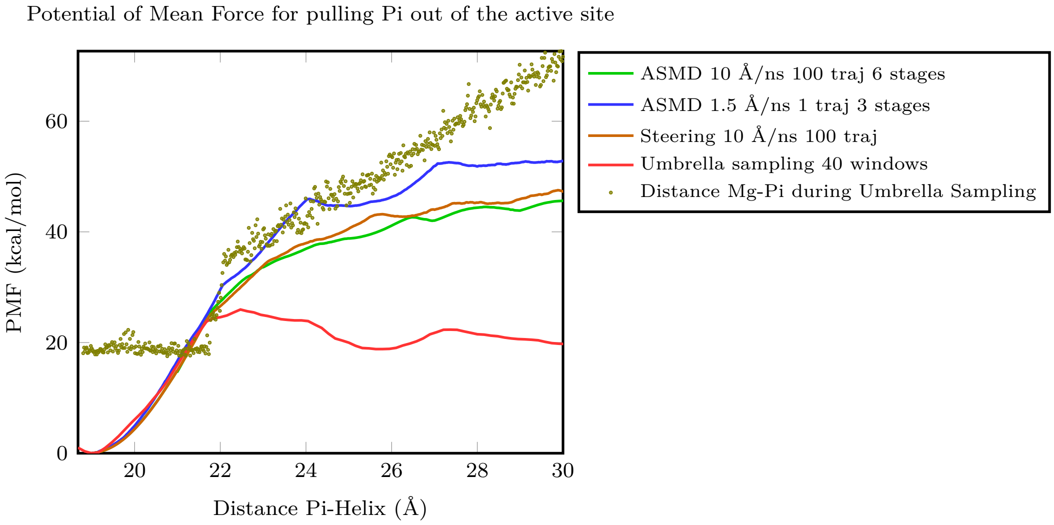


Figure F.3: Potential of Mean force using ASMD.

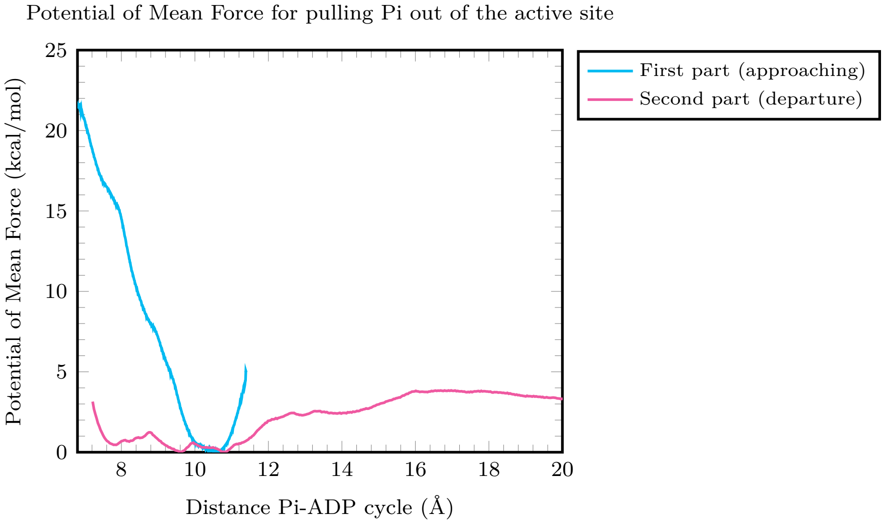
start of the simulation (see Fig. F.2, color bar in upper part). This means that the increase of the distance $P_i$-helix is mostly due to the unfolding of the helix. This unfolding results in displacing the COM of the helix farther from $P_i$, while this $P_i$ stays in the active site.

Yet, although this was not expected, this can be interpreted as the fact that $P_i$ displacement from the active site necessitates more energy than the deformation of the helix.

### F.1.3 Pi departure with Adaptive Steering Molecular Dynamics using first constraint

Fig. F.3 shows some PMF obtained using ASMD (and one PMF obtained with US) for the departure of the $P_i$. All the free energy calculations leading to these PMF were designed to have similar computational cost. What can be particularly noted is that among these PMF, ASMD gave way higher free energy barriers than what was obtained with Umbrella Sampling at similar computational cost. This might be due to the great difficulty to release the $P_i$, which necessitates some time for the system to equilibrate and change conformation along the departure. The system is not given the time to equilibrate during ASMD in comparison with US simulations that is supposed to be an equilibration sampling. Particularly, the ASMD might be confronted to the difficulty for the active site to equilibrate with all its important interactions and notably the coordination of $Mg^{2+}$, especially the separation of $Mg^{2+}$ and $P_i$, that is often seen correlated to a peak in free energy.

After a fail in the design of the constraint which simulation was unwillingly pulling the $P_i$ in the direction of $Mg^{2+}$ and ADP before pulling it out of the active site, it was performed a simulation of $P_i$ departure in two parts. The simulation was performed using as collective variable the distance between the center of mass of $P_i$ with the center of mass of the heavy



atoms from the cycle nearest from the phosphates in ADP. The computation of the Potential of Mean force associated to the two parts was performed separately and are presented on Fig. **??**.

Moreover, some tests were made with a constraint on the number of water in the coordination of $Mg^{2+}$.

## F.2 Impact of adding water molecules in coordination with $Mg^{2+}$

**Description of how we simulate** We bring the closest water molecule from $Mg^{2+}$ into the coordination of $Mg^{2+}$ by forcing the displacement of the water molecule with a constraint on the distance between the $Mg^{2+}$ and the water oxygen.

**Bringing Water molecules in the coordination of $Mg^{2+}$- PPS // where do they come from** The first water molecule comes from a position between C-terminal part of the Switch I and the N-terminal part of the Switch II. This water molecule takes the place of the oxygen from $P_i$, and so the $P_i$ molecule is no longer coordinated to $Mg^{2+}$. The second water molecule comes from a position between C-terminal part of the P-Loop and the N-terminal part of the Switch II. This water molecule replaces THR158.

**Bringing Water molecules in the coordination of $Mg^{2+}$- $P_i$R// where do they come from** The first water molecule comes from a position between C-terminal part of the Switch I and the N-terminal part of the Switch II. This water molecule replaces SER204. The second water molecule comes from a place between Switch-II and P-Loop. This water molecule replaces the $P_i$.

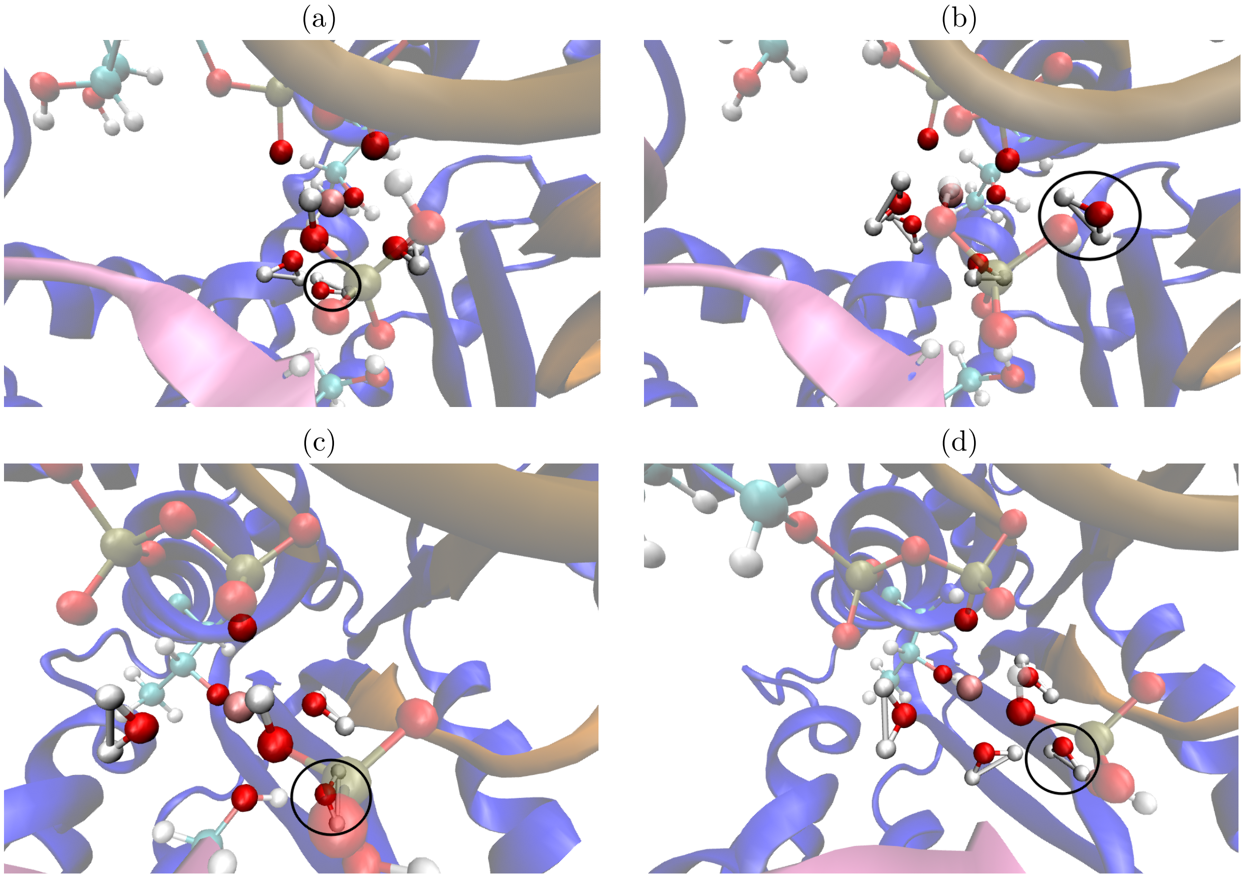
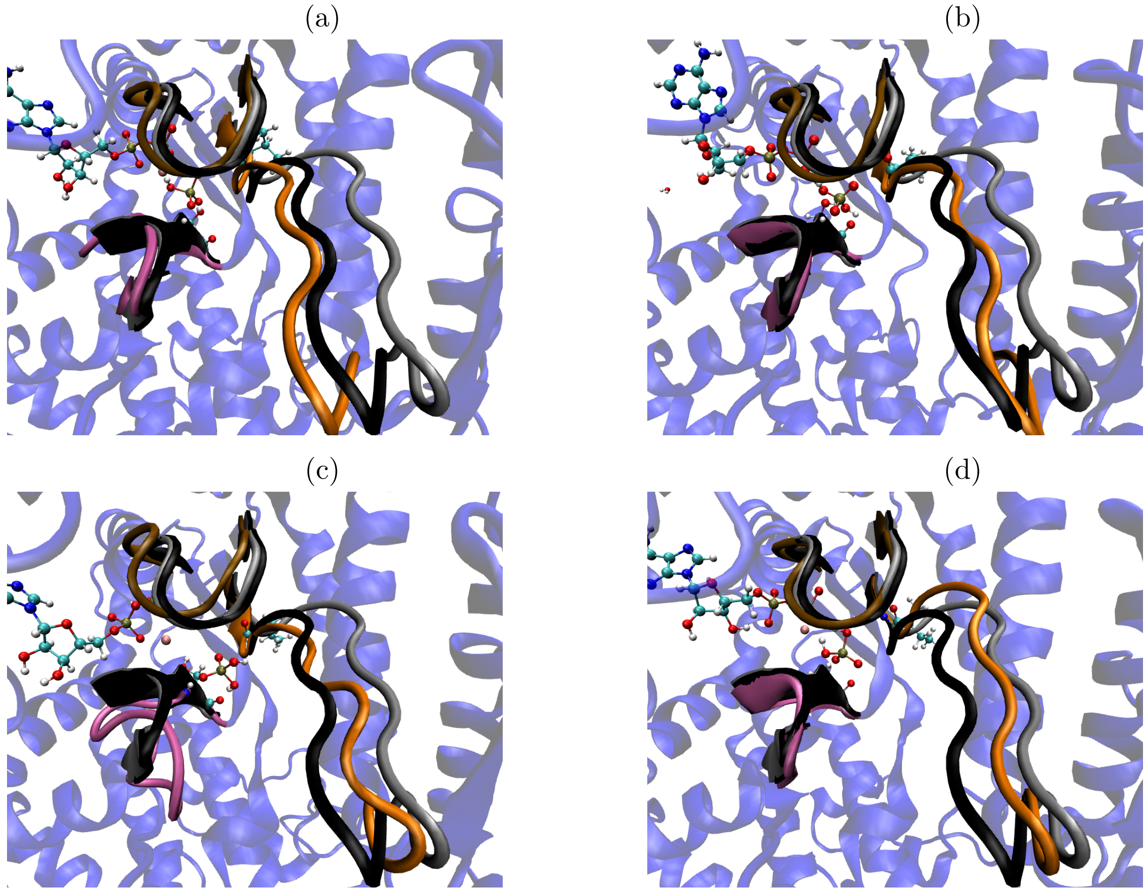


Figure F.4: Coordination state of $Mg^{2+}$ and location of selected water molecules (black circled) chose to be brought in the coordination of $Mg^{2+}$ for PPS and $P_iR$. (a) PPS 3rd water molecule, (b) PPS 4th water molecule, (c) $P_iR$ 3rd water molecule, (d) $P_iR$ 4th water molecule.

**Impact on the active site conformation** For the simulation starting with PPS, while the first brought water molecule had not much impact on the NBL interactions, the second one destabilised it completely. Notably the interactions between ARG199 and TYR462 and between SER153 and PHE460 have broken.

For the simulation starting with $P_iR$, the way the water molecules were brought in the coordination of $Mg^{2+}$had not much impact on the NBL interactions.

The potential of Mean Force obtained after forcing the departure from the structures obtained after bringing the water molecules are represented on Fig. **??**.



(a) (b) (c) (d)

Figure F.5: Representation of nucleotide binding loops states in the relaxed structures. In orange is represented the Switch II, Switch I in magenta, P-Loop in brown. Are also represented the crystallographic states of the NBL from PPS (black) and P$_i$R (gray). (a) PPS-R, (b) P$_i$R-R, (c) PPS-4W-R, (d) P$_i$R-4W-R.