

**Molecular Dynamics Studies on the Essential  
Mitotic Protein Kinase – Aurora B**

by

Oda Selvåg Hovet

**Thesis**

for the degree of

**Master of Science**



Department of Biosciences

Faculty of Mathematics and Natural Sciences

University of Oslo

June 2020

This master's thesis is submitted under the master's program *Bioscience*, with program option *Molecular Biology and Biochemistry*, at the Department of Biosciences, University of Oslo. The scope of the thesis is 60 credits.

© Oda Selvåg Hovet, 2020

[www.duo.uio.no](http://www.duo.uio.no)

Print production: Representralen, University of Oslo

# Acknowledgements

The work in this thesis was conducted in a joint collaboration between the Cascella group at Hylleraas Centre for Quantum Molecular Sciences and the Sekulic group at Centre for Molecular Medicine Norway (NCMM).

Without some people, this thesis would not have been possible. I first want to express my gratitude towards my supervisors associate professor Nikolina Sekulic and professor Michele Cascella, and to researcher at the Sekulic lab, Dr. Dario Segura-Pena, who contributed the complementary experimental analysis with HDX and kinetic studies. Thank you all for your great support, guidance and meaningful discussions. I have learned a lot about scientific processes and in great part thanks to you. Thank you Michele for showing me the world of molecular dynamics, I believe I have gained a higher understanding of biochemistry because of it.

Working from home during the Covid-19 outbreak has been a challenging experience, but I believe it in some ways has taught me a lot about discipline and communication. I would like to thank my friends and family, especially my parents for their unconditional support and love in my years as a student in molecular biology and biochemistry. I would also like to express my gratitude towards my samboer Nicolai for his endless support throughout my degree. Your background in physics and informatics has also helped me achieve a better understanding for, and become more interested in, multiple areas of science I never would have otherwise.

Oda Selvåg Hovet  
Oslo, June 2020

# Abstract

Aurora B is a serine/threonine protein kinase. Its enzymatic activity is required for accurate chromosome segregation, spindle midzone stability and cell division. Aurora B kinase is always in association with the activator protein, INCENP, and full activity requires auto-phosphorylation of both, a threonine in the activation loop of Aurora B and the serines in the TSS-motif (Threonine-Serine-Serine) present on INCENP. Structural studies have revealed profound conformational changes in the enzyme complex upon auto-phosphorylation and initial hydrogen-deuterium exchange studies indicate structural dynamic changes associated with auto-phosphorylation. In this study, we have used molecular dynamics to simulate three phosphorylation states of the enzyme complex (non-phosphorylated enzyme complex, enzyme complex phosphorylated only in the activation loop of Aurora B and enzyme complex phosphorylated in both loop of Aurora B and INCENP). We found that phosphorylation has a profound impact on the structure and dynamics of Aurora B/INCENP-complex inducing allosteric communication between INCENP and Aurora B. Upon phosphorylation of the threonine in the activation loop of Aurora B, INCENP adopts a dramatic conformational change and comes into close contact with the activation loop region. Other analyzed regions undergoing noticeable structural changes upon phosphorylation are the active site where ATP becomes more stably bound, the activation loop that becomes more rigid and the  $\alpha$ G-helix experiencing a shift towards the core of Aurora B. Principle component analysis revealed a synchronized motion in the fully phosphorylated state of the enzyme complex with coupled open-close and activation loop motions, a feature that might contribute to increased activity. The insights into regulation of Aurora B/INCENP-complex provided in our study might help towards generating a more specific inhibitor of Aurora B in future cancer treatment.

# Contents

<b>Acknowledgements</b>	<b>i</b>
<b>Abstract</b>	<b>ii</b>
<b>1 Introduction and Objective of the Study</b>	<b>1</b>
1.1 Protein Kinases . . . . .	1
1.1.1 Classification of Protein Kinases . . . . .	2
1.1.2 Regulation of Protein Kinases . . . . .	4
1.1.3 Aurora Kinase Family . . . . .	5
1.2 Objective of the Study . . . . .	12
<b>2 Theoretical Background and Computational Approach</b>	<b>14</b>
2.1 Molecular Dynamics Simulations . . . . .	14
2.1.1 Basic Principles of Molecular Dynamics . . . . .	15
2.1.2 Integrating Newton's Equations of Motion . . . . .	15
2.1.3 Potential Energy Function . . . . .	16
2.1.4 Force Fields . . . . .	19
2.1.5 Initial Conditions . . . . .	19
2.1.6 Periodic Boundary Conditions . . . . .	20
2.1.7 Solvent Treatment . . . . .	20
2.1.8 Limitations of Molecular Dynamics . . . . .	21
2.1.9 Analysis of Molecular Dynamics Trajectories . . . . .	21
2.2 Computational Approach . . . . .	23
2.2.1 System Setup . . . . .	23
2.2.2 Simulation Parameters . . . . .	24
<b>3 Results and Discussion</b>	<b>26</b>
3.1 Properties of Conformation and Fluctuation . . . . .	27
3.2 INCENP Drastically Changes Conformation Upon Aurora B <sup>Thr248</sup> Phosphorylation . . . . .	30
3.3 PCA Reveals Global Kinase Motions of Aurora B . . . . .	32
3.4 Effect of Phosphorylation on Regions of the Complex . . . . .	34
3.4.1 Active Site . . . . .	35
3.4.2 Conformation of the Activation Loop . . . . .	45
3.4.3 $\alpha$ G-Helix . . . . .	54
3.5 Comparison of fully Phosphorylated Aurora C/INCENP and AURKB <sup>full-P</sup>	59

<b>4 Conclusion &amp; Future Research</b>	<b>63</b>
<b>A Abbreviations and Units</b>	<b>66</b>
A.1 List of Abbreviations . . . . .	66
A.2 Units . . . . .	67
A.3 Overview of Amino Acids . . . . .	68
<b>B GROMACS Commands</b>	<b>69</b>
B.1 Overview of GROMACS Commands . . . . .	69
<b>C Supplemental Figures</b>	<b>71</b>
C.1 Contact Maps . . . . .	71
<b>References</b>	<b>72</b>

# List of Figures

1.1.1	The final steps of the EGF pathway. Ras-GTP activates Raf which activates MEK which again activates MAPK. MAPK then phosphorylates TFs to initiate gene expression promoting cell growth. Retrieved from Figure 4.5 in [5, p.80]. . . . .	2
1.1.2	The phylogenetic human kinome tree shown with all groups except Atypical, RGC and PKL. Aurora B kinase is highlighted in the Other group (grey) of kinases. Modified from [7]. . . . .	4
1.1.3	Structural organization of CPC and its components in the localization module (PDB ID: 2QFA [26]) and catalytic module (PDB ID: 2BFX [25]). INCENP encompasses more than 800 residues and its N- and C-terminal regions bind the localization and catalytic modules, respectively. Retrieved from Figure 3B in [24]. . . . .	7
1.1.4	Structure of <i>X. laevis</i> Aurora B/INCENP-complex with highlighted structural regions (PDB entry: 4C2W [27]). . . . .	8
1.1.5	Alignment of INCENP with secondary structures from a variety of organisms (Hs – <i>Homo sapiens</i> , Ss – <i>Sus scrofa</i> , Mm – <i>Mus musculus</i> , Xl – <i>Xenopus laevis</i> , Ce – <i>Caenorhabditis elegans</i> , Dm – <i>Drosophila melanogaster</i> ). The secondary structures (from PDB entry: 4C2W [27]) are colored according to structural regions in Figure 1.1.4. Substrate residues Ser849 and Ser850 are highlighted in yellow triangles. . . . .	8
1.1.6	Alignment of Aurora B with secondary structures from a variety of organisms (Hs – <i>Homo sapiens</i> , Ss – <i>Sus scrofa</i> , Mm – <i>Mus musculus</i> , Xl – <i>Xenopus laevis</i> , Ce – <i>Caenorhabditis elegans</i> , Dm – <i>Drosophila melanogaster</i> ). The secondary structures (from PDB entry: 4C2W [27]) are colored according to structural regions in Figure 1.1.4. Substrate residue Thr248 is highlighted in a yellow triangle. . . . .	9
1.2.1	The three states of phosphorylation of Aurora B. AURKB <sup>no-P</sup> , AURKB <sup>loop-P</sup> and AURKB <sup>full-P</sup> represent basally active, semi active and fully active enzymes, respectively. . . . .	13
2.1.1	Illustrated bonded and non-bonded potential energy terms. Retrieved from [42]. . . . .	17

2.1.2	The Lennard-Jones potential model (blue curve). The repulsive force, due to quantum mechanical effects, gives a potential proportional to $(1/r)^{12}$ (dotted orange line), and the attractive force, due to a dipole-dipole interaction, gives a potential proportional to $(1/r)^6$ (dotted green line). Here, $\sigma$ is a characteristic length and $\epsilon$ a characteristic energy. . . . .	18
2.1.3	Periodic boundary conditions with the center box in the middle with 8 periodic images surrounding it. When an atom moves out of the box, another moves in to replace it. Figure retrieved from [49]. . . .	20
2.2.1	System of AURKB and INCENP in the initial periodic box after solvation of TIP3P water molecules and adding salt to a concentration of 0.15 M. Sodium and chloride ions are red and green, respectively.	24
2.2.2	Benchmark of equilibration run on daily performance (ns/day) versus the number of nodes. Simulations were run on 6 nodes which is where the curve slightly starts to flatten. . . . .	25
3.1.1	Plot showing changes in RMSD of $C_\alpha$ -atoms with respect to the equilibrated conformation of Aurora B <sup>89-350</sup> (a) and INCENP <sup>798-839</sup> (b) during the course of MD simulation (2.8 $\mu$ s). The plots are made from full simulation trajectories of all states; AURKB <sup>no-P</sup> (green), AURKB <sup>loop-P</sup> (blue) and AURKB <sup>full-P</sup> (magenta). . . . .	28
3.1.2	RMSF on $C_\alpha$ of all residues of Aurora B (a) and INCENP (b) computed for the last $\mu$ s of simulation. The states are AURKB <sup>no-P</sup> (green), AURKB <sup>loop-P</sup> (blue) and AURKB <sup>full-P</sup> (magenta). Also to be noticed is the higher values on the y-axis in INCENP (b) compared to Aurora B (a). . . . .	29
3.2.1	Structures of Aurora B/INCENP and Aurora C/INCENP where INCENP is yellow in all structures. Left (white area): Crystal starting structure Aurora B/INCENP (grey/yellow), middle (grey area): AURKB <sup>no-P</sup> (green/yellow), AURKB <sup>loop-P</sup> (blue/yellow), AURKB <sup>full-P</sup> (magenta/yellow) and right (white area): AuroraC/INCENP (pink/yellow). The TSS-motif (shown with * if phosphorylated) is pointed out for all structures except starting structure (empty circles). Phosphorylation (*) in loop and/or INCENP is stated at the bottom for each structure. . . . .	31
3.3.1	The eigenvalue (nm <sup>2</sup> ) plotted against eigenvector index. The most significant motion is contained in the first eigenvector with high eigenvalues. . . . .	32
3.3.2	The two most significant modes with corresponding motions with arrows are shown for each state. (a): For mode 1, AURKB <sup>no-P</sup> has a kinase-like motion, AURKB <sup>loop-P</sup> has a loop motion and AURKB <sup>full-P</sup> has both kinase-like and loop motions. (b): For mode 2, AURKB <sup>no-P</sup> and AURKB <sup>full-P</sup> have twist-like motions, while AURKB <sup>loop-P</sup> has the kinase-like motion. . . . .	33



3.4.1	(a): RMSF plot for all three phosphorylation forms of Aurora B (modified from original RMSF plot in Figure 3.1.2). Regions with higher fluctuations are encircled with different colors. (b): Corresponding encircled regions from RMSF shown on final conformation of AURKB <sup>loop-P</sup> following the same color-scheme as in (a). . . . .	35
3.4.2	Conformation of the P-loop (black, residues 102-105) in the aligned three states at end of simulations. The states are AURKB <sup>no-P</sup> (green), AURKB <sup>loop-P</sup> (blue) and AURKB <sup>full-P</sup> (magenta). The P-loop is more unstructured in AURKB <sup>no-P</sup> . . . . .	36
3.4.3	Important hydrogen bonds formed in the active site between ATP and surrounding residues for all states at the end of simulations. Mg <sup>2+</sup> and Na <sup>+</sup> are depicted as green and purple balls, respectively. . . . .	37
3.4.4	ATP forms hydrogen bonds to surrounding water molecules in the active site in all states at the end of simulations. . . . .	37
3.4.5	Top (a, b, c): ATP in the active site aligned for initial (light grey) and end conformation (AURKB <sup>no-P</sup> (green), AURKB <sup>loop-P</sup> (blue) and AURKB <sup>full-P</sup> (magenta)). Bottom (d, e, f): ATP-protein interaction map for each state (AURKB <sup>no-P</sup> (left), AURKB <sup>loop-P</sup> (middle) and AURKB <sup>full-P</sup> (right)). Hydrogen bonds are represented with green dashed lines and residues involved in hydrophobic interactions are represented as red half circles. . . . .	38
3.4.6	Distance between Aurora B <sup>Glu141</sup> and Aurora B <sup>Lys122</sup> measured in the final structure of each of the enzyme complex phosphorylation states (AURKB <sup>no-P</sup> (green), AURKB <sup>loop-P</sup> (blue) and AURKB <sup>full-P</sup> (magenta)). Green spheres are Mg <sup>2+</sup> ions and purple spheres are Na <sup>+</sup> ions. Distances in AURKB <sup>full-P</sup> around Na <sup>+</sup> are 2.5, 2.5, 2.6 and 3.1 Å. . . . .	41
3.4.7	Histogram showing probabilities of distances between Glu141:Cδ–Lys122:Nζ for each state (AURKB <sup>no-P</sup> (green), AURKB <sup>loop-P</sup> (blue) and AURKB <sup>full-P</sup> (magenta)). Distances below 4.5 Å are considered to be a hydrogen bond between the ion-pair. . . . .	42
3.4.8	The histograms of probabilities of distances between Glu141:Cδ–Lys122:Nζ AURKB <sup>loop-P</sup> (a) and AURKB <sup>full-P</sup> (b) fitted to a sum of two Gaussian distributions. The black stairs represent the original histograms (Figure 3.4.7), the magenta curve is the sum of two Gaussian distributions for State 1 and 2 represented in blue and orange, respectively. . . . .	43
3.4.9	The defined activation segment. (a): Partially phosphorylated Aurora B with activation segment shown in red at end of simulations. (b): The nomenclature of the activation segment from conserved DFG to APE motifs. Modified from Figure 1 in [9]. . . . .	45
3.4.10	Structures of the states at 2.8 μs with loop regions shown in detail with interactions formed by surrounding residues for all states. Aurora B and INCENP are green and yellow, blue and cyan and magenta and pink in AURKB <sup>no-P</sup> , AURKB <sup>loop-P</sup> and AURKB <sup>full-P</sup> , respectively. . . . .	47

3.4.11	Structures and interactions of the activation loop region around Aurora B <sup>Thr248</sup> shown in all states in stick figure representation at the end of simulations: AURKB <sup>no-P</sup> (green), AURKB <sup>loop-P</sup> (Aurora B blue and INCENP yellow) and AURKB <sup>full-P</sup> (Aurora B magenta and INCENP yellow). Distances of hydrogen bonds in AURKB <sup>loop-P</sup> and AURKB <sup>full-P</sup> are depicted. . . . .	48
3.4.12	Interactions between Aurora B <sup>TPO248</sup> – Aurora B <sup>Arg140</sup> and Aurora B <sup>TPO248</sup> – INCENP <sup>Arg843</sup> in AURKB <sup>loop-P</sup> with formed hydrogen bonds and respective distances. INCENP is represented in yellow. Distances are shown in Å. . . . .	49
3.4.13	Monitored hydrogen bonds in red and black with corresponding colors in plot of the distance (Å) over time (ns) in AURKB <sup>loop-P</sup> . Also to be noticed are the different values on the y-axis (Å). (a): Hydrogen bonds between Aurora B <sup>TPO248</sup> – Aurora B <sup>Arg140</sup> and (b): Hydrogen bonds between Aurora B <sup>TPO248</sup> – INCENP <sup>Arg843</sup> . . . . .	50
3.4.14	Interactions between Aurora B <sup>TPO248</sup> – INCENP <sup>Arg847</sup> , INCENP <sup>Arg843</sup> – INCENP <sup>SPO849</sup> , INCENP <sup>Lys846</sup> – INCENP <sup>SPO849</sup> and Aurora B <sup>Arg245</sup> – INCENP <sup>His854</sup> in AURKB <sup>full-P</sup> with formed hydrogen bonds and respective distances. INCENP is represented in yellow. Distances are shown in Å. . . . .	51
3.4.15	Monitored hydrogen bonds in red and black with corresponding colors in plot of the distance (Å) over time (ns) in AURKB <sup>full-P</sup> . Also to be noticed are the different values on the y-axis (Å). (a): Hydrogen bonds between Aurora B <sup>TPO248</sup> – INCENP <sup>Arg847</sup> , (b): Hydrogen bonds between INCENP <sup>Arg843</sup> – INCENP <sup>SPO849</sup> , (c): Hydrogen bonds between INCENP <sup>Lys846</sup> – INCENP <sup>SPO849</sup> and (d): Hydrogen bond between Aurora B <sup>Arg245</sup> – INCENP <sup>His854</sup> . . . . .	52
3.4.16	Aligned structures at end of simulations of all states: AURKB <sup>no-P</sup> (green), AURKB <sup>loop-P</sup> (blue) and AURKB <sup>full-P</sup> (magenta). The lighter shade represents their respective bound INCENP. The αG-helix (residues 293-304) is encircled in red. . . . .	54
3.4.17	AURKB <sup>full-P</sup> is more structured than AURKB <sup>no-P</sup> . (a): The difference in deuteration between AURKB <sup>full-P</sup> and AURKB <sup>no-P</sup> are plotted in ribbon diagrams representing the crystal structure used as starting point for the simulation studies. The regions colored in blue represent parts of the protein where HDX is faster for AURKB <sup>no-P</sup> and slower for AURKB <sup>full-P</sup> . The darker blue color represents larger difference in deuterium uptake. The αG-helix is enclosed by the green ellipse. (b): Deuteration uptake plot for a specific peptide of Aurora B covering most of the αG-helix region (residues 298-305). The red trace represents the uptake for AURKB <sup>full-P</sup> and blue trace for AURKB <sup>no-P</sup> . . . . .	55

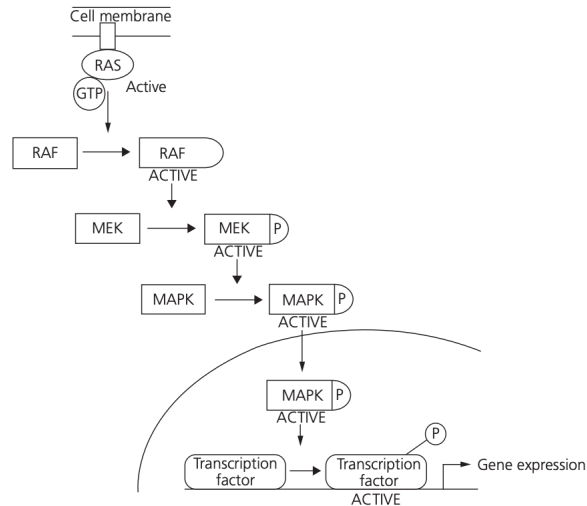
3.4.18	The $\alpha$ G-helix in the final structures (AURKB <sup>no-P</sup> (green), AURKB <sup>loop-P</sup> (blue) and AURKB <sup>full-P</sup> (magenta)) are aligned and arrows are drawn through the center of each helix. Helix is shown in front, back and side views to illustrate helix displacement. Notice bigger displacement between unphosphorylated forms of the enzyme in comparison with the phosphorylated enzyme complex. . . . .	56
3.4.19	Hydrogen bond stabilizing the $\alpha$ G-helix in the phosphorylated forms of the enzyme. The helix for AURKB <sup>full-P</sup> and AURKB <sup>no-P</sup> are showed in magenta and green, respectively. The hydrogen bonds showed are only present in the phosphorylated form of the enzyme, probably contributing towards helix stability observed in HDX experiments. . . . .	57
3.4.20	(a): Globally aligned proteins of AURKB <sup>loop-P</sup> and AURKB <sup>no-P</sup> with residues located around 4 Å of the $\alpha$ G-helix included. On the right, the $\alpha$ G-helix is further focused in the frame. (b): Difference contact map for the region in (a) with encircled region of gained and lost contacts visualized in the structures on the right. Yellow color indicates that AURKB <sup>loop-P</sup> has gained contacts, purple color that AURKB <sup>no-P</sup> has gained contacts and red that neither has gained contacts. . . . .	58
3.5.1	AURKB <sup>full-P</sup> (magenta) at end of simulation globally aligned to fully phosphorylated Aurora C and INCENP (gold). The region of the C-terminal end of INCENP Aurora B and C are highlighted on the right. . . . .	60
3.5.2	AURKB <sup>full-P</sup> , with Aurora B (grey) and INCENP (green) with structural regions $\alpha$ G-helix (orange), activation loop (red) and phosphorylated residues (TPO248, SPO849 and SPO850) highlighted in yellow on the figure. The peptide modelled from PKA structure (cyan) is in sticks. The Ser/Thr (in the original structure mutated to Ala) that accepts the phosphate transfer is shown in black. . . . .	61
3.5.3	AURKB <sup>full-P</sup> (magenta) at the end of simulation aligned globally to fully phosphorylated Aurora C (gold) with encircled regions of the active site (1), activation loop region (2) and $\alpha$ G-helix (3). . . . .	62
C.1.1	35 × 35 distance matrices between nearby residues (see Table 3.4.3) for residue index in state AURKB <sup>no-P</sup> in (a), and state AURKB <sup>loop-P</sup> in (b) with truncation cut-off 1 nm. The range of white to black corresponds to shortest to longest distances between residues. . . . .	71

# 1

## Introduction and Objective of the Study

### 1.1 Protein Kinases

Post-translational modifications (PTMs) are biochemical modifications that occur on a protein after it has been translated by a ribosome. Typical PTMs include phosphorylation, glycosylation, ubiquitination, methylation, proteolysis and many more. Protein kinases (PKs) are enzymes that phosphorylate proteins by catalyzing transfer of the  $\gamma$ -phosphate of adenosine triphosphate (ATP) to the hydroxyl group of amino acid serine, threonine or tyrosine on the protein substrate (in non-animal kingdom also to histidine) [1]. Phosphorylation usually leads to changes in the target protein with consequences on its enzymatic activity, cellular localization, turnover or association with other proteins [2]. The human genome contains >500 identified PKs which make up  $\sim 2\%$  of all human genes [3]. Protein phosphorylation is one of the most studied PTMs and it affects more than 30% of all human proteins. Phosphorylation is involved in regulating the majority of cellular pathways, especially those involved in signal transduction. In addition, phosphorylation has important roles in growth, metabolism, cell differentiation, cytoskeletal rearrangement, cell movement, apoptosis and membrane transport [4]. Many kinases interact in cascade pathways where one kinase phosphorylates the next and so forth downstream in order to initiate a cellular response. The epidermal growth factor (EGF) pathway is an example of such a pathway: a substrate kinase, Raf, is phosphorylated by membrane-bound Ras-GTP and becomes active in order to phosphorylate MEK which again phosphorylates MAPK. MAPK enters the nucleus and phosphorylates transcription factors (TFs) which turn on different sets of genes for cell growth [5, p.78-81]. In such a way, kinases are able to bring forward important physiological responses.



**Figure 1.1.1:** The final steps of the EGF pathway. Ras-GTP activates Raf which activates MEK which again activates MAPK. MAPK then phosphorylates TFs to initiate gene expression promoting cell growth. Retrieved from Figure 4.5 in [5, p.80].

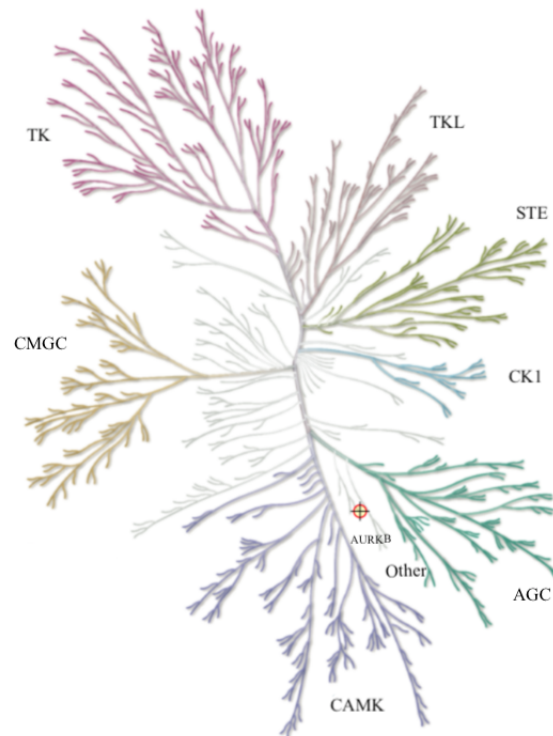
### 1.1.1 Classification of Protein Kinases

PKs represent one of the largest protein families corresponding to approximately 2% of the eukaryotic genome, also referred to as the kinome [4]. There are more than 518 human kinases that can be grouped into typical (90%) and atypical (10%) groups based on conserved sequence similarity in their catalytic domain [2]. Typical PKs can further be divided into three subgroups based on the residue(s) they phosphorylate: ser/thr kinases, tyr kinases and dual-specific kinases where the latter is able to phosphorylate all three residues. A system to classify PKs was developed by Manning et. al in 2002 [3]. According to that classification, kinases are divided into 10 groups which are subdivided into multiple families and subfamilies [3]. The classification is primarily based on sequence similarity of the catalytic domain and biological functions. Groups are classified based on the type of residue becoming phosphorylated, families on the kinase's broad function and sequence similarity and subfamilies on even more similar sequence and functions. The ten groups are listed below where one comprises atypical kinases and the rest represent typical [6]:

- **AGC** – named after the kinases PKA, PKG and PKC. Cytoplasmic ser/thr kinases that consists of 16 families.
- **CMGC** – named after MAPK and Cdk kinases. Includes kinases involved in key physiological processes and consists of 9 families.

- **CAMK** – named after kinases regulated by calcium and calmodulin. Consists of 22 families.
- **CK1** – named after the original Casein Kinase 1. This group has very similar sequence and includes 4 families.
- **Other** – includes 32 families of kinases that do not fit in the other groups. Most of them have the kinase fold.
- **STE** – named after homologs of *STE*-genes in yeast and consists of three families that activate themselves before activating kinases in the MAPK family.
- **Tyrosine Kinase (TK)** – phosphorylates mainly tyrosine residues. Includes 30 families where most of them are membrane receptors.
- **Tyrosine Kinase-Like (TKL)** – similar to TK-group but phosphorylates serine and threonine instead of tyrosine.
- **Receptor Guanylate Cyclase (RGC)** – possesses an active guanylate cyclase domain and makes cGMP. They are present in the plasma membrane. There are no defined families, but many classes.
- **Protein Kinase-Like (PKL)** – consists of families with the protein kinase-like fold and catalytic mechanism with very different sequences. Includes 14 families.
- **Atypical** - split into three structural categories instead of families. All kinases in this group lack sequence similarity to the catalytic domain, but have shown kinase activity experimentally.

The 10 groups comprise all known kinases and a corresponding phylogenetic tree highlights most of them in Figure 1.1.2. All groups are included except from RGC, PKL and Atypical which do not share clear sequence similarities with the other groups. However, they still function as kinases transferring phosphate groups from ATP to serine, threonine or tyrosine on the target protein. Although PKs are diverse, most of them share sequential and structural similarities such as the typical kinase-fold important for function.



**Figure 1.1.2:** The phylogenetic human kinome tree shown with all groups except *Atypical*, *RGC* and *PKL*. *Aurora B* kinase is highlighted in the *Other* group (grey) of kinases. Modified from [7].

### 1.1.2 Regulation of Protein Kinases

PKs possess a broad range of functions affecting almost every cellular process from cell cycle regulation to DNA damage response. In order to control these necessary processes properly, kinases must be tightly regulated [8]. Kinases are prone to a range of intricate regulation mechanisms controlling their catalytic activity and substrate specificity. Some examples include transcriptional regulation, regulation at the level of protein degradation, inhibition or activation of their protein partners, phosphorylation (by other kinases or auto-phosphorylation) and dephosphorylation (by phosphatases) and changes in cell localization that limit presence of specific substrates and/or activators [2]. In terms of cellular localization, PKs are located on the cell surface as transmembrane receptors, inside the cell as intracellular transducers or inside the nucleus [5, p.80-83]. All PKs usually exist in a basal state and become activated when needed for activity. They can therefore act as molecular switches adopting conformations characterizing states of activation. The two most extreme conformations are *on* and *off*, describing maximal or minimal activity of the kinase, respectively. Active, *on* conformations, are characterized as structurally similar in kinases where the substrates and catalytic domain become properly aligned for  $\gamma$ -phosphate transfer. *Off* conformations, on the other hand,

are diverse due to less chemical constraints applied [9]. The first kinase structure solved, in 1991, was active and phosphorylated PKA, where it was shown that a phosphorylated residue in the activation segment interacts with a pocket of positively charged residues on the surface of the kinase [10]. On the other hand, the earliest inactive structures such as CDK2 showed that an unphosphorylated activation loop can adopt a range of conformations distinct from those found in active PKA [11][2].

Even though kinases are tightly regulated, they are still prone to failed regulation mechanisms. Deregulation of kinase activity is associated with many diseases such as cancer, inflammatory diseases, cardiovascular disorders and many more [12]. PKs thus represent important drug targets for treating human disorders through selective inhibitors. However, design of small molecule drugs targeting particular PKs is very challenging due to similar catalytic domains between many PKs. Today, there are about 52 small-molecule PK inhibitors with 46 of these related to cancer treatment [13]. An example of a successful PK inhibitor is Imatinib. Imatinib is a tyrosine kinase PK inhibitor approved for use in patients in 2001 [14]. The inhibitor interacts with tyrosine kinase ABL whose activity is uncontrolled in patients with chronic myelogenous leukemia (CML). Tumor cells of those patients have reciprocal translocation between chromosome 9 and chromosome 22 generating the so called “Philadelphia chromosome” with fusion gene *BCR-ABL1* that escapes normal physiological control. Imatinib is a small molecule that binds in the ATP-binding site of ABL enzyme in the inactive state of the kinase [5, p.98-100]. Currently, there are 37 kinase inhibitors FDA approved for treatment of malignancies and research on protein kinases which accounts for a quarter of all current developmental cancer drug efforts [15]. A major challenge with targeting protein kinases is assuring specificity (and thus decreasing side effects). Many kinase inhibitors are small molecules mainly targeting the ATP binding site of the kinase. However, with more than 500 kinases (with similar ATP binding sites) identified in the human body, it is impossible to avoid cross-reactivity. Because of this, development of allosteric kinase inhibitors has a great potential in cancer therapy. For example, Trametinib, an allosteric inhibitor of MEK1 and 2, was FDA approved in 2013. Allosteric drugs are valuable both in limiting off-target effects through high protein specificity, but also in overcoming drug resistance. Malignant cells often already have or develop mutations in the active site after treatments with active site inhibitors. Combinatorial therapy using active site and allosteric inhibitors have therefore proven to be more efficient than either alone [16].

### 1.1.3 Aurora Kinase Family

The Aurora kinases are a family of Serine/Threonine protein kinases categorized as “Other” in classification of the kinome due to their low sequence homology to other categories (Figure 1.1.2). Aurora kinase genes were discovered in late 1980s as part of a search for *Drosophila* genes regulating cell cycle progression [17]. Discovered genes were given names related to polar regions (Aurora, Borealin, Polo

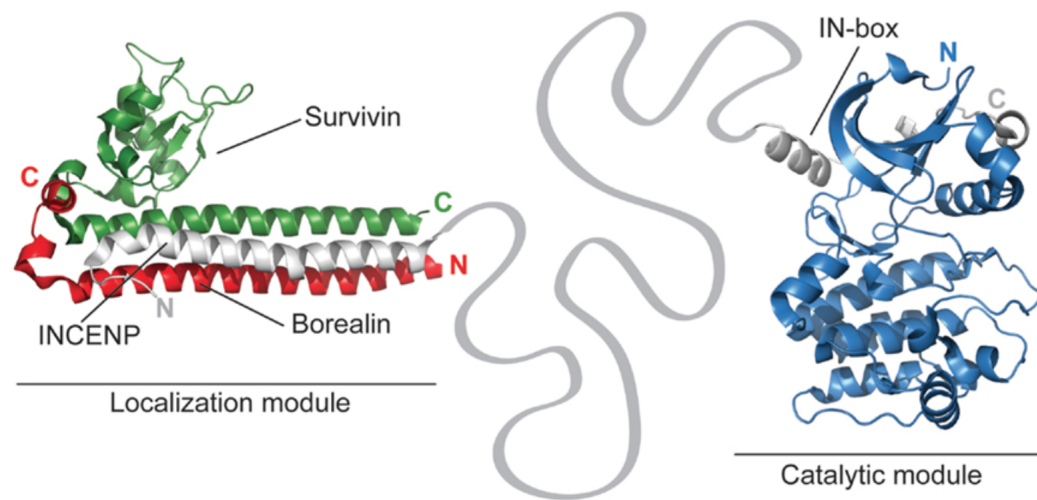


etc.), supposedly because defective mitotic spindle in the mutant bears resemblance to the polar lights. Soon after, Aurora paralogs were identified in different species [18]. In humans, the Aurora family is comprised of three members: Aurora kinase A, B and C (AURKA, AURKB and AURKC) that are all critical for the regulation of multiple steps of the mitotic program. Aurora A is essential for mitotic spindle assembly and centrosome maturation [19], whereas Aurora B kinase activity is required for accurate chromosome segregation, spindle midzone stability and cell division [17]. The B and C isoforms are 76 % identical and differ primarily in regards to tissue localization. Aurora B is expressed in somatic cells whereas Aurora C is expressed in the germ cells. Aurora C is the least studied in the Aurora family and has also been shown to have overlapping functions with Aurora B in the cell cycle [20]. Aurora A is little bit more distant with 71% sequence identity to Aurora B [21][22].

The activity of Aurora kinases is regulated through (auto)phosphorylation and interaction with binding partner proteins. TPX2 (Targeting Protein for Xklp2) is a binding partner for Aurora A and INCENP (INner CENTromere Protein) is a binding partner for Aurora B and C. Despite of the high level of similarity in sequence and structure, Aurora A and B are regulated differently in respect to auto-phosphorylation and interaction with unique binding partner proteins. Aurora A is enzymatically active when at least one of the following requirements is fulfilled: either binding of TPX2 or phosphorylation of the activation loop. In difference to this, Aurora B activity is inseparable from INCENP binding. In addition to phosphorylation of the Aurora B auto-activation loop the binding partner INCENP also needs to be phosphorylated for complete Aurora B kinase auto-activation [20].

### **Physiological Importance of Aurora Kinase B**

Aurora B is a constant component of a larger molecular complex - the Chromosomal Passenger Complex (CPC). The CPC consists of a catalytic unit, Aurora B and a localization unit, composed of 3 proteins: Survivin, Borealin and INCENP [23][24]. The N-terminal region of INCENP forms a 3-helix bundle with Survivin and Borealin, while its C-terminal region forms a crown that surrounds the N-terminal domain of Aurora B through hydrophobic interactions (Figure 1.1.3). INCENP is a larger protein than Aurora B containing >800 residues. The sequence between the N- and C-terminal ends is probably unstructured with low homology to other known proteins [25].

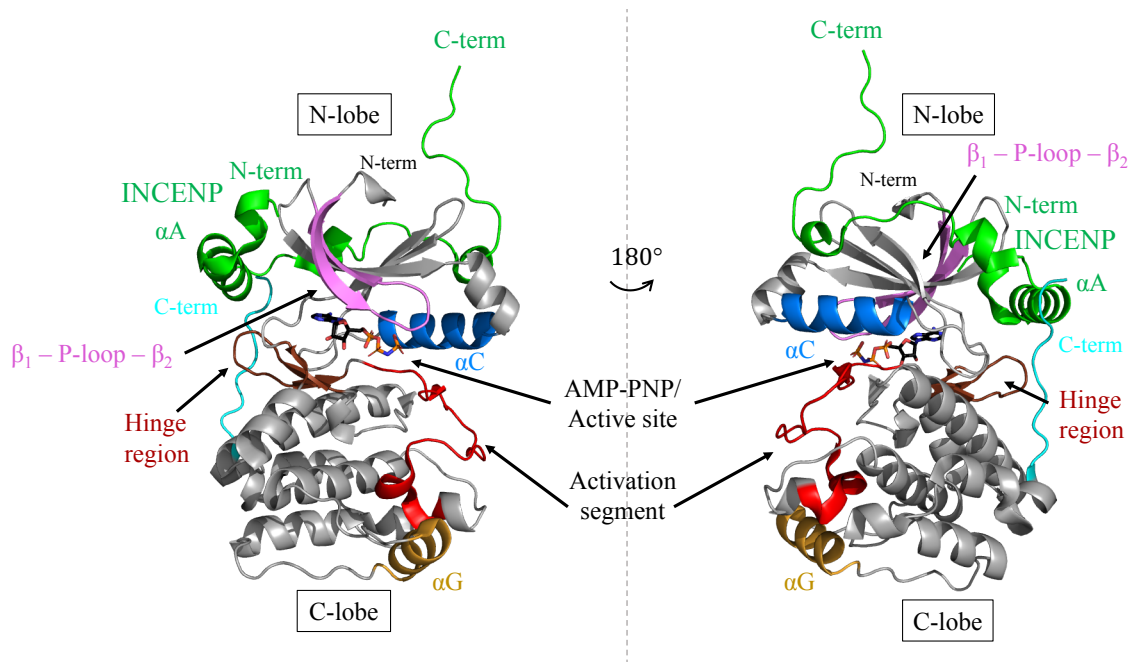


**Figure 1.1.3:** Structural organization of CPC and its components in the localization module (PDB ID: 2QFA [26]) and catalytic module (PDB ID: 2BFX [25]). INCENP encompasses more than 800 residues and its N- and C-terminal regions bind the localization and catalytic modules, respectively. Retrieved from Figure 3B in [24].

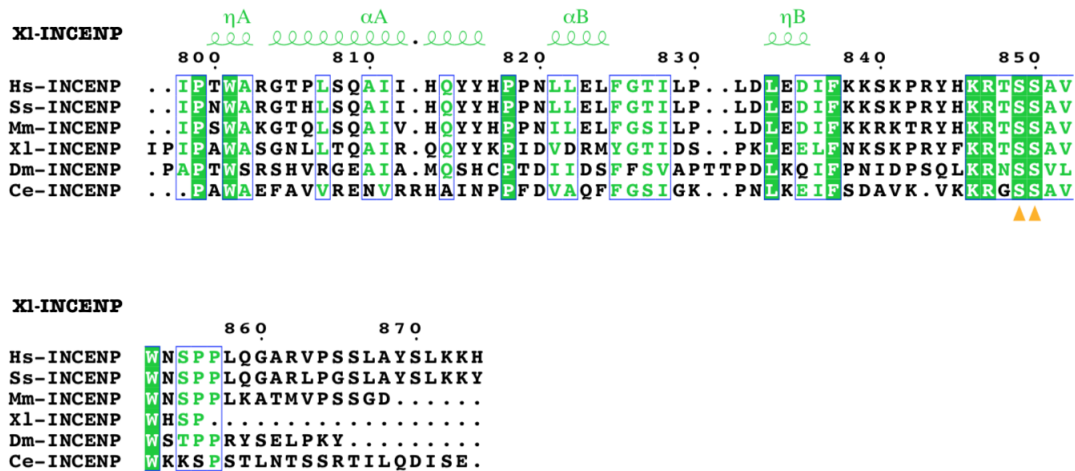
The CPC is a dynamic complex with different subcellular localization and specialized functions during the cell cycle. In the prophase, CPC resides on chromosome arms and starts accumulating at the inner centromeres, between kinetochores. In prometaphase and metaphase, the complex is concentrated at the inner centromere. The kinase activity of the centromeric pool of Aurora B is necessary for the correct kinetochore-microtubule attachment and subsequently accurate chromosome segregation. In anaphase, the complex relocates to the spindle midzone. Finally, in telophase, while cells progress to cytokinesis, CPC concentrates at the cleavage furrow between the dividing cells. In each stage of the cell cycle the Aurora B phosphorylates a range of substrate proteins in its vicinity. These include histone H3, centromeric histone CENP-A, Dam 1 complex, HEK1/Ndc80, MCAK, shugoshin, intermediate filaments and more [23] [17].

### Structure of Aurora B/INCENP-Complex

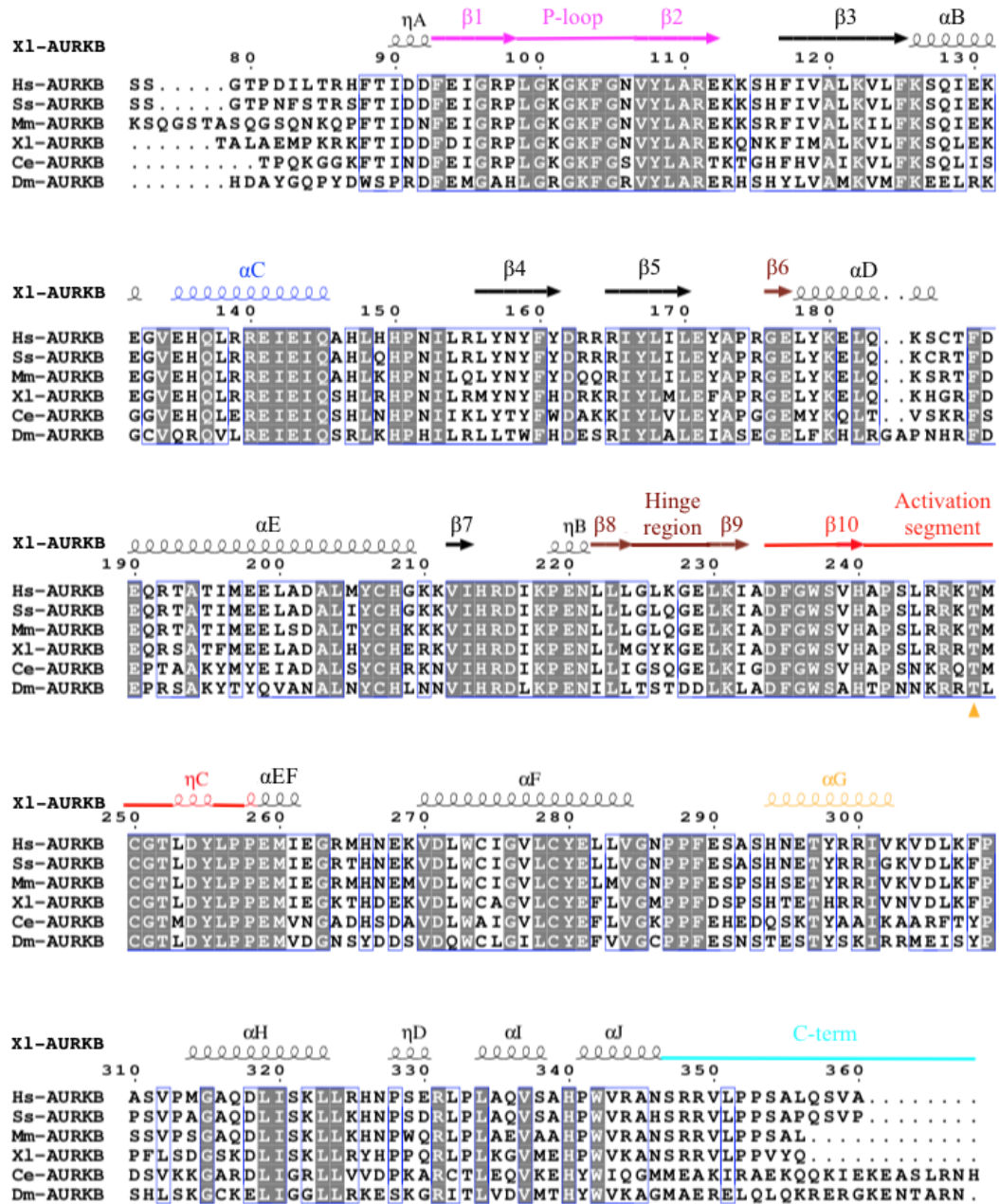
Aurora B kinase has three regions: N-terminal region (residues 1-85), catalytic region (residues 86-347) and C-terminal region (residues 348-356), where the catalytic domain makes up the largest and most conserved part [25]. The largest catalytic domain consists of two lobes: a smaller N-terminal lobe featuring a 5-stranded  $\beta$ -sheet and an  $\alpha$ C-helix and a larger C-terminal lobe consisting mostly of  $\alpha$ -helices and loops. The two lobes are connected by a flexible hinge and the active site is located between the lobes. Several evolutionary conserved parts of the enzyme (Figure 1.1.4, Figure 1.1.6 and Figure 1.1.5) are known to be important for regulation of kinase activity and for catalysis. These are:



**Figure 1.1.4:** Structure of *X. laevis* Aurora B/INCENP-complex with highlighted structural regions (PDB entry: 4C2W [27]).



**Figure 1.1.5:** Alignment of INCENP with secondary structures from a variety of organisms (Hs – *Homo sapiens*, Ss – *Sus scrofa*, Mm – *Mus musculus*, Xl – *Xenopus laevis*, Ce – *Caenorhabditis elegans*, Dm – *Drosophila melanogaster*). The secondary structures (from PDB entry: 4C2W [27]) are colored according to structural regions in Figure 1.1.4. Substrate residues Ser849 and Ser850 are highlighted in yellow triangles.



**Figure 1.1.6:** Alignment of Aurora B with secondary structures from a variety of organisms (Hs – *Homo sapiens*, Ss – *Sus scrofa*, Mm – *Mus musculus*, Xl – *Xenopus laevis*, Ce – *Caenorhabditis elegans*, Dm – *Drosophila melanogaster*). The secondary structures (from PDB entry: 4C2W [27]) are colored according to structural regions in Figure 1.1.4. Substrate residue Thr248 is highlighted in a yellow triangle.

- **Activation segment (residues 234-258)** – a stretch of around 30 residues located in the C-lobe between DFG to PPE-motif. Activation loop phosphorylation is the most common method for regulating kinase activity. Aurora B<sup>Thr248</sup> which becomes auto-phosphorylated is located in this loop.
- **DFG-motif (residues 234-236)** – a stretch of three amino acids (Aspartic acid–Phenylalanine–Glycine) located at the N-terminal part of the activation loop. The Aurora B<sup>Asp234</sup> residue binds Mg<sup>2+</sup> and stabilizes the phosphates of ATP. Analysis of other kinases has revealed that orientation of Phe in this motif correlates with kinase activity [28].
- **Residues Aurora B<sup>Lys122</sup> and Aurora B<sup>Glu141</sup>** – these two residues are conserved in many studied Ser/Thr kinases and their interaction is associated with the active form of the enzyme. Both residues are located in the N-lobe of the catalytic domain.
- **$\alpha$ C-helix (residues 135-145)** – the  $\alpha$ C-helix (catalytic helix) is located in the N-terminal lobe. In other kinases, residues from  $\alpha$ C-helix interact with both phosphates of ATP and activation loop. Aurora B<sup>Glu141</sup> is located in the this helix and helix rotation is proposed as a mechanism for regulating Aurora B/INCENP activity [25].
- **$\beta_1$ –P-loop– $\beta_2$  (residues 93-112)** – The  $\beta$ -hairpin is located in the N-lobe and forms the upper part of the ATP binding site. The classical conserved sequence of the P-loop of Aurora kinases is GKGKFG. The Gly residues form interactions with the  $\beta$ - and  $\gamma$ -phosphate groups of ATP to stabilize ATP binding by coordinating the phosphate groups through backbone interactions [4].
- **Hinge Region** – The hydrophobic hinge region connects the N and C lobes and can form hydrogen bonds to the purine ring of adenosine of ATP for stabilization in catalysis. The hinge region is important for a catalytic breathing motion of the kinase where the lobes act as rigid bodies [29].

Only C-terminal part of INCENP, also called IN-box (residues 798-840) is bound by Aurora B. Crystal structures [25] revealed that INCENP binds to the N-lobe of Aurora B through hydrophobic interactions. IN-box is mainly unstructured except for the helix,  $\alpha$ A (residues 804-816). The most conserved part in the IN-box is encompassing TSS (Threonine-Serine-Serine) (residues 848-850) motif where two serine residues are subject to auto-phosphorylation that regulates kinase activity. Interestingly, Aurora B<sup>Trp853</sup> is also fully conserved and has been reported to be important for a stable active conformation in Aurora kinase B and C, so much that the TSS-motif in fact might be redefined as TSSxxW involved in activation mechanism [20].

## Aurora B/INCENP Regulation & Activation

The Aurora B/INCENP-complex is subjected to several regulation mechanisms involving transcription, degradation and auto-phosphorylation.

Transcription of Aurora kinases depend on cell cycle-dependent element (CDE) and CDE/cell cycle gene homology region (CHR) sequences in Aurora kinase promoters. Aurora B is induced upon binding to various transcription factors, E2F-1, E2F-4, DP-2 and FoxM1, with the CDE/CHR sequences within the *Aurora B* promoter during prophase [30]. Overexpressed Aurora B is associated with impaired chromosome segregation and cell division. The overexpressed Aurora B enhances aneuploidy, genetic instability and risk of cancer. Degradation of Aurora kinases is mediated by anaphase-promoting complex which is activated by Cadherin 1 (Cdh1) protein [31]. The activated anaphase-promoting complex ubiquitinates Aurora kinases and initiates proteasome-mediated degradation. Aurora kinase B is usually degraded after cytokinesis [32].

Aurora B/INCENP-complex is regulated by auto-phosphorylation. The enzymatic complex without any phosphorylated amino acids has only minimal basal catalytic activity. Full activity requires both, phosphorylation of Aurora B<sup>Thr248</sup> in the activation loop and phosphorylation of Serines 849-850 in Threonine-Serine-Serine (TSS) motif on INCENP. Upon autophosphorylation, the activity of the enzyme complex increases more than two orders of magnitude [33]. While phosphorylation of the Thr in the activation loop is a feature common to many protein kinases, requirement for phosphorylation of INCENP is unique for Aurora B. At the beginning of this study in 2018, the molecular understanding of Aurora B auto-activation was based on a model derived from the structure of a partially phosphorylated enzyme where Aurora B<sup>Thr248</sup> in the activation loop is phosphorylated, but the TSS-motif of INCENP is missing [25]. Last year, a structure where both activation loop and INCENP are phosphorylated and structured became available [20], indicating the importance of phosphorylated TSS-motif of INCENP in stabilizing the activation loop of Aurora B. The structure contains Aurora C/INCENP-complex, that is meiosis-specific, but otherwise highly similar to Aurora B/INCENP-complex. The structure of the enzymatic complex in the dephosphorylated state is not available, probably due to the high flexibility of this form of the enzyme.

Although very informative, crystal structures represent a snapshot of the enzyme complex in the particular phosphorylation state. In solution, there is most likely assemblies of conformations representing each state. The amount of different conformations a macromolecule adopts reflects its flexibility or dynamics. To understand the (auto)-activation mechanism, it is important to obtain insight into the structural and dynamic changes linked to the phosphorylation status of the kinase. Although, there are several experimental approaches that can assess dynamics, the only manner to obtain a high resolution view of the structural changes taking place during the activation process is through a computational approach - molecular dynamics simulation. In Aurora B/INCENP-complex, INCENP is loosely

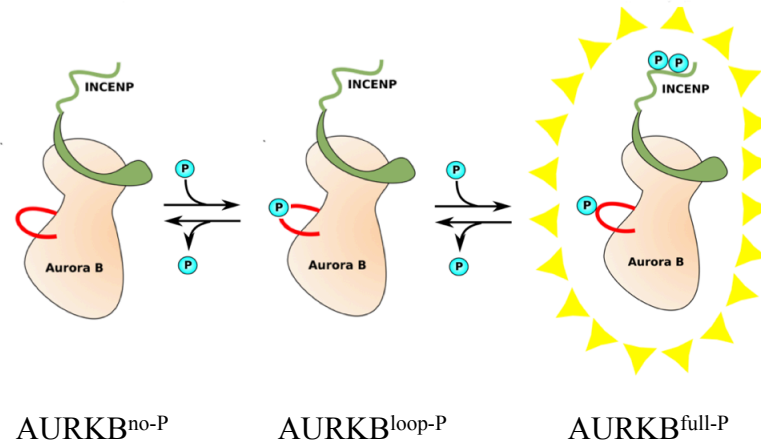
structured and it would be important to understand if the dynamics of the enzyme complex changes in response to phosphorylation. Also, it remains unclear to which extent phosphorylation at different sites contributes to enzyme activity and how the activities are differently regulated by for instance INCENP.

### **Aurora B Kinase and Cancer**

Aurora B is up-regulated and/or overexpressed in many cancer types and this is always associated with poor prognosis [34]. Due to its important regulatory role in cell division, Aurora B has been identified as an attractive target for cancer chemotherapy and several inhibitors are already in clinical trials [35]. The identified inhibitors of Aurora B show promising results in clinic but exhibit high side effects. Part of the reason is their cross-reactivity with other signaling pathways in cells, as current inhibitors are mostly targeting the ATP binding site. The development of a highly specific Aurora B drug has a potential to overcome this problem but progress have been hampered due to lack of understanding of the Aurora B's activation mechanism and oncogenic activities.

## **1.2 Objective of the Study**

The objective of the present study is to obtain a better understanding of the mechanism of auto-activation of Aurora B/INCENP-complex through phosphorylation. The aim was to investigate how different phosphorylation states of the Aurora B/INCENP-complex affect dynamics and structure important for kinase activity. To achieve this, we used molecular dynamics (MD) to simulate the enzyme complex in three phosphorylation states: unphosphorylated ( $\text{AURKB}^{\text{no-P}}$ ), partially phosphorylated ( $\text{AURKB}^{\text{loop-P}}$ ) and fully phosphorylated ( $\text{AURKB}^{\text{full-P}}$ ) structures schematically presented in Figure 1.2.1.



**Figure 1.2.1:** The three states of phosphorylation of Aurora B.  $AURKB^{no-P}$ ,  $AURKB^{loop-P}$  and  $AURKB^{full-P}$  represent basally active, semi active and fully active enzymes, respectively.

We explore conformational changes taking place in the complex as a consequence of phosphorylation and monitor parts known to be important for catalysis. The results of the computational approach are discussed and compared with the results of the complementary experimental approach, hydrogen-deuterium exchange (HDX), and kinetic analysis performed by Dr. Dario Segura, researcher in Nikolina Sekulic's group. Together, the knowledge generated from the work presented in this thesis contributes towards understanding of physiological regulation of Aurora B/INCENP and could be exploited in future studies to design Aurora B-specific (and thus less toxic) cancer treatment.



# 2

## Theoretical Background and Computational Approach

### 2.1 Molecular Dynamics Simulations

Kinases represent dynamical systems with important flexibility when switching from inactive to active states and abilities to form complex interactions with other proteins and molecules. The conformational dynamics is strongly linked to their structures and functions. In order to study the structure of a protein, a 3D-structure obtained experimentally by X-ray crystallography or nuclear magnetic resonance (NMR) can be used, but the structure is a static representation. In order to understand more about the protein's internal dynamics and how this is connected to the protein's activity, experimental information can be combined with computer modelling. Molecular dynamics (MD) is a simulation method that allows the exploration of the conformational space of a molecular system by direct integration of its equations of motion, typically in the classical Newtonian mechanics formulation. Atomistic MD simulations based on classical mechanical laws and a well-refined force field is considered one of the most accurate computational techniques for studying protein flexibility and for describing a system's dynamics over time [36].

The first MD simulation of a protein was in 1977 by McCammon et. al. The simulation was performed on bovine pancreatic trypsin inhibitor and replaced the views of proteins as rigid molecules [37]. The protein was around 500 atoms and the trajectory of the experiment was 9.2 ps. Today, computer power has increased exceptionally and larger systems containing  $10^6$  atoms or more can be simulated in a

multi ns/ $\mu s$ -timescale, which is the timescale for many biological processes including conformational transitions of most kinases [38][36]. With more powerful computers and cheaper simulation times, the popularity of MD is increasing and predictions of properties can be made more accurately. As a statistical sampling tool, MD is not just used to explore direct time-evolution behavior from pre-defined initial conditions; rather, its main scope is to predict thermodynamically equilibrated ensemble averages for any observable of interest.

### 2.1.1 Basic Principles of Molecular Dynamics

MD is a simulation method for computing the time evolution of a molecular system by numerical integration of Newton's equations of motion [39, p.63-65]. The technique generates trajectories for the interacting particles in a system with certain initial and boundary conditions while satisfying thermodynamical constraints. Following the laws of classical mechanics, one has:

$$\mathbf{F}_i = m_i \mathbf{a}_i \quad (2.1.1)$$

$$\mathbf{F}_i = m_i \frac{d^2 \mathbf{r}_i}{dt^2} \quad (2.1.2)$$

for each atom  $i$  in a system of  $N$  atoms. Here,  $m_i$ ,  $\mathbf{a}_i$  and  $\mathbf{F}_i$  represent the mass, acceleration and force acting on atom  $i$ , respectively.

In order to model the physical system, a potential energy function,  $V(\mathbf{r}_1, \dots, \mathbf{r}_N)$  of the atomic positions  $r$  must be defined. The force on atom  $i$  can be derived as the negative of the gradient of the potential energy

$$\mathbf{F}_i = -\nabla_{\mathbf{r}_i} V(\mathbf{r}_1, \dots, \mathbf{r}_N) \quad (2.1.3)$$

The sum of the potential energy,  $V$ , and the kinetic energy,  $K$ , is the total conserved energy of the system  $E = K + V$  [39, p.63-65].

### 2.1.2 Integrating Newton's Equations of Motion

Solving the Newton's equations of motion requires two independent initial conditions: positions and velocities. The forces of the many-body system depend on the positions of the particles. By discretization of the time axis in finite steps  $\Delta t$ , positions and velocities along a physical trajectory can be computed in an iterative manner. A smaller time-step is more accurate and can reduce errors, but is also more computationally expensive to accumulate long trajectories.

Historically, the prototype of modern integration algorithms used in MD is the Verlet algorithm, which is derived from Taylor expansion of the positions:

$$\begin{aligned} \mathbf{r}(t + \Delta t) &= \mathbf{r}(t) + \mathbf{v}(t)\Delta t + \frac{1}{2}\mathbf{a}(t)\Delta t^2 + \frac{1}{6}\mathbf{b}(t)\Delta t^3 + O(\Delta t^4) \\ \mathbf{r}(t - \Delta t) &= \mathbf{r}(t) - \mathbf{v}(t)\Delta t + \frac{1}{2}\mathbf{a}(t)\Delta t^2 - \frac{1}{6}\mathbf{b}(t)\Delta t^3 + O(\Delta t^4) \end{aligned}$$

Adding the two expansions gives

$$\mathbf{r}(t + \Delta t) = 2\mathbf{r}(t) - \mathbf{r}(t - \Delta t) + \mathbf{a}(t)\Delta t^2 + O(\Delta t^4) \quad (2.1.4)$$

where  $\mathbf{r}(t)$ ,  $\mathbf{v}(t)$ ,  $\mathbf{a}(t)$  are the position, velocity and acceleration in the present, respectively,  $\mathbf{r}(t \pm \Delta t)$  are the positions forward and backward in time, and  $\mathbf{b}$  is the third derivative of the position with respect to  $t$ . Errors related to the accuracy of the algorithm using the finite difference method with respect to the true solution are defined as truncation errors, which in case of the Verlet, it is of the fourth order in  $\Delta t$  [40].

An important drawback of the Verlet algorithm is that the velocities do not enter in the propagation of the equations of motion. In fact, the velocities are required for computing the kinetic energy,  $K$ , thus controlling the temperature in simulations of systems at thermodynamic equilibrium with the environment. Therefore, other Verlet-like variants and higher-order predictor-corrector algorithms have been developed. For example, the leap-frog and velocity Verlet algorithms take positions and velocities into consideration in the equations. The leap-frog algorithm will be described in the next section since it was used in the simulations in this study [40].

### The Leap-Frog Algorithm

The leap-frog algorithm is a popular algorithm for integrating the equations of motion. Unlike the Verlet algorithm, the velocities are taken into consideration in this case, but at half-integer time-steps. In other words, the velocities and positions are "leapfrogging" over each other. The leap-frog algorithm needs initial information on positions and velocities to generate the first position in the future by standard Taylor expansion. Then, by midpoint extrapolation, the velocities and accelerations at half time-step can be estimated:

$$\mathbf{v}(t + \frac{1}{2}\Delta t) = \frac{1}{\Delta t}(\mathbf{r}(t + \Delta t) - \mathbf{r}(t)) \quad (2.1.5)$$

$$\mathbf{a}(t) = \frac{1}{\Delta t}(\mathbf{v}(t + \frac{1}{2}\Delta t) - \mathbf{v}(t - \frac{1}{2}\Delta t)) \quad (2.1.6)$$

reorganizing the equations, one has:

$$\mathbf{v}(t + \frac{1}{2}\Delta t) = \mathbf{v}(t - \frac{1}{2}\Delta t) + \mathbf{a}(t)\Delta t + O(\Delta t^2) \quad (2.1.7)$$

$$\mathbf{r}(t + \Delta t) = \mathbf{r}(t) + \mathbf{v}(t + \frac{1}{2}\Delta t)\Delta t \quad (2.1.8)$$

As the velocity estimate contains a truncation error in the second order, the error for  $\mathbf{r}$  is expected to be in the third order [41, p.26-27].

### 2.1.3 Potential Energy Function

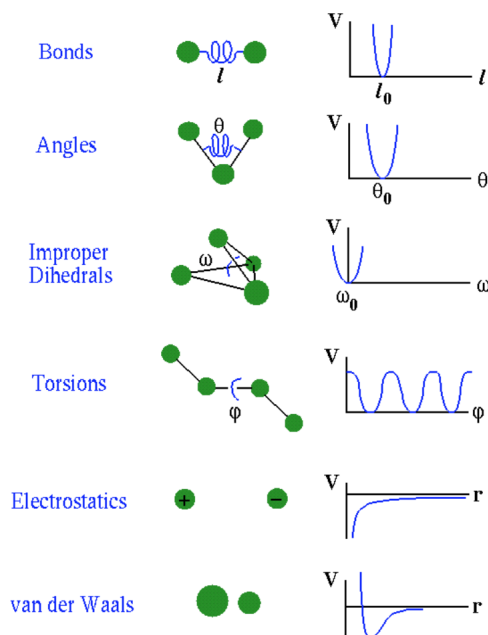
The potential energy function is, in general, a function of the atomic positions where the energy is comprised of bonded- and non-bonded terms characterizing

different interactions in the system. The potential energy function,  $V$ , can be described in the following form:

$$\begin{aligned}
 V(r) &= E_{bonded} + E_{non-bonded} \\
 &= V_{bond} + V_{angle} + V_{dihedral} + V_{improper} + V_{vdW} + V_{elec} \\
 &= \sum_{bond} k_b(b - b_0)^2 + \sum_{angle} k_\theta(\theta - \theta_0)^2 + \sum_{dihedral} \sum_{n=1}^N K_\varphi^n [1 + \cos(n\varphi - \varphi_0)] \\
 &+ \sum_{impropers} K_\omega(\omega - \omega_0)^2 + \sum_{i,j} 4\epsilon_{i,j} \left[ \left( \frac{\sigma_{i,j}}{r_{i,j}} \right)^{12} - \left( \frac{\sigma_{i,j}}{r_{i,j}} \right)^6 \right] + \sum_{i,j} \left[ \frac{q_i q_j}{4\pi\epsilon_0 r_{i,j}} \right]
 \end{aligned} \tag{2.1.9}$$

where the bonded terms are comprised of  $V_{bond}$ ,  $V_{angle}$ ,  $V_{dihedral}$  and  $V_{improper}$  and the non-bonded  $V_{vdW}$  and  $V_{elec}$ . The first term in Equation 2.1.9, bond stretching, represents the energy that is required to stretch or compress a covalent bond for a pair of atoms (1,2-pairs). The second term, angle bending, represents the energy needed to bend a bond from its equilibrium angle,  $\theta$ , for three atoms present. The third term, torsional rotation, means the sum of the rotations of four atoms in chain (1,4-pairs). This potential is represented by the dihedral angle between two planes and the energy of torsion required for the rotation about the bonds. The non-bonded terms describe inter- or intramolecular interactions between far-separated atoms such as van der Waals and electrostatic interactions.

The interaction terms above are illustrated in Figure 2.1.1.



**Figure 2.1.1:** Illustrated bonded and non-bonded potential energy terms. Retrieved from [42].

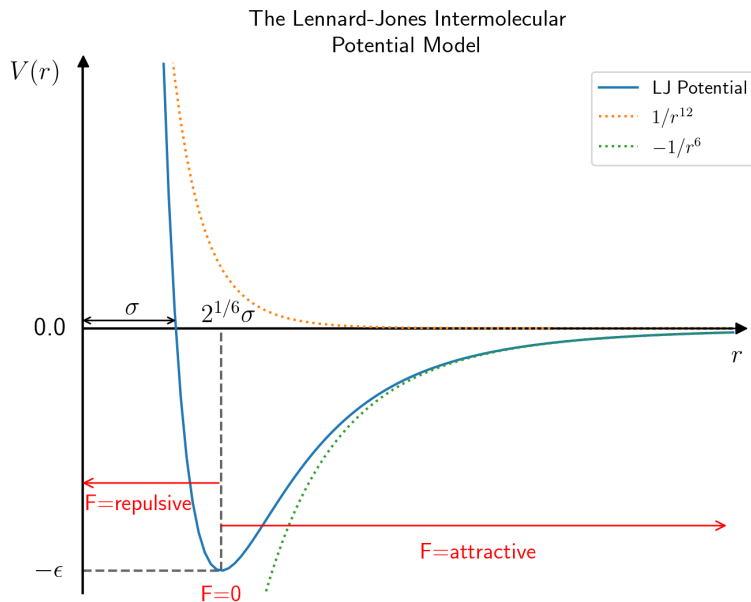
We will now explain the terms further starting with the non-bonded potential energy terms.

### Non-Bonded Interactions

The non-bonded terms for intermolecular forces are modelled by the Lennard-Jones potential and the Coloumb potential. The terms are more computationally intensive because they include in principle all pair interactions, thus they grow proportionally to  $N^2$ . The Lennard-Jones potential models the dispersion interactions and steric repulsion between atom pairs. The potential is given by the equation

$$V_{i,j}^{LJ}(r) = 4\epsilon_{i,j} \left[ \left( \frac{\sigma_{i,j}}{r_{i,j}} \right)^{12} - \left( \frac{\sigma_{i,j}}{r_{i,j}} \right)^6 \right] \quad (2.1.10)$$

where the first and second terms correspond to the repulsion and attraction parts, respectively. Parameters  $\epsilon$  corresponds to the depth of the energy minimum and  $\sigma$  represents the characteristic length. The potential model is shown in Figure 2.1.2.



**Figure 2.1.2:** The Lennard-Jones potential model (blue curve). The repulsive force, due to quantum mechanical effects, gives a potential proportional to  $(1/r)^{12}$  (dotted orange line), and the attractive force, due to a dipole-dipole interaction, gives a potential proportional to  $(1/r)^6$  (dotted green line). Here,  $\sigma$  is a characteristic length and  $\epsilon$  a characteristic energy.

The other non-bonded term is the Coloumb potential, which models the electrostatic interactions between two charged atoms. The Coloumb potential is given by the equation

$$V_{i,j}^{Coloumb}(r) = \frac{q_i q_j}{4\pi\epsilon_0 r_{i,j}} \quad (2.1.11)$$

where  $4\pi\epsilon_0$  is the electric conversion factor,  $r$  is the distance between atoms  $i$  and  $j$  with corresponding charges  $q_i$  and  $q_j$ . The Lennard-Jones potential is a well-behaved, fast decaying potential, and it is usually numerically estimated by truncation at a specific cut-off distance. The Coulomb potential is a slow-decaying potential, for which truncation methods are not possible, and it is typically estimated by reciprocal space method approaches like Particle-Mesh Ewald [39, p.65-69].

## Bonded Interactions

The intramolecular potentials are typically represented as harmonic or simple periodic potentials. The bonding potentials represent three types of local molecular deformations: bond stretching, angle bending and torsional rotation. A last term, improper dihedral, can be defined in order to describe special (i.e., flat) molecular organization promoted by quantum mechanical electron delocalization over more than two atoms (for example, aromatic groups), which cannot easily be reproduced by simple mechanical terms.

### 2.1.4 Force Fields

The mechanical potential function  $V$  is determined by a large number of parameters including equilibrium structural values and force constants for the bonded terms, Coulomb point charges, and Lennard Jones parameters. The term *force field* refers to any consistent set of these parameters, derived by arbitrary means, which can be employed for molecular mechanics modelling. In all-atom force fields, the smallest interacting unit is single atoms and parameters for every type of atom in a system is provided by the specific force field. In the past decade, several groups have been active in developing accurate force fields for simulations of biosystems, the most popular being AMBER [43], GROMOS [44], OPLS/AA[45] and CHARMM [46] [41]. In our simulations, the CHARMM36 force field (2018) was used due to its accuracy and versability.

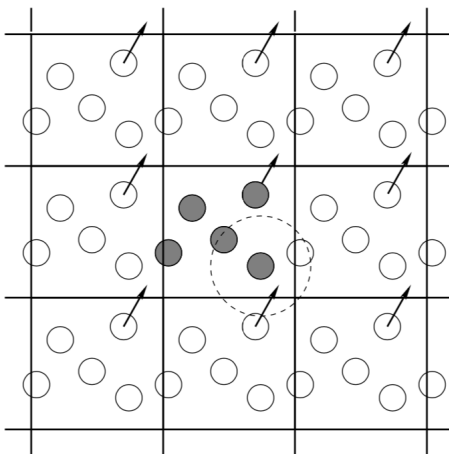
### 2.1.5 Initial Conditions

The environment for the simulated protein must be, to some extent, similar to the experimental conditions, specifically regarding temperature, pressure and solvent. In MD simulations of biomolecules, the most utilized thermodynamic ensembles are the isothermal-isobaric ensemble  $NPT$ , where the number of particles  $N$ , the temperature  $T$ , and pressure  $P$  are kept constant. In order to mimic the temperature

and pressure coupling to the exterior, additional thermostat and barostat algorithms must be coupled to the standard MD time integration [47].

### 2.1.6 Periodic Boundary Conditions

It is beyond computational capabilities to model macroscopic systems of the order of  $10^{23}$  atoms. However, it is possible to gain some insight into how large or even infinite systems behave. In MD simulations, a solute is solvated in a three-dimensional box. This box has external boundaries that causes surface effects to have a high impact on the calculations where it is favorable to simulate as few molecules as possible. Therefore, periodic boundary conditions (PBC) pose a solution to this problem. With PBC, the system can be fooled into believing it is infinite by replicating the box through space. In this way, an atom at the left side of the box can be removed and reinserted at the right side, so the atom can "move through" the box [48]. Following this, it would also mean that an atom can interact with their images in nearby boxes. This would cause the number of interacting atom pairs in the system to become high due to PBC which would not be suitable for the short interaction range from the potentials. The minimum image criterion ensures that each atom interacts with the closest atom among all the possible periodic images (Figure 2.1.3) [40].



**Figure 2.1.3:** Periodic boundary conditions with the center box in the middle with 8 periodic images surrounding it. When an atom moves out of the box, another moves in to replace it. Figure retrieved from [49].

### 2.1.7 Solvent Treatment

In biomolecular simulations, water is the most used solvent. Solvent atoms compose more atoms than the solute and therefore, the treatment of solvent is an important factor. There are two main solvent models available; implicit and explicit. Implicit solvent models treat the solvent as a continuous medium with dielectric properties.

The potential mean force can be applied to describe the average behavior of properties. In explicit solvent models, on the other hand, solvent molecules are treated explicitly and can interact with solute molecules. Therefore, explicit models are more physically relevant but also less computationally efficient due to computing interactions between all pairs of solute and solvent atoms [50].

### 2.1.8 Limitations of Molecular Dynamics

MD simulations are in many respects quite similar to real experiments, as it allows to measure the statistical properties of many-body systems [48]. With the advancements in computing, MD simulations are becoming increasingly able to resemble the collective effects seen in traditional experiments, and are also able to study regions not experimentally available [40]. Regardless, the realism of MD simulations must be put under scrutiny. The method inevitably makes assumptions and approximations whose impact need to be carefully assessed.

First, in MD simulations atoms follow classical mechanics dynamical laws, which are correct only within the Born-Oppenheimer approximation of the quantum mechanical Schrödinger equation. Second, the accuracy of MD simulations depends on the capability of the force field to reproduce the electronic energy dependent on the position of the particles. Finally, the simulations are also restricted regarding time and size of the systems that depend on the computational power available. This may severely limit a well-balanced exploration of the relevant conformational space. Moreover, simulation times may be often too short than the characteristic relaxation times of the quantities of interest [40]. The well-known problems related to poor sampling affect especially the quantitative estimation of ensemble-averaged quantities like free-energy or related thermodynamic properties. These issues may be overcome by coupling MD to enhanced sampling techniques.

### 2.1.9 Analysis of Molecular Dynamics Trajectories

#### RMSD

The root-mean-square deviation (RMSD) is a measure of how much the protein has changed over the course of the trajectory. The coordinates of the reference structure of the protein and relaxed structure are superimposed for the averaged atom pairs and the deviation is calculated by least-squares fit to the reference structure. The RMSD is calculated as:

$$RMSD(t) = \sqrt{\frac{1}{N} \sum_{i=1}^N (\mathbf{r}_i(t) - \mathbf{r}_i^{ref})^2} \quad (2.1.12)$$

where  $\mathbf{r}_i(t)$  is the current position of atom  $i$  at time  $t$ ,  $\mathbf{r}_i^{ref}$  is the position of the same atom in the reference (typically: initial) structure. The RMSD of proteins



is usually computed from the backbone  $C_\alpha$  atoms only, to avoid noisy data from larger fluctuations of the side-chains [41].

### RMSF

The root-mean-square fluctuation is a measure of flexibility of the residues of the protein:

$$RMSF_i = \sqrt{\langle (\mathbf{r}_i - \langle \mathbf{r}_i \rangle)^2 \rangle} \quad (2.1.13)$$

where  $\mathbf{r}_i$  is the instantaneous current atom position, and the  $\langle \rangle$  brackets indicate a time average operation computed over the whole trajectory. For the same reasons as RMSD, RMSF is usually computed on the  $C_\alpha$  atoms only in proteins.

### Principal Component Analysis

Principal component analysis (PCA) is a method for determining the global motion of proteins from MD trajectories. The protein conformations in the trajectories are represented as a vector space in  $3N$  dimensions where  $N$  is the number of atoms. The number of dimensions is equal to the degrees of freedom (DOF). The DOF characterize the motion of the protein, however, only a few DOF contribute significantly to the global fluctuation. To find the DOF contributing most to the global motion, a covariance matrix can be used [51]. A covariance matrix is a square, symmetric matrix that shows the correlation among residues averaged over a MD trajectory. The covariance matrix is obtained from the expression:

$$C_{i,j} = \langle (r_i - \langle r_i \rangle)(r_j - \langle r_j \rangle) \rangle \quad (2.1.14)$$

where  $r_i$  and  $r_j$  represent the atomic coordinates. By diagonalizing the covariance matrix through orthogonal transformation, the PCA can be found as the eigenvectors of the matrix. Practically, PCA filters the real-time complex noisy and apparently chaotic motion into simple global rigid-body components:

$$C_{i,j} = R \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_N) R^T \quad (2.1.15)$$

where the columns of matrix  $R$  correspond to the eigenvectors [51]. Because of the unitary transformation property, the respective eigenvalues  $\lambda$  correspond to the amplitude of each eigenvector, summing to the total mean square fluctuation of the system. Typically, low-frequency global deformations contain most of the motion of the systems, therefore, it is common to analyze only the *essential dynamics* composed by a small subset of PCs (up to 20 modes) to have a clear picture of the global motion of the system [52, p.194-195].

## 2.2 Computational Approach

### 2.2.1 System Setup

At the beginning of this study, the only available structures of Aurora B were the X-ray crystal structures of *Xenopus laevis* [27] [25] that is 77.42% identical to human Aurora B. To build the initial molecular structure for the simulations, the structure was taken from protein data bank 4C2W with most complete chains B and D, refined to 1.7 Å resolution (PDB entry: 4C2W<sup>1</sup>[27]). This structure encompasses Aurora B(78-356) and INCENP(798-847) with ATP-analog AMP-PNP (Adenylyl-imidodiphosphate) in the active site. Aurora B<sup>Thr248</sup> is phosphorylated, indicating that the enzymatic complex is in the partially active conformation. Other partially active structures were available, but they all contained inhibitors in the active site. Due to low resolution and containing an ATP-analog in the active site, 4C2W was chosen. However, ATP is the natural substrate of kinases and in order to run simulations with ATP, AMP-PNP was replaced with ATP-Mg<sup>2+</sup> from homologous Aurora kinase A (PDB entry: 5DN3<sup>2</sup>[53]). The C-terminal end of INCENP was extended as a coiled coil by including TSSAVWHSP. This extension contains the TSS-motif that becomes phosphorylated in the process of auto-activation. The structures used in simulations therefore have residues Aurora B<sup>78-356</sup> and INCENP<sup>798-856</sup>. Titratable groups were protonated at standard positions at pH 7. The all-atom additive CHARMM36 protein force field was used to parametrize the proteins and ligands of the system [54]. The system was solvated by ~36000 water molecules and with 104 and 113 sodium and chloride ions, respectively, to reach the physiological salt concentration of 0.15 M, and charge neutrality. The total number of atoms contained in the system was ~112800. The initial periodic box dimensions were 105.6 × 118.4 × 92.5 Å<sup>3</sup>, as shown in Figure 2.2.1.

Aurora B/INCENP-complex in fully phosphorylated state was generated in PyMOL [55] by addition of phosphate groups to Aurora B<sup>Thr248</sup> and INCENP<sup>Ser849-Ser850</sup>. In unphosphorylated state, the phosphate group was removed from Aurora B<sup>Thr248</sup>. The phosphorylation states of the enzymatic complex are listed in Table 2.2.1.

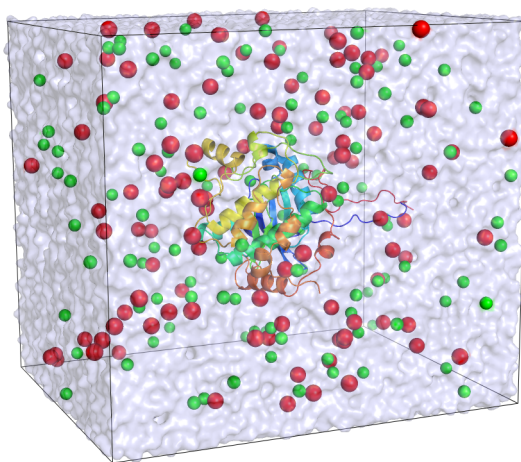
---

<sup>1</sup><http://www.rcsb.org/structure/4C2W>

<sup>2</sup><http://www.rcsb.org/structure/5DN3>

**Table 2.2.1:** The three Aurora B/INCENP phosphorylation states simulated in the study:  $AURKB^{no-P}$  (not phosphorylated),  $AURKB^{loop-P}$  (partially phosphorylated, only on Aurora B<sup>Thr248</sup>) and  $AURKB^{full-P}$  (fully phosphorylated, on Aurora B<sup>Thr248</sup> and  $INCENP^{Ser849-Ser850}$ ). Phosphorylated or not phosphorylated substrate residues are stated with yes or no (Y/N).

State	P-Thr248 (Y/N)	P-Ser849 – P-Ser850 (Y/N)	Atoms in complex (#)
$AURKB^{no-P}$	N	N	5223
$AURKB^{loop-P}$	Y	N	5226
$AURKB^{full-P}$	Y	Y	5232

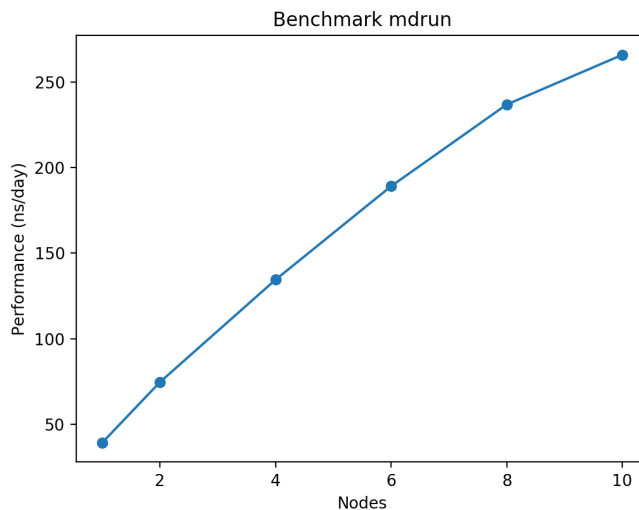


**Figure 2.2.1:** System of  $AURKB$  and  $INCENP$  in the initial periodic box after solvation of  $TIP3P$  water molecules and adding salt to a concentration of 0.15 M. Sodium and chloride ions are red and green, respectively.

## 2.2.2 Simulation Parameters

**TIP3P** All systems were relaxed to a state of low energy by the steepest descent method with a tolerance of 1000 kJ/mol/nm. The energy minimization was run for 0.5 ns with step size  $\Delta t=10$  fs and converged after 826 steps. Equilibration of the systems was carried out in NVT- and NPT ensembles. The system was equilibrated by simulated annealing in three steps to reach the temperature of 300 K after 2 ns with the Berendsen thermostat [56] with a coupling constant of 2 ps. For the NPT ensemble, the Berendsen barostat was applied to keep the pressure at 1 bar with coupling constant of 2 ps. The particle mesh Ewald method [57] was used to treat the long-range electrostatic interactions, Lennard-Jones terms were computed with a cutoff distance of 10 Å. The nonbonded list was updated every 10 steps in a total of 700 ns simulation with time step  $\Delta t=2$  fs. All bonds were constrained with the LINCS algorithm [58].

The relaxed partially phosphorylated system was used as a starting point for 2.8  $\mu$ s long MD simulations. The temperature was kept constant by the Nosé-Hoover thermostat [59][60][61] with coupling constant of 1 ps. The barostat Parrinello-Rahman [62] was used for keeping the pressure constant in the NPT-ensemble with coupling constant of 2 ps. The long-range van der Waals cut-off was 10 Å. The leap-frog algorithm was used for integrating Newton's equations of motion. All MD runs were performed with GROMACS 2018.3 [63][64][65][66][67]. The simulations were run on the supercomputer FRAM [68] with 6 nodes with 32 tasks per node. The number of nodes were verified after benchmarking the performance on several nodes based on equilibration dynamics runs shown in Figure 2.2.2.



**Figure 2.2.2:** Benchmark of equilibration run on daily performance (ns/day) versus the number of nodes. Simulations were run on 6 nodes which is where the curve slightly starts to flatten.

Analysis of trajectories was performed using different software tools available within the GROMACS package, VMD [69] and PyMOL [55]. PyMOL and VMD were also used to create figures from simulation trajectories. Alignments were generated with ESPript3 [70]. LigPlot<sup>+</sup> was used to create ligand-protein interaction diagrams of the active site [71]. An overview of GROMACS-commands used in the study to analyze MD simulations is given in section B.1.

# 3

## Results and Discussion

In this study, the main focus was to investigate how different phosphorylation states of the Aurora B/INCENP-complex affects conformations and dynamics of regions important for kinase activity. The crystal structure represents a partially active and phosphorylated state used as a template for constructing the two remaining states to perform molecular dynamics on.

Residues of Aurora B and INCENP are referred to as for example Aurora B<sup>Thr248</sup> and INCENP<sup>Ser849</sup>, respectively. The states concerning the Aurora B/INCENP-complex will be referred to in the following sections as listed:

1. AURKB<sup>no-P</sup> – without any phosphate groups in Aurora B and INCENP
2. AURKB<sup>loop-P</sup> – one phosphate group in the activation loop of Aurora B
3. AURKB<sup>full-P</sup> – one phosphate group in the activation loop of Aurora B and two in the TSS-motif of INCENP

Some regions of kinases presented in section 1.1.3 have functions involved in activation and the following analysis investigate some of these. In this chapter, results from 2.8  $\mu$ s of MD simulations mainly focusing on the last  $\mu$ s and end of simulations of each state, are presented with focus on their differences in fluctuating properties and how they are affected by phosphorylation. Experimental results from hydrogen-deuterium exchange (HDX) and kinetic analysis are added to complement the computational results providing insights of behavior of Aurora B/INCENP-complex.

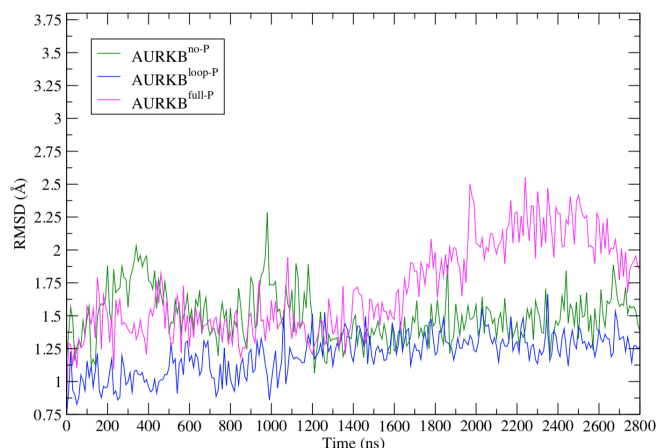
Throughout the analysis, plots and structural figures of the states follow a general color scheme where AURKB<sup>no-P</sup> is green, AURKB<sup>loop-P</sup> is blue and AURKB<sup>full-P</sup> is

magenta. An overview of GROMACS commands used for creating figures and tables from simulation trajectories are provided in section B.1 in the Appendix.

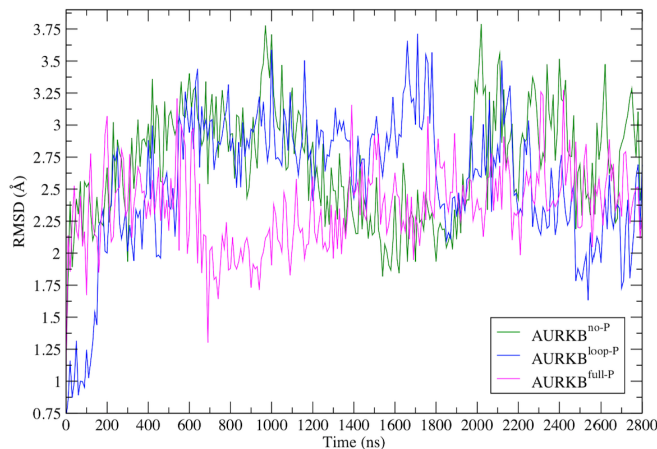
### 3.1 Properties of Conformation and Fluctuation

First, we wanted to understand what regions of the enzyme complex exhibit increased overall conformational changes and fluctuations in different phosphorylation states. As RMSD provides information about global conformational changes, RMSF gives information about local dynamics and these quantities were measured over the simulation trajectory.

The RMSD analysis follows root-mean-square deviation of the C $\alpha$ -atoms in the structure during MD simulations of 2.8  $\mu$ s. The RMSD plots for all phosphorylation states of the complex were computed for Aurora B<sup>89-350</sup> and INCENP<sup>798-839</sup>. The terminal parts for both proteins were excluded from the analysis due to extreme flexibility that would otherwise dominate in the plot. The output plots are shown in Figure 3.1.1.



(a)



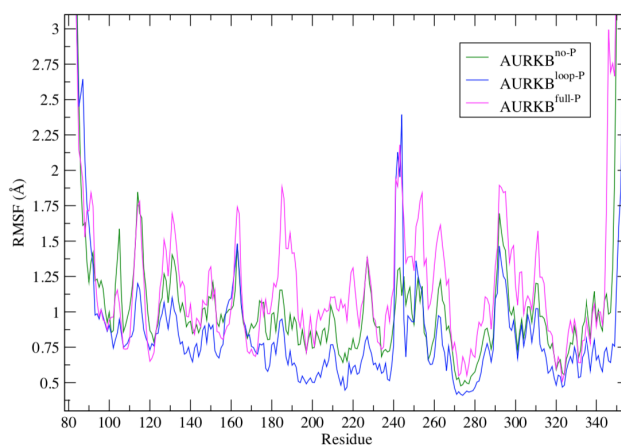
(b)

**Figure 3.1.1:** Plot showing changes in RMSD of  $C_{\alpha}$ -atoms with respect to the equilibrated conformation of Aurora  $B^{89-350}$  (a) and  $INCENP^{798-839}$  (b) during the course of MD simulation ( $2.8 \mu s$ ). The plots are made from full simulation trajectories of all states;  $AURKB^{no-P}$  (green),  $AURKB^{loop-P}$  (blue) and  $AURKB^{full-P}$  (magenta).

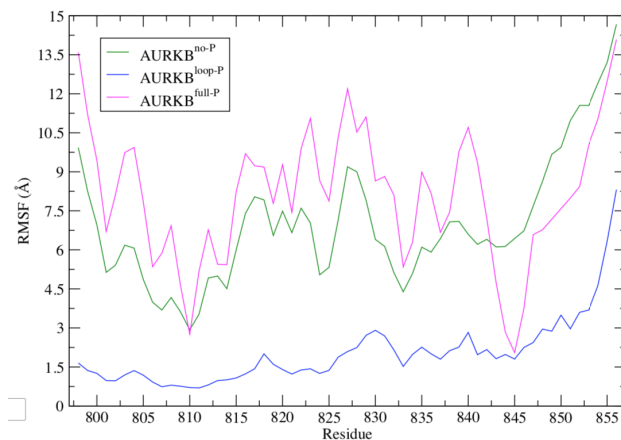
In general, the RMSD of INCENP is higher than that of Aurora B in all analyzed structures indicating a higher overall flexibility of INCENP. This is expected taking into account that INCENP lacks secondary structure elements except from an  $\alpha A$ -helix and wraps around Aurora B in a crown-like structure using mainly hydrophobic interactions. All three phosphorylation states exhibit high conformational variability in INCENP. For Aurora B, the state with the overall lowest RMSD is  $AURKB^{loop-P}$  (blue in Figure 3.1.1). This is also expected given that the crystal structure used for initial model generation was actually partially phosphorylated. The unphosphorylated state also shows steady RMSD for Aurora B indicative of

conformationally stable structure that has reached equilibrium. Interestingly, the RMSD for Aurora B in  $\text{AURKB}^{\text{full-P}}$  increases substantially in the range 1.4 - 2.2  $\mu\text{s}$  most likely because the enzyme complex needs to undergo larger conformational changes in order to find a more stable resting state. Together, these results identify INCENP as the conformationally unstable part of the complex in all phosphorylation states and point towards  $\text{AURKB}^{\text{full-P}}$  as the most conformationally divergent structure among those analyzed.

Root-mean-square fluctuation (RMSF) was calculated for each  $C_{\alpha}$ -atom on the last  $\approx 1\mu\text{s}$  of MD simulations. While RMSD analysis points to global conformational changes, RMSF analysis identifies flexible parts of the protein. Figure 3.1.2 shows the RMSF for all states of Aurora B.



(a)



(b)

**Figure 3.1.2:** RMSF on  $C_{\alpha}$  of all residues of Aurora B (a) and INCENP (b) computed for the last  $\mu\text{s}$  of simulation. The states are  $\text{AURKB}^{\text{no-P}}$  (green),  $\text{AURKB}^{\text{loop-P}}$  (blue) and  $\text{AURKB}^{\text{full-P}}$  (magenta). Also to be noticed is the higher values on the y-axis in INCENP (b) compared to Aurora B (a).



For Aurora B (Figure 3.1.2(a)), it can be observed that the N and C-terminal ends have higher fluctuations which is expected as they are unstructured loops. Other particular regions within Aurora B also have high fluctuations, which will be considered in section 3.4. Interestingly, while structure rigidification was reported for AURKB<sup>full-P</sup> based on HDX-MS experiments, the RMSF analysis of AURKB<sup>full-P</sup> had generally higher flexibility than the other phosphorylation states. This could be because the molecular system did not reach equilibrium and requires longer simulation. Alternatively, since RMSF measures fluctuations between C<sub>α</sub>-atoms during MD simulations, it is possible that high RMSF-values are a consequence of bigger and more global conformational change, rather than intrinsic flexibility. To distinguish between the two possibilities we have performed a principle component analysis (PCA) to investigate concerted motions in each of the phosphorylation states of the Aurora B/INCENP-complex (section 3.3).


The INCENP structures are not in stable equilibria due to its unstructured nature, but the RMSF was still plotted for main chain C<sub>α</sub>-atoms to obtain information about its fluctuation in the different states (Figure 3.1.2(b), notice the much higher value on the y-axis). We elaborated on the conformational changes of INCENP taking place during the course of the simulations in each of the states (section 3.2).

## 3.2 INCENP Drastically Changes Conformation Upon Aurora B<sup>Thr248</sup> Phosphorylation

INCENP (INner CENTromere Protein) is a long (in human 918 residues) adaptor protein. It has a localization module (binds Survivin and Borealin) at the N-terminus and catalytic module that binds Aurora B kinase at the C-terminus. The C-terminal region of INCENP, spanning 84 residues, is involved in binding of Aurora B. The interaction between INCENP and Aurora B is constant and necessary for kinase activity. In vitro experiments have shown that Aurora B, although soluble, in the absence of INCENP cannot catalyze phosphate transfer [25].

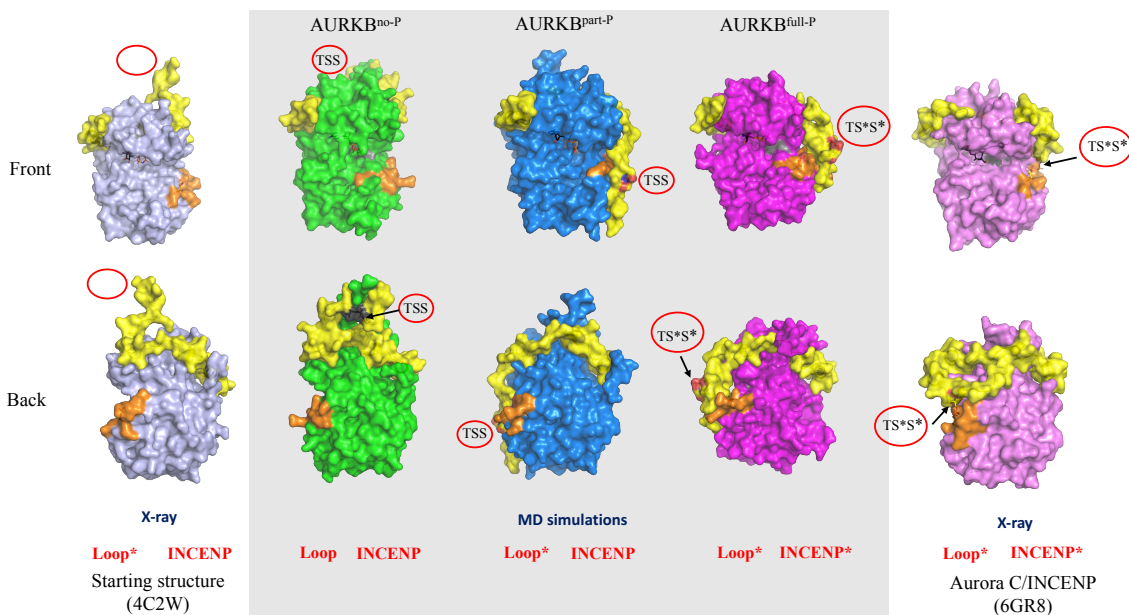
As it can be seen from the videos of full simulations (available by downloading the public GitHub repository in folder *movies\_full\_sim*<sup>1</sup>), the structure of INCENP is fluctuating until it finally adopts stable conformations, first for AURKB<sup>full-P</sup> after 0.8  $\mu$ s, for AURKB<sup>loop-P</sup> after 1  $\mu$ s, while for AURKB<sup>no-P</sup>, it takes 2  $\mu$ s out of 2.8  $\mu$ s that full MD simulations were run for. In both AURKB<sup>full-P</sup> and AURKB<sup>loop-P</sup> where Aurora B<sup>Thr248</sup> is phosphorylated, INCENP stabilizes in the neighborhood of the loop region, albeit in different conformations. For AURKB<sup>no-P</sup>, the C-terminal end of INCENP fluctuates between the loop region and the N-terminal end of Aurora B, but it finally stabilizes in the interaction with Aurora B N-terminus, which is a flexible unstructured region of Aurora B, mainly through main chain hydrogen bonding and hydrophobic interactions. It is possible that negatively charged phosphorylated Aurora B<sup>TPO248</sup> provides a good docking point for INCENP

---

<sup>1</sup>  odasho/Master-thesis

that gets stabilized in a “downward” conformation when the activation loop is phosphorylated. Interestingly, phosphorylation of the TSS-motif on INCENP adds negative charge to it, but it still electrostatically interacts with the Aurora activation loop, shielding it from the solvent. The protection of the activation loop has earlier been shown for Aurora A with TPX2 [20]. The interactions between INCENP and activation loop are going to be analyzed in detail in subsection 3.4.2.

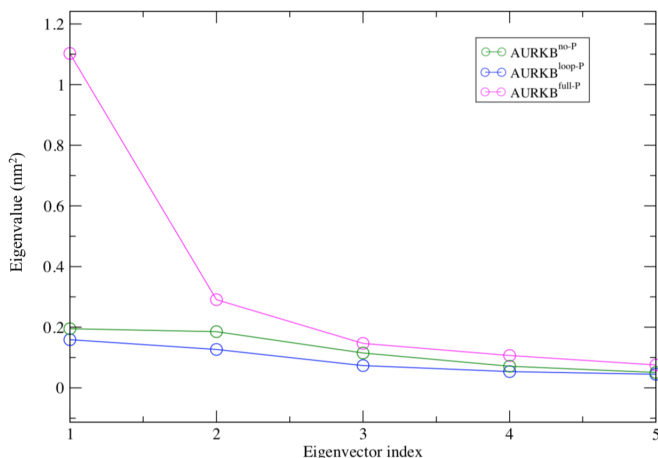
While we were first to observe this dramatic change in the INCENP conformation coupled to activation loop phosphorylation, concomitantly with this study, last summer, a crystal structure of highly similar human Aurora C/INCENP-complex was published [20]. Interestingly, in this structure, INCENP also interacts with the activation loop albeit in a slightly different conformation. The different conformations of INCENP with marked TSS-motifs in Aurora C/INCENP-complex, our simulated structures in all states and original crystal structure of Aurora B/INCENP (PDB entry: 4C2W [27]) are presented in Figure 3.2.1. In general, the experimental data is in agreement with our simulation and it confirms high flexibility of INCENP and its involvement in regulation of the activation loop.



**Figure 3.2.1:** Structures of Aurora B/INCENP and Aurora C/INCENP where INCENP is yellow in all structures. Left (white area): Crystal starting structure Aurora B/INCENP (grey/yellow), middle (grey area):  $AURKB^{\text{no-P}}$  (green/yellow),  $AURKB^{\text{loop-P}}$  (blue/yellow),  $AURKB^{\text{full-P}}$  (magenta/yellow) and right (white area): AuroraC/INCENP (pink/yellow). The TSS-motif (shown with \* if phosphorylated) is pointed out for all structures except starting structure (empty circles). Phosphorylation (\*) in loop and/or INCENP is stated at the bottom for each structure.


### 3.3 PCA Reveals Global Kinase Motions of Aurora B

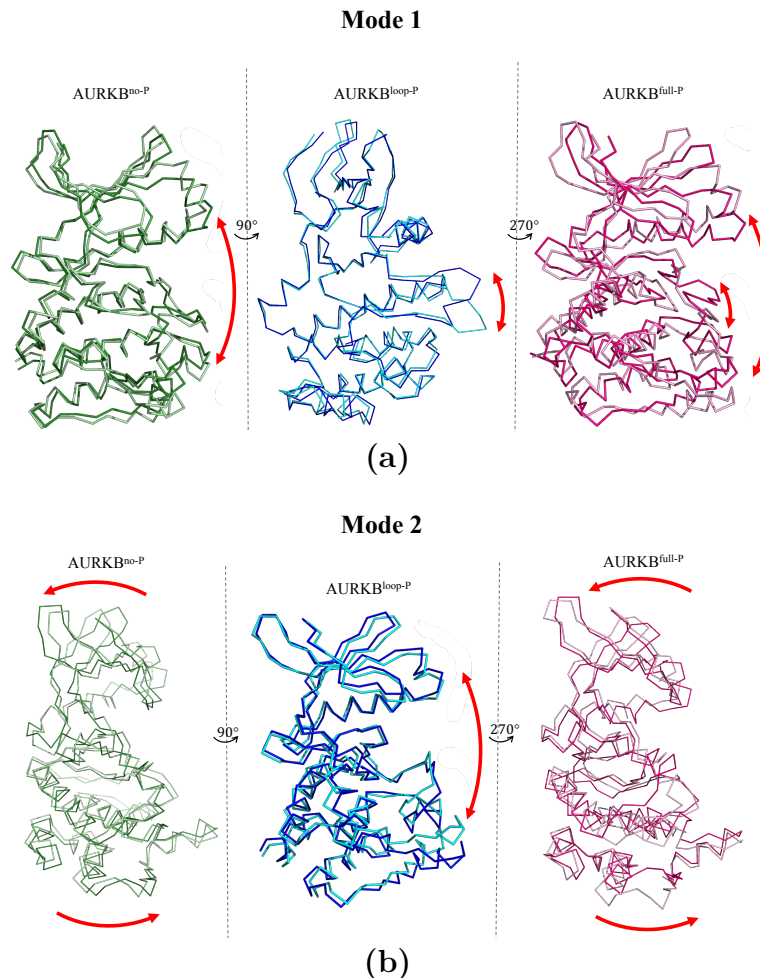
PCA is a statistical method to filter out the global motion of proteins in modes that sort the motions from larger to smaller spatial scales. Here, we used it to analyze if there are global rigid body movements taking place in the enzymatic complex in each state and obtain a better understanding of the motions contributing to the high RMSF observed for the fully phosphorylated state (Figure 3.1.2). The eigenvectors were obtained by diagonalizing the  $\sim 753 \times \sim 753$  covariance matrix of the  $C_\alpha$  atoms of the core folded region of Aurora B<sup>89-350</sup>. Figure 3.3.1 reports the corresponding eigenvalues, from highest to lowest representing the most and least significant motions, respectively. In particular, the figure reports the first five eigenvectors, which alone make up a very a large portion of the global motion, 41.4%, 39% and 62.3%, for states AURKB<sup>no-P</sup>, AURKB<sup>loop-P</sup> and AURKB<sup>full-P</sup>, respectively.



**Figure 3.3.1:** The eigenvalue ( $\text{nm}^2$ ) plotted against eigenvector index. The most significant motion is contained in the first eigenvector with high eigenvalues.

It can be observed from Figure 3.3.1 that AURKB<sup>full-P</sup> has a substantially higher eigenvalue in the first eigenvector than the other two states, indicating the activation of a global rigid body motion for this state. We will focus on the two eigenvectors with highest values, referred to as mode 1 and mode 2. The two most extreme projections of the motion of mode 1 and 2 were aligned for each state (Figure 3.3.2). Videos of the motions are also available in the GitHub repository [odasho/Master-thesis/movies\\_PCA\\_states](https://github.com/odasho/Master-thesis/movies_PCA_states)<sup>2</sup>.

<sup>2</sup>  [odasho/Master-thesis](https://github.com/odasho/Master-thesis)



**Figure 3.3.2:** The two most significant modes with corresponding motions with arrows are shown for each state. (a): For mode 1,  $AURKB^{no-P}$  has a kinase-like motion,  $AURKB^{loop-P}$  has a loop motion and  $AURKB^{full-P}$  has both kinase-like and loop motions. (b): For mode 2,  $AURKB^{no-P}$  and  $AURKB^{full-P}$  have twist-like motions, while  $AURKB^{loop-P}$  has the kinase-like motion.

It was found that each state has different types of motion describing most of the global dynamics of the system. In  $AURKB^{no-P}$ , the first two modes correspond to (1) an open-close *breathing* motion and (2) a twist motion between the two lobes of the protein defining the active site cleft. In the open-close motion, the N and C lobes display rigid body periodic motions that alternate approximations and distancing of the two kinase lobes. In the twist-motion, the N and C lobes move in opposite directions. These two motions are typical topological modes of proteins characterized by folding in two domains, and in particular of kinases [72]. The relatively low values of the corresponding eigenvalues indicate that these modes are overall well at equilibrium, and that the protein can be considered in a resting state ( $AURKB^{no-P}$ ). Interestingly, in  $AURKB^{loop-P}$  we observed the appearance of a different dominant mode (mode 1), which involves the displacement of the activation

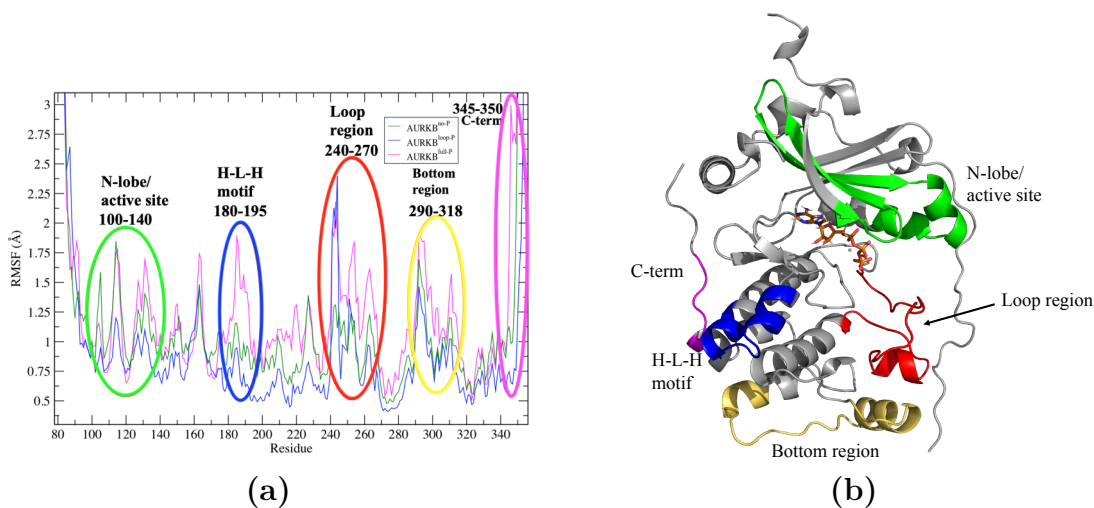
loop. In AURKB<sup>full-P</sup>, the open-close and the loop movements are coupled as the first highly activated, coherent mode, responsible in largest part for the motion of the protein. Therefore, the motion of the activation loop is coupled with the open-close motion of the lobes when both Aurora B<sup>Thr248</sup> and INCENP<sup>Ser849-Ser850</sup> are phosphorylated.

Our analysis has confirmed that Aurora B<sup>86-174</sup> comprise the N lobe and Aurora B<sup>175-347</sup> the C lobe with residues Aurora B<sup>Glu177</sup>, Aurora B<sup>Leu222</sup>, Aurora B<sup>Leu223</sup>, Aurora B<sup>Ile232</sup> and Aurora B<sup>Ala233</sup> as hinges.

PCA reveals that phosphorylation induces enhancement of the motion around the active site cleft, and in particular full phosphorylation, which promotes the appearance and activation of a coherent motion between the loop and the two lobes for Aurora B. This indicates that the flexibility for AURKB<sup>full-P</sup> is associated with the promotion of a conformational change along this global deformation coordinate, possibly resembling a more catalytically competent state. This suggests that mode 1 in AURKB<sup>full-P</sup> might be functionally relevant for achieving maximal kinase activity. Mode 1 and 2 take place with higher amplitude in the fully phosphorylated system, which resembles the fully active state, albeit the fact that the system might not have reached equilibrium in the fully phosphorylated state. The PCA analysis also explained the apparent conflict between high RMSF-values observed for the fully phosphorylated state and the lower hydrogen exchange rate of this form of the enzyme (unpublished data Sekulic group).

### 3.4 Effect of Phosphorylation on Regions of the Complex

When the Aurora B/INCENP-complex becomes phosphorylated, it undergoes conformational changes as well as changes in dynamics. We have used PCA to analyze rigid body motions (section 3.3), but we turn back to RMSF analysis to observe the regions with high local mobility which are highlighted in Figure 3.4.1.



**Figure 3.4.1:** (a): RMSF plot for all three phosphorylation forms of Aurora B (modified from original RMSF plot in Figure 3.1.2). Regions with higher fluctuations are encircled with different colors. (b): Corresponding encircled regions from RMSF shown on final conformation of  $AURKB^{loop-P}$  following the same color-scheme as in (a).

Several parts of the enzyme complex are experiencing high RMSF. Based on their well-established function and our own experimental data (HDX-MS, Sekulic group unpublished) we decided to focus our analysis on the following regions: active site including ATP binding and ion-pair Lys-Glu, activation loop (that also interacts with INCENP in  $AURKB^{loop-P}$  and  $AURKB^{full-P}$ ) and  $\alpha$ G-helix.

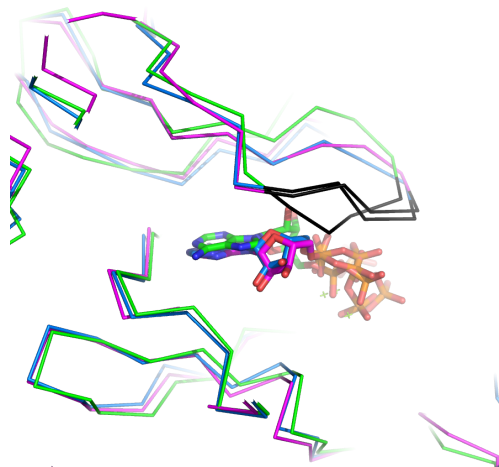
### 3.4.1 Active Site

The active site of Aurora B is located between the N and C lobes where substrates  $ATP-Mg^{2+}$  and the protein substrate bind. For the catalytic reaction to take place, the enzyme needs to form a stable complex with ATP in the active site and the protein substrate. The catalytic reaction occurs where  $\gamma$ -phosphate of ATP is transferred to Ser or Thr of the substrate protein to yield the products ADP and phosphorylated substrate protein. The Aurora B/INCENP-complex phosphorylates its own residues Aurora B<sup>Thr248</sup> in the activation loop and INCENP<sup>Ser849-Ser850</sup> in the TSS-motif of INCENP, which means that the activation loop and INCENP can occupy the substrate position in the active site. Whether these phosphorylation events happen in *cis* (same molecule performs catalysis and serves as a substrate) or in *trans* (one molecule performs catalysis and one serves as a substrate) is still not clear. According to kinetic analysis, Aurora B<sup>Thr248</sup> in the activation loop is most likely to first become phosphorylated, which happens in *cis*, and subsequent phosphorylation of the Serines in the TSS-motif of INCENP most likely happens in *trans* [33][25].

As mentioned before, we have modelled ATP-Mg<sup>2+</sup> in the active site, but the residue substrate site is left unoccupied in our simulations. The x-ray structure of the partially phosphorylated Aurora B is characterized by the presence of ATP-analog AMP-PNP which has NH-group instead of a bridging oxygen between  $\beta$ - and  $\gamma$ -phosphates and no coordinating Mg<sup>2+</sup> ion. Therefore, ATP was modelled in Aurora B, using the structure of ATP observed in the x-ray structure of the active, phosphorylated homologous Aurora A (PDB: 5DN3)[53]. An overview of the ATP binding mode for AURKB<sup>no-P</sup>, AURKB<sup>loop-P</sup> and AURKB<sup>full-P</sup> follows.

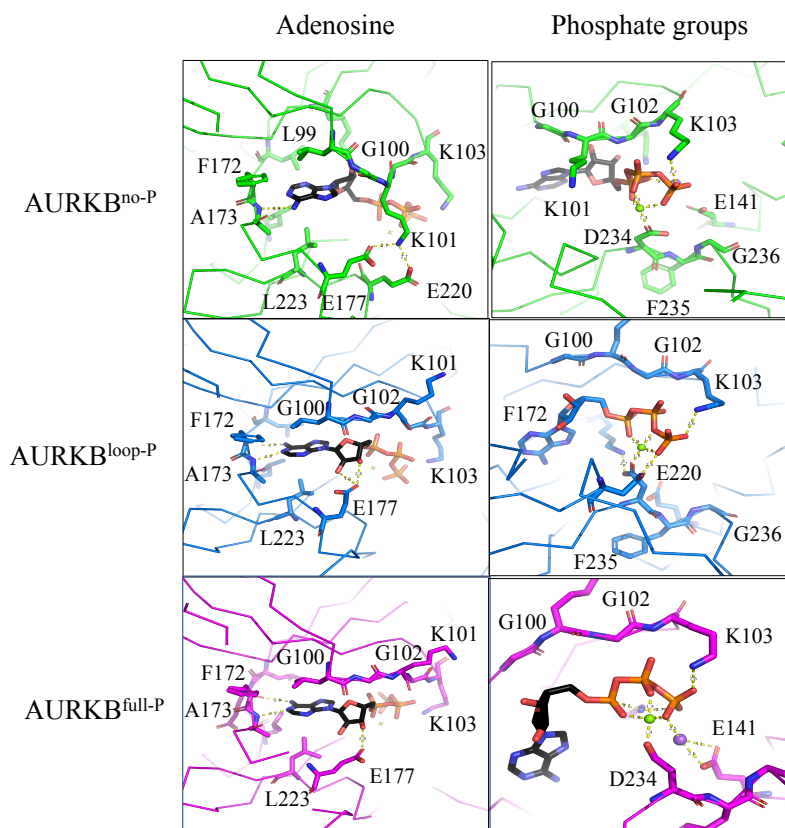
### ATP Binding

The binding site of ATP is sandwiched between the N and C lobes in the active site. The phosphate binding loop (P-loop) (black in Figure 3.4.2), also called Walker A motif, is located between two  $\beta$  strands in the N-lobe. The function of the P-loop is to interact with the  $\beta$ - and  $\gamma$ -phosphate groups of ATP through the amide nitrogen of the backbone chain and Lys side chain. Upon alignment of the enzyme complex in the three states, it became clear that P-loop in AURKB<sup>no-P</sup> adopts a drastically different conformation than AURKB<sup>loop-P</sup> and AURKB<sup>full-P</sup> (Figure 3.4.2).

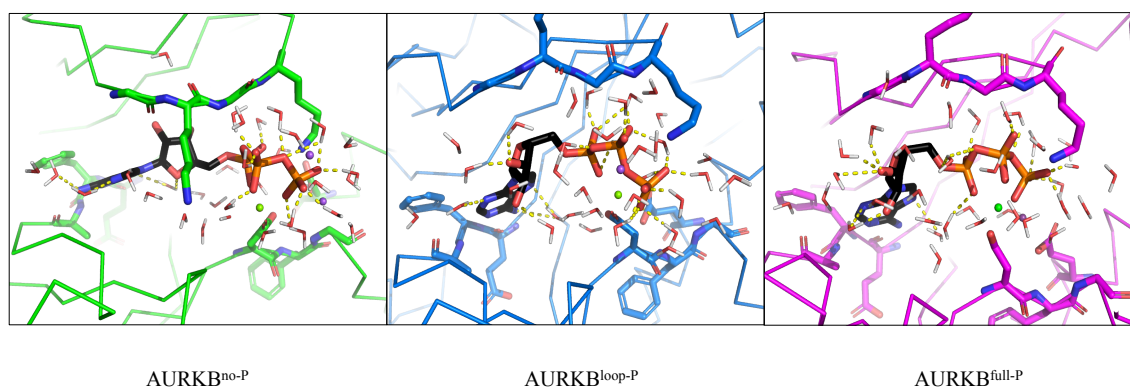


**Figure 3.4.2:** Conformation of the P-loop (black, residues 102-105) in the aligned three states at end of simulations. The states are AURKB<sup>no-P</sup> (green), AURKB<sup>loop-P</sup> (blue) and AURKB<sup>full-P</sup> (magenta). The P-loop is more unstructured in AURKB<sup>no-P</sup>.

The different conformation observed in AURKB<sup>no-P</sup> is due to the electrostatic interaction between Aurora B<sup>Lys101</sup> in the P-loop with Aurora B<sup>Glu177</sup> and Aurora B<sup>Glu220</sup>. In the other two phosphorylation states (AURKB<sup>loop-P</sup> and AURKB<sup>full-P</sup>), Aurora B<sup>Lys101</sup> is facing the solvent while Aurora B<sup>Glu177</sup> is involved in ATP binding through electrostatic interaction with ribose (Figure 3.4.3). ATP in the active site also forms hydrogen bonds to surrounding water molecules in all states as seen in Figure 3.4.4.



**Figure 3.4.3:** Important hydrogen bonds formed in the active site between ATP and surrounding residues for all states at the end of simulations.  $Mg^{2+}$  and  $Na^{+}$  are depicted as green and purple balls, respectively.

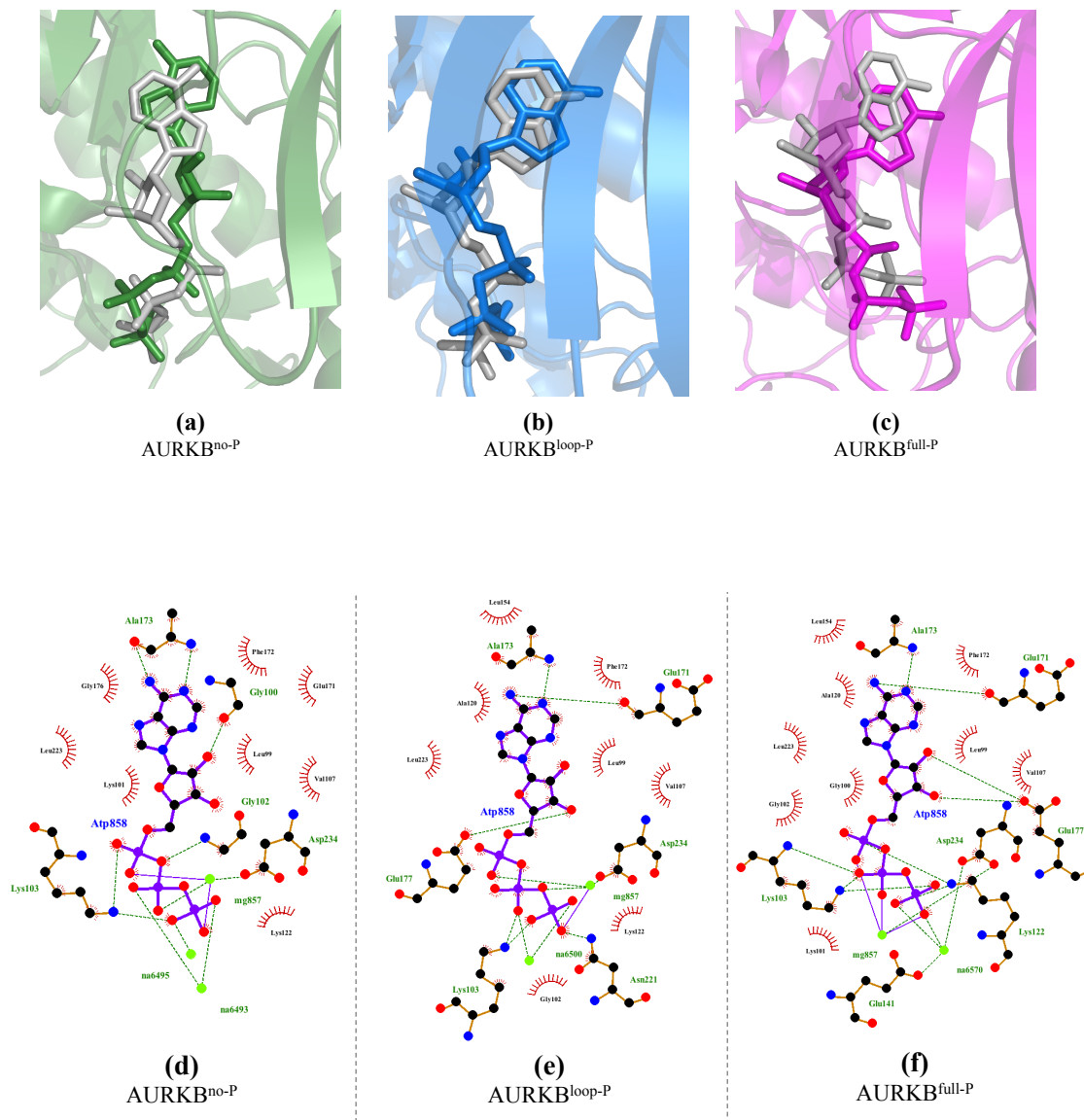


**Figure 3.4.4:** ATP forms hydrogen bonds to surrounding water molecules in the active site in all states at the end of simulations.

Due to these altered interactions between the states, ATP adopts a completely different orientation in AURKB<sup>no-P</sup> as shown in Figure 3.4.5 (top figures a, b and c), where the initial and end conformations of ATP were aligned for each state. To understand how ATP is oriented in relation to its surrounding residues, ligand-protein



interaction diagrams for end conformations are presented in Figure 3.4.5(bottom figures d, e and f).



**Figure 3.4.5:** Top (a, b, c): ATP in the active site aligned for initial (light grey) and end conformation (AURKB<sup>no-P</sup> (green), AURKB<sup>loop-P</sup> (blue) and AURKB<sup>full-P</sup> (magenta)). Bottom (d, e, f): ATP-protein interaction map for each state (AURKB<sup>no-P</sup> (left), AURKB<sup>loop-P</sup> (middle) and AURKB<sup>full-P</sup> (right)). Hydrogen bonds are represented with green dashed lines and residues involved in hydrophobic interactions are represented as red half circles.

We noticed from the top figures in Figure 3.4.5 that the conformations of ATP in AURKB<sup>loop-P</sup> and AURKB<sup>full-P</sup> are fairly maintained along the whole trajectory,

which is not the case for  $\text{AURKB}^{\text{no-P}}$  where ATP adopts a completely different conformation. In particular, we also observed a flipped orientation of the purine ring toward the end of the simulation as well as the ribose sugar. The changed orientation of ATP suggests that it is more tightly bound in  $\text{AURKB}^{\text{loop-P}}$  and  $\text{AURKB}^{\text{full-P}}$  than in  $\text{AURKB}^{\text{no-P}}$ , where loose interactions allow for dramatic displacement in the binding cavity.

The adenine ring of ATP is placed in the hydrophobic pocket formed by residues Aurora B<sup>Phe172</sup>, Aurora B<sup>Leu223</sup>, Aurora B<sup>Leu99</sup>, Aurora B<sup>Val107</sup> and Aurora B<sup>Ala120</sup> (Figure 3.4.5(bottom)). The orientation of the base is secured through interactions with the main chain Aurora B<sup>Ala173</sup> ( $\text{AURKB}^{\text{no-P}}$ ) and Aurora B<sup>Glu171</sup> in  $\text{AURKB}^{\text{loop-P}}$  and  $\text{AURKB}^{\text{full-P}}$ . In  $\text{AURKB}^{\text{no-P}}$ , the adenosine (base and sugar) is flipped and this conformation is stabilized mainly by electrostatic interaction between hydroxyl groups on ribose and main chain carbonyl of Aurora B<sup>Gly100</sup> and  $\epsilon$ -amino group of Aurora B<sup>Lys122</sup>. Aurora B<sup>Lys122</sup> is an important catalytic residue and its involvement in stabilization of ATP ribose is most likely preventing catalysis.

The phosphate groups are located towards the water-filled part of the active site in all three structures and are tightly binding  $\text{Mg}^{2+}$ . The protein substrate in the substrate-enzyme complex would likely bind in this region and the pre-organization of the ATP phosphate group likely facilitates  $\gamma$ -phosphate transfer during the catalytic event. Interestingly, although we started simulations with the ATP molecule where  $\text{Mg}^{2+}$  interacts with  $\alpha$ - and  $\beta$ -phosphates, in all simulations, triphosphate of ATP adopts an altered conformation where  $\text{Mg}^{2+}$  gets involved with all three phosphates ( $\alpha$ ,  $\beta$  and  $\gamma$ ). Similar ATP conformation in the active site is also observed in protein kinase A (PKA) [10] soaked with  $\text{Mn}^{2+}$ , but not in the highly homologous Aurora kinase A [53]. In all structures,  $\gamma$ -phosphate interacts with Aurora B<sup>Lys103</sup> from the P-loop and Aurora B<sup>Asp234</sup> from the DFG-motif (Figure 3.4.3). This is expected since Aurora B<sup>Asp234</sup> is part of the conserved DFG-motif in the active site of kinases known to chelate  $\text{Mg}^{2+}$  [29].

With an understanding of how ATP is oriented in the active site, we wanted to analyze how well it is bound by calculating the number of hydrogen bonds that are formed on average between ATP- $\text{Mg}^{2+}$  and surrounding residues or water molecules from the last  $\mu\text{s}$  of simulations (Table 3.4.1).

**Table 3.4.1:** Average number of hydrogen bonds that ATP in  $\text{AURKB}^{\text{no-P}}$ ,  $\text{AURKB}^{\text{loop-P}}$  and  $\text{AURKB}^{\text{full-P}}$  is forming with protein and water molecules in the last  $\mu\text{s}$  for all states.

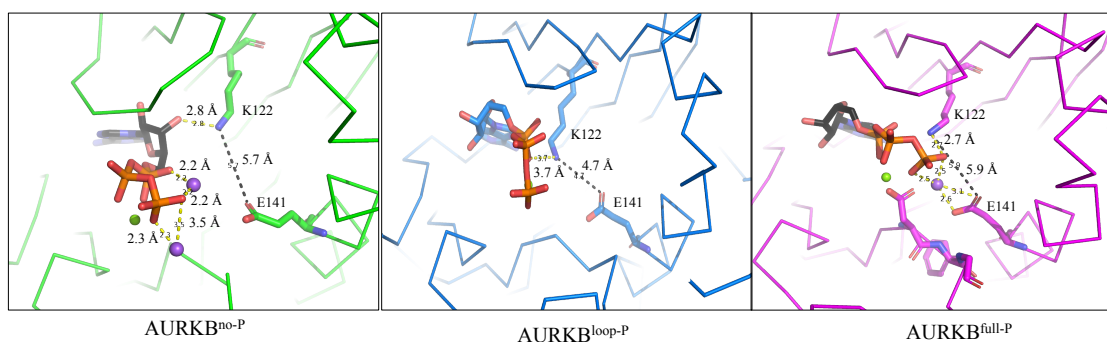
State	Avg. H-bonds to Protein (#)	Avg. H-bonds to Water (#)
$\text{AURKB}^{\text{no-P}}$	$6.080 \pm 0.862$	$13.998 \pm 2.184$
$\text{AURKB}^{\text{loop-P}}$	$5.592 \pm 1.148$	$15.402 \pm 2.116$
$\text{AURKB}^{\text{full-P}}$	$7.295 \pm 3.499$	$13.242 \pm 1.911$

From results summarized in Table 3.4.1, it can be noted that ATP-Mg<sup>2+</sup> in the active site of AURKB<sup>full-P</sup> is forming more hydrogen bonds to protein residues on average than in the other two states. This makes sense because we know that this is the most active form of the enzyme complex. Interestingly, both AURKB<sup>no-P</sup> and AURKB<sup>full-P</sup> have less water molecules bound than AURKB<sup>loop-P</sup>, although this could not be observed at the end of simulations in Figure 3.4.4. For AURKB<sup>full-P</sup> with less water molecules bound, this corresponds well with what was observed in Figure 3.4.3 and Figure 3.4.5(f) where AURKB<sup>full-P</sup> was surrounded by more residues forming interactions than in the other two states (AURKB<sup>no-P</sup> and AURKB<sup>loop-P</sup>). This indicates that AURKB<sup>full-P</sup> tends to have a more closed conformation containing less water and ions than the other two states. A more closed conformation correlates well with the findings of the global motion where the opening and closing motion was coupled with the loop motion forming a substantially higher amplitude for the state.

From Figure 3.4.5(f), we observed Aurora B<sup>Glu141</sup> and Aurora B<sup>Lys122</sup> being closely oriented in the active site. A hallmark of active kinases is a salt bridge between conserved Lys in the  $\beta_3$ -sheet of the N-lobe and Glu from the  $\alpha$ C-helix [28]. In the literature, this ionic lock is widely represented in many active kinases and shown to be related to activity [73][25][74]. The  $\alpha$ C-helix also makes direct contact with the N-terminal region of the activation loop, and its conformation is often linked to that of the conserved DFG-motif. This interaction, along with the Glu-Lys ion pair, directly couples the conformation of the helix to nucleotide binding. However, in the AURKB<sup>full-P</sup> conformation, the two residues are coming very close, but the presence of a Na<sup>+</sup>-ion in the active site is preventing the direct interaction (Figure 3.4.5(f)). To investigate if the Na<sup>+</sup>-ion could be an artifact in the simulations and if it is present over the whole trajectory preventing the formation of a salt bridge, we analyzed the distances between Aurora B<sup>Glu141</sup> and Aurora B<sup>Lys122</sup>.

### **Analysis of the Aurora B<sup>Lys122</sup>–Aurora B<sup>Glu141</sup> Ion-Pair Reveals Two Structural States in the Active Site of Phosphorylated Enzyme Complex**

In the active site of protein kinases, ionic interaction between two residues in the N lobe is a hallmark of active kinase [28]. Lys is located in the  $\beta$ -sheet and Glu in the  $\alpha$ C-helix. In Aurora B, these two residues correspond to Aurora B<sup>Lys122</sup> and Aurora B<sup>Glu141</sup> [25] (Figure 3.4.6). The salt bridge is formed between one of the hydrogens of the  $\epsilon$ -amine group of Lys and one of the two oxygen atoms of the carboxyl group of Glu, positioning Aurora B<sup>Lys122</sup> in a close interaction with the  $\alpha$ - and  $\beta$ -phosphates of ATP. Sessa et. al suggests a mechanism for activation of Aurora B where the allosteric binding of INCENP causes the  $\alpha$ C-helix to move closer to the active site [25], thus adopting a catalytically competent conformation. We investigated the final conformation from all three simulated states for presence of this hydrogen bond, but we found that residues are too far away in each of the conformations (Figure 3.4.6).

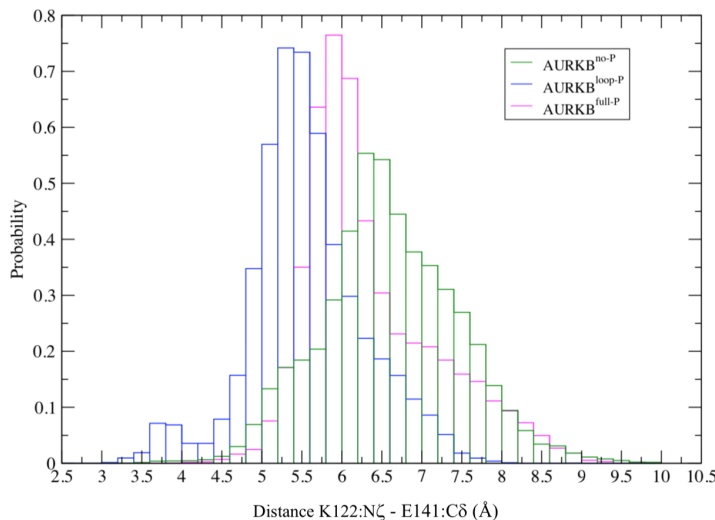


**Figure 3.4.6:** Distance between Aurora B<sup>Glu141</sup> and Aurora B<sup>Lys122</sup> measured in the final structure of each of the enzyme complex phosphorylation states (AURKB<sup>no-P</sup> (green), AURKB<sup>loop-P</sup> (blue) and AURKB<sup>full-P</sup> (magenta)). Green spheres are Mg<sup>2+</sup> ions and purple spheres are Na<sup>+</sup> ions. Distances in AURKB<sup>full-P</sup> around Na<sup>+</sup> are 2.5, 2.5, 2.6 and 3.1 Å.

In AURKB<sup>no-P</sup> the distance between the ion-pair is 5.7 Å. In addition, Aurora B<sup>Lys122</sup> is involved in the hydrogen bond formation with the hydroxyl group of the ribose in ATP. This interaction, together with presence of two additional Na<sup>+</sup> ions interacting with phosphates of ATP, probably due to high flexibility of the active site, is rendering AURKB<sup>no-P</sup> a catalytically incompetent conformation.

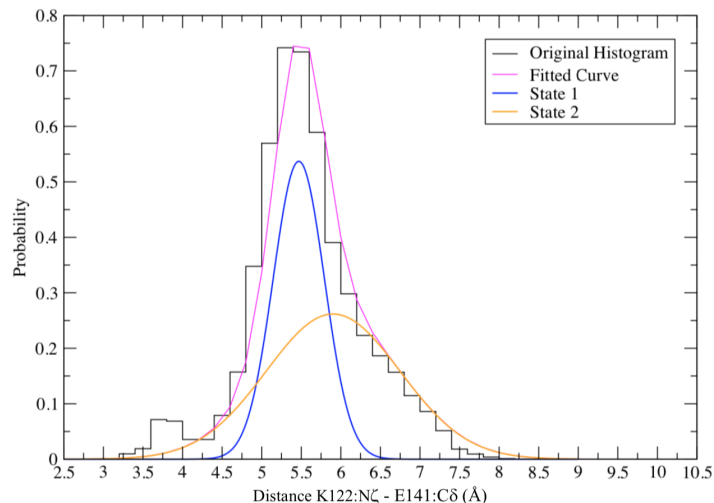
In AURKB<sup>loop-P</sup> the distance between Aurora B<sup>Lys122</sup> and Aurora B<sup>Glu141</sup> is closer (4.7 Å), but still too far for hydrogen bond formation. Finally, in completely phosphorylated form of the enzyme complex AURKB<sup>full-P</sup>, residues are further away (5.9 Å). This is surprising because we expect AURKB<sup>full-P</sup> to be the state resembling the catalytically competent form of the enzyme. However, we notice presence of Na<sup>+</sup> ion in the active site that interacts with  $\gamma$ -phosphate and Aurora B<sup>Glu141</sup>. Also, Aurora B<sup>Lys122</sup> interacts with the same  $\gamma$ -phosphate. Further investigations are needed to elaborate if the observed conformation is stable or physiologically relevant, but we still wanted to analyze if Aurora B<sup>Lys122</sup> and Aurora B<sup>Glu141</sup> make an ion-pair during our simulations. To do so, we monitored the distance between N $\zeta$ -atom of Aurora B<sup>Lys122</sup> and C $\delta$ -atom of Aurora B<sup>Glu141</sup>. The choice of measuring this distance instead of direct hydrogen-oxygen distance was taken so to avoid encompass interactions with any of the two oxygens on the Aurora B<sup>Glu141</sup>. We assumed that distances below 4.5 Å between N $\zeta$ -atom and C $\delta$ -atom signify hydrogen

bond formation. Distribution of distances for each of the phosphorylation forms is reported in Figure 3.4.7.

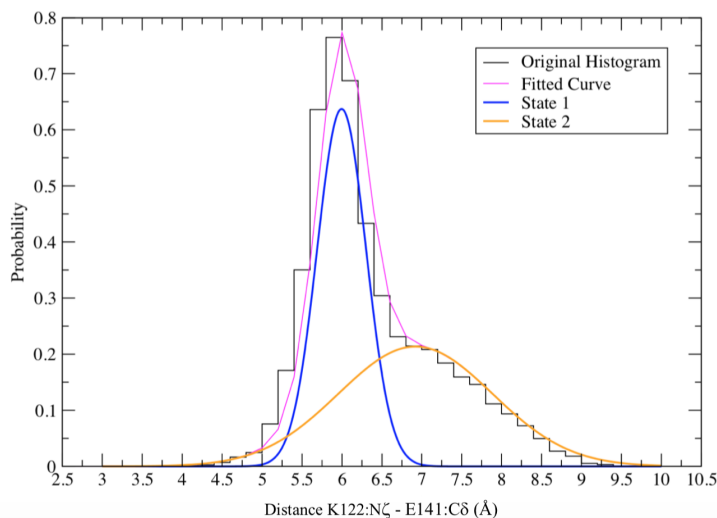


**Figure 3.4.7:** Histogram showing probabilities of distances between *Glu141:Cδ*–*Lys122:Nζ* for each state ( $AURKB^{no-P}$  (green),  $AURKB^{loop-P}$  (blue) and  $AURKB^{full-P}$  (magenta)). Distances below 4.5 Å are considered to be a hydrogen bond between the ion-pair.

From the histogram distribution in Figure 3.4.7, it can be noticed that the most probable distances are around 5.5 Å for  $AURKB^{loop-P}$  and 6 Å for  $AURKB^{no-P}$  and  $AURKB^{full-P}$ , respectively. Thus, according to our criteria we find that all the three Aurora B states reside in unlocked conformations with no ion-pair formation. Nonetheless, the left tails of the distributions give a measurement of the presence of transient hydrogen bond formation. From our plots,  $AURKB^{loop-P}$  seems more prone to facilitate these events, even though their likelihood remains very marginal. By further inspection of the distributions in Figure 3.4.7, it was observed that while the distance distribution in  $AURKB^{no-P}$  is rather regular, the same is not true for  $AURKB^{loop-P}$  and  $AURKB^{full-P}$ . To investigate the possibility that the distributions obtained from simulation are a convolution of two distributions corresponding to two different conformations of the protein, we fitted them with the sum of two Gaussians (Figure 3.4.8).



(a)



(b)

**Figure 3.4.8:** The histograms of probabilities of distances between Glu141:Cδ-Lys122:Nζ AURKB<sup>loop-P</sup> (a) and AURKB<sup>full-P</sup> (b) fitted to a sum of two Gaussian distributions. The black stairs represent the original histograms (Figure 3.4.7), the magenta curve is the sum of two Gaussian distributions for State 1 and 2 represented in blue and orange, respectively.

The two histograms were fitted into the expression below that represents the sum of two Gaussian distributions:

$$H(i) = a_0 e^{-a_1(i-a_2)^2} + a_3 e^{-a_4(i-a_5)^2} \quad (3.4.1)$$

The population of each state is obtained by (normalized) integration of the corresponding Gaussian. Defining the expressions:

$$\begin{aligned}
 I_1 &= a_0 \sqrt{\frac{\pi}{a_1}} \\
 I_2 &= a_3 \sqrt{\frac{\pi}{a_4}} \\
 P(\text{State1}) &= \frac{I_1}{I_1 + I_2}
 \end{aligned}
 \tag{3.4.2}$$

The histograms for the phosphorylated states fit excellent with the bimodal distributions with low chi-square values of 0.008 and 0.001 for AURKB<sup>loop-P</sup> and AURKB<sup>full-P</sup>, respectively. From Figure 3.4.8, the blue and orange graphs represent two conformational states, State1 and State2, confirming that the structures have two main conformations they transition between. The narrow blue graph, State1, indicates that the distance between Lys122:N $\zeta$  and Glu141:C $\delta$  has high probability of being between distances 4.5 to 6.5 Å in AURKB<sup>loop-P</sup> and 5 to 7 Å in AURKB<sup>full-P</sup>. The broad orange graph, State2, indicates that distances between 3 to 8.5 Å in AURKB<sup>loop-P</sup> and 3 to 10 Å in AURKB<sup>full-P</sup> occur with low probability. When the enzymatic complex is in State2, it is implied that the hydrogen bond is able to form transiently, but it is also the conformation where the residues are the farthest apart. Therefore, State1 is more rigid and State2 more dynamic for the phosphorylated states. We also wanted to find the probability that the structure of each phosphorylation state is in State1 or State2.

From the expression above, the probability that AURKB<sup>loop-P</sup> is in State1 and State2 is 44% and 56%, while for AURKB<sup>full-P</sup>, the probability is 49% and 51%, respectively. The structures visit both conformations with more or less the same probability, but AURKB<sup>loop-P</sup> is slightly more in the dynamic State2 which is why the hydrogen bond is formed more frequently in this state.

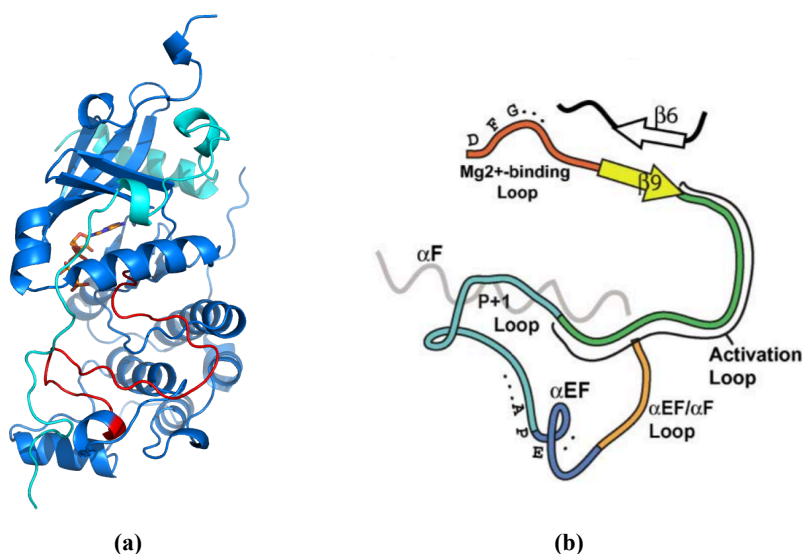
**Cation interference in the salt-bridge formation** The lack of formation of any stable salt bridge between Aurora B<sup>Lys122</sup> and Aurora B<sup>Glu141</sup> is likely connected to the intrusion of Na<sup>+</sup> ions in the ATP binding cavity. In fact, regardless of the phosphorylation state of Aurora B, we systematically observe a rapid diffusion of Na<sup>+</sup> within the first 50-100 ns of MD simulations. The diffusion of positive ions is likely favored by the strong negative charge near the phosphate groups. Conformations of the Glu-Lys ion-pair in relation to ATP are shown in each state in Figure 3.4.6.

In AURKB<sup>no-P</sup>, up to three Na<sup>+</sup> are present at the same time in the pocket, while in AURKB<sup>full-P</sup>, one is mostly observed throughout the simulation in the same orientation as in Figure 3.4.6. The systems were simulated with the physiological salt concentration of 0.15 M, excluding that there could be too much salt in the system. The fact that Na<sup>+</sup> enters the cavity and in all states with similar time scales also suggests that the event has solid physical basis and is not an artifact produced from the simulations' conditions.

With the available data, it is not possible to provide any direct rationalization of the functional role of these ions for Aurora B (if any). Nonetheless, kinases with multiple bound ions have been reported in the past. For example, in the literature, the presence of two  $Mg^{2+}$  increase the binding affinity of ATP to the protein kinase substantially which is important in catalysis [10].

### 3.4.2 Conformation of the Activation Loop

The activation loop of Aurora B is part of the activation segment, which is defined by two conserved tripeptide motifs, DFG-motif ( $Mg^{2+}$  binding) and APE-motif (beginning of  $\alpha$ EF-helix) (Figure 3.4.9 [9]).

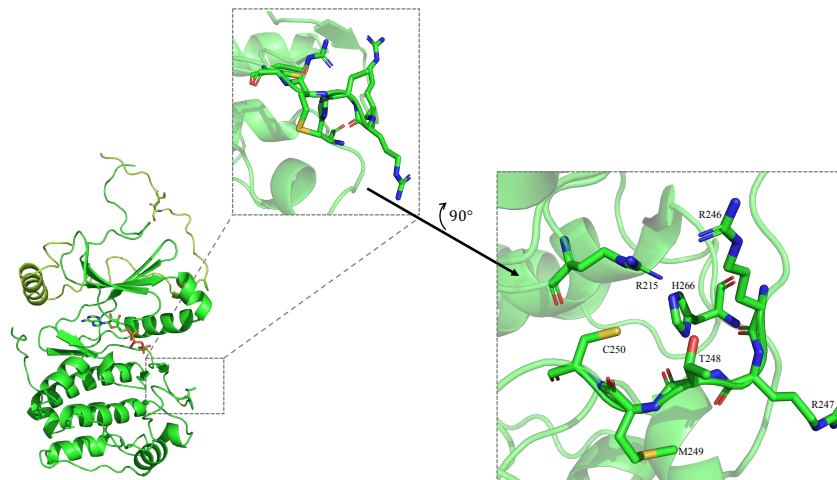


**Figure 3.4.9:** The defined activation segment. (a): Partially phosphorylated Aurora B with activation segment shown in red at end of simulations. (b): The nomenclature of the activation segment from conserved DFG to APE motifs. Modified from Figure 1 in [9].

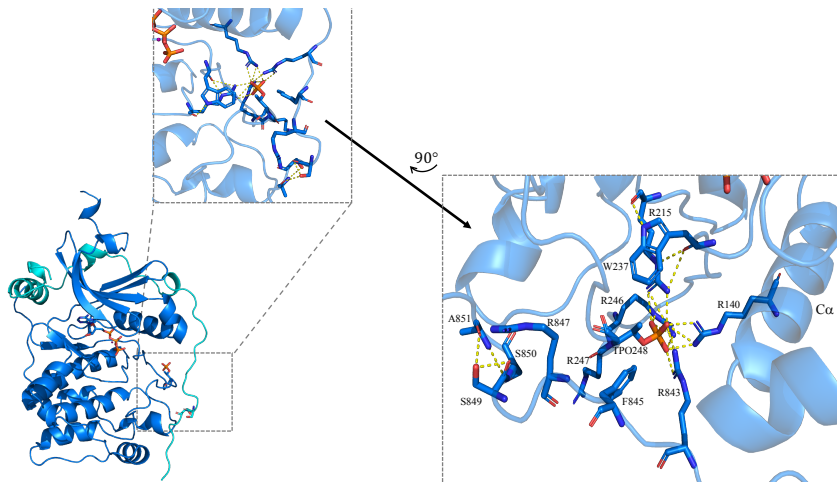
In *X. laevis*, the Aurora B activation loop is comprised of residues (240)HAPSLR-RRTM(249) structured as a flexible loop exposed on the surface of the C-lobe, right next to the active site (Figure 3.4.10). The loop conformation facilitates positioning of the substrate in the catalytically favorable conformation. The activation loop is highly dynamic in all investigated enzyme complex conformations (see RMSF-plot in Figure 3.4.1), which might be because of the absence of substrate in our simulations. Conformational changes in the activation loop in  $AURKB^{no-P}$  are limited, but with phosphorylation of the Aurora B<sup>Thr248</sup> (as present in the  $AURKB^{loop-P}$  and  $AURKB^{full-P}$ ), the loop is stabilized in two different defined conformations (see Figure 3.3.2 middle panel (a)). Furthermore, while in  $AURKB^{loop-P}$ , the enzyme switches between these two conformations in a random manner. In  $AURKB^{full-P}$ , the loop conformation is coupled to opening and closing of the active site and



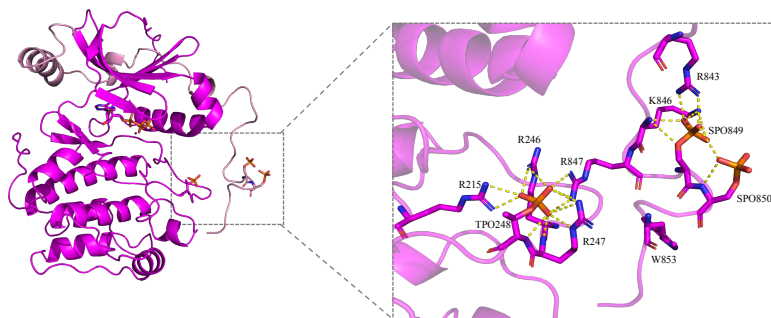
INCENP conformation where the C-terminal region binds there. The activation loop-INCENP interactions are further analyzed and Figure 3.4.10 presents an overview of the whole structures of the states with interactions and structures around the phosphorylated residues. Figure 3.4.11 shows the activation loop region around Aurora B<sup>Thr248</sup> and interactions in the region more in detail.



(a)

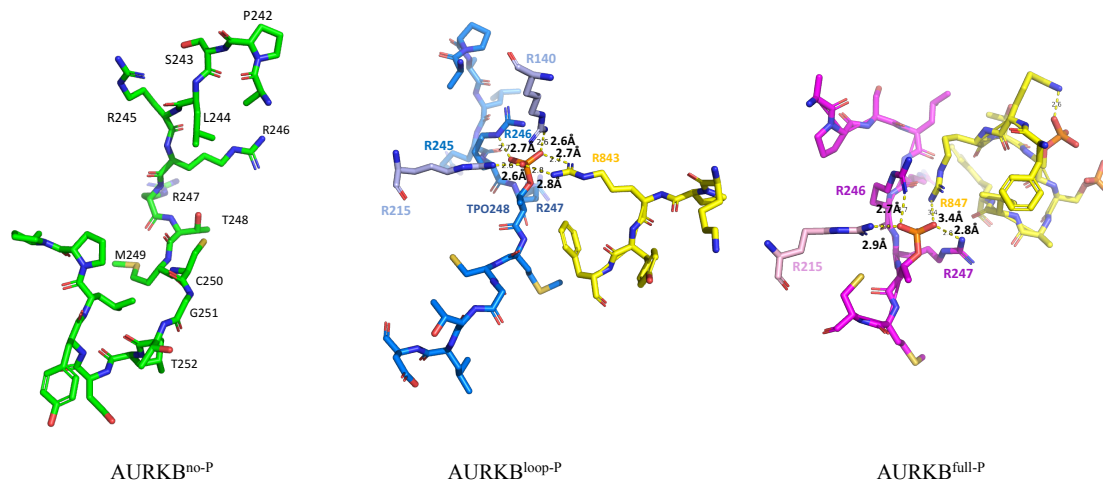


(b)



(c)

**Figure 3.4.10:** Structures of the states at  $2.8 \mu\text{s}$  with loop regions shown in detail with interactions formed by surrounding residues for all states. Aurora B and INCENP are green and yellow, blue and cyan and magenta and pink in  $\text{AURKB}^{\text{no-P}}$ ,  $\text{AURKB}^{\text{loop-P}}$  and  $\text{AURKB}^{\text{full-P}}$ , respectively.

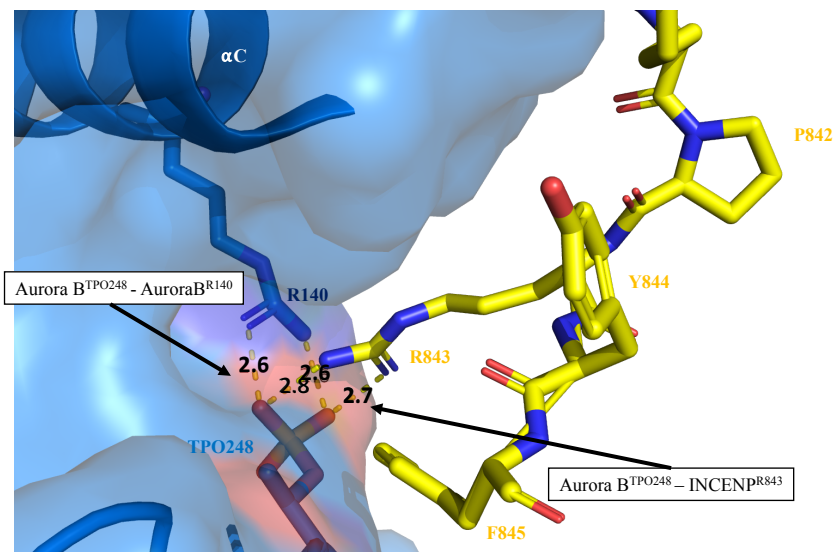


**Figure 3.4.11:** Structures and interactions of the activation loop region around Aurora B<sup>Thr248</sup> shown in all states in stick figure representation at the end of simulations: AURKB<sup>no-P</sup> (green), AURKB<sup>loop-P</sup> (Aurora B blue and INCENP yellow) and AURKB<sup>full-P</sup> (Aurora B magenta and INCENP yellow). Distances of hydrogen bonds in AURKB<sup>loop-P</sup> and AURKB<sup>full-P</sup> are depicted.

In AURKB<sup>no-P</sup>, the loop is in the resting state and no residues are involved in hydrogen bonding. However, when Aurora B<sup>Thr248</sup> in the activation loop becomes phosphorylated, four Arg residues are recruited to stabilize its negative charge in both AURKB<sup>loop-P</sup> and AURKB<sup>full-P</sup> (see Figure 3.4.10). In the final conformation of AURKB<sup>loop-P</sup>, one of these Arg residues is from the activation loop (Aurora B<sup>Arg246</sup>), one is from  $\alpha$ C-helix (Aurora B<sup>Arg140</sup>), one is from the end of the  $\alpha$ E-helix (Aurora B<sup>Arg215</sup>) and one is contributed by INCENP (INCENP<sup>Arg843</sup>) (Figure 3.4.11). In case of AURKB<sup>full-P</sup>, two Arg residues are from the activation loop (Aurora B<sup>Arg246</sup> and Aurora B<sup>Arg247</sup>), one is from the end of the  $\alpha$ E-helix (Aurora B<sup>Arg215</sup>) and again one is contributed by INCENP (INCENP<sup>Arg847</sup>) (Figure 3.4.11). Notably though, a different Arg residue is involved in phosphate shielding in AURKB<sup>loop-P</sup> and AURKB<sup>full-P</sup> conformations. INCENP<sup>Arg843</sup> that stabilizes Aurora B<sup>TPO248</sup> in AURKB<sup>loop-P</sup> is recruited to stabilize phosphorylated TSS-motif on INCENP in AURKB<sup>full-P</sup> conformation. Since INCENP is quite dynamic during simulation (see Figure 3.1.2(b)), we wanted to understand if those interactions observed at the end of simulations are only transient or stable throughout the simulations. We decided to focus on important interactions observed in Figure 3.4.11 and these are considered in detail for AURKB<sup>loop-P</sup> and AURKB<sup>full-P</sup> below.

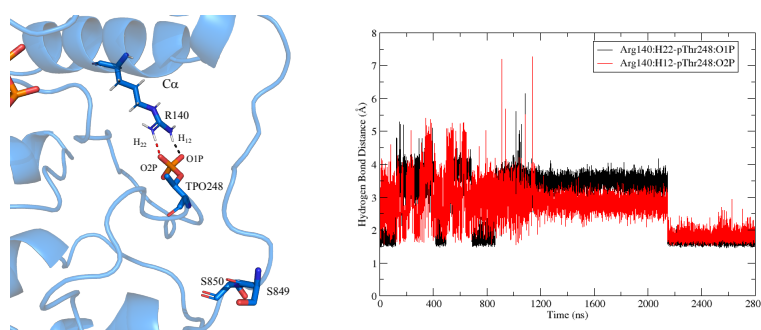
In AURKB<sup>loop-P</sup>, two important interactions were considered (see Figure 3.4.12):

1. Aurora B<sup>TPO248</sup> – Aurora B<sup>Arg140</sup>

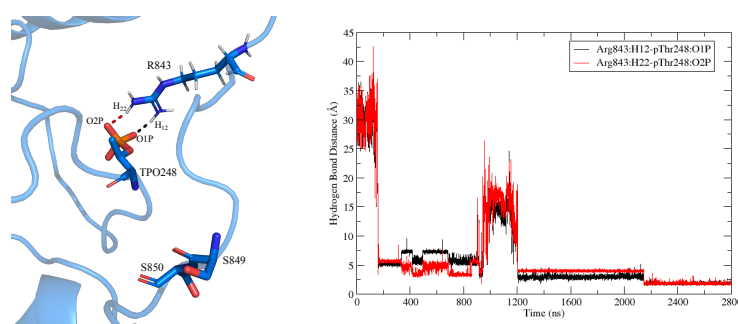
2. Aurora B<sup>TPO248</sup> – INCENP<sup>Arg843</sup>

**Figure 3.4.12:** Interactions between Aurora B<sup>TPO248</sup> – Aurora B<sup>Arg140</sup> and Aurora B<sup>TPO248</sup> – INCENP<sup>Arg843</sup> in AURKB<sup>loop-P</sup> with formed hydrogen bonds and respective distances. INCENP is represented in yellow. Distances are shown in Å.

These two distances were monitored over the course of the simulation of 2.8  $\mu$ s (Figure 3.4.13) and they are considered below.



(a)



(b)

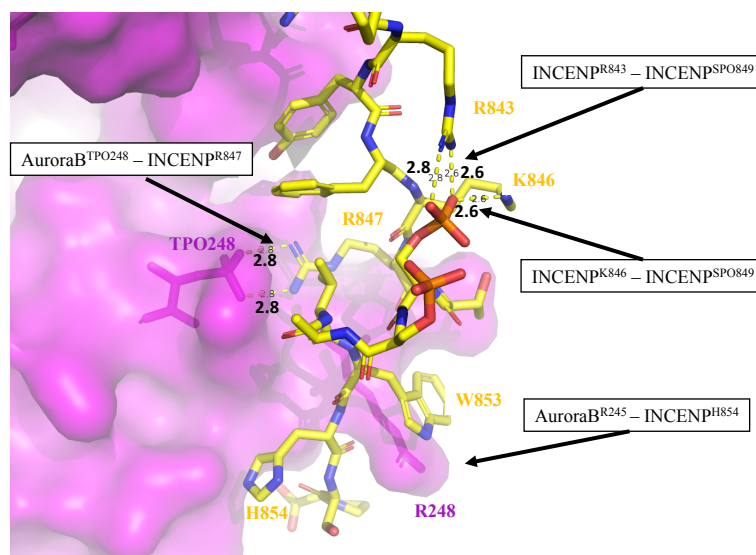
**Figure 3.4.13:** Monitored hydrogen bonds in red and black with corresponding colors in plot of the distance ( $\text{\AA}$ ) over time (ns) in  $\text{AURKB}^{\text{loop-P}}$ . Also to be noticed are the different values on the y-axis ( $\text{\AA}$ ). (a): Hydrogen bonds between Aurora  $B^{\text{TPO248}}$  – Aurora  $B^{\text{Arg140}}$  and (b): Hydrogen bonds between Aurora  $B^{\text{TPO248}}$  –  $\text{INCENP}^{\text{Arg843}}$ .

- **Aurora  $B^{\text{TPO248}}$  – Aurora  $B^{\text{Arg140}}$**  **Figure 3.4.13(a)** – Aurora  $B^{\text{Arg140}}$  is located in the  $\alpha\text{C}$ -helix right next to Aurora  $B^{\text{Glu141}}$  that participates in the Glu-Lys ion-pair (section 3.4.1). Furthermore, Aurora  $B^{\text{Arg140}}$  is evolutionary conserved (see Figure 1.1.6). Therefore, it is likely that Aurora  $B^{\text{Arg140}}$  could have a function in stabilizing Aurora  $B^{\text{TPO248}}$  before the protein would be further phosphorylated and become  $\text{AURKB}^{\text{full-P}}$  where these interactions are not present. In the plot in Figure 3.4.13 on the right, we observe that the distance between Aurora  $B^{\text{TPO248}}$  and Aurora  $B^{\text{Arg140}}$  varies between 1.5  $\text{\AA}$  and 6  $\text{\AA}$  in the first 2.1  $\mu\text{s}$ , indicating that this interaction is transient. Once the bidentate interaction is established it is stable for the last 0.7  $\mu\text{s}$  of simulation.

- **Aurora B<sup>TPO248</sup> – INCENP<sup>Arg843</sup>** **Figure 3.4.13(b)** – The distance between Aurora B<sup>TPO248</sup> – INCENP<sup>Arg843</sup> drops to  $\sim 5$  Å in the beginning of the simulation (after 200 ns) when interactions with both oxygens of the phosphate are established. Distance stays in the range of 3-7 Å for the next  $\sim 800$  ns, then it gets destabilized between 1-1.2  $\mu$ s. Finally, in the last half of the simulation, the distance is reduced to  $\sim 5$  Å again and in the last 600 ns it is forming a true hydrogen bond with under 2.5 Å distance. In summary, analysis indicate that this is a very stable interaction in our simulation.

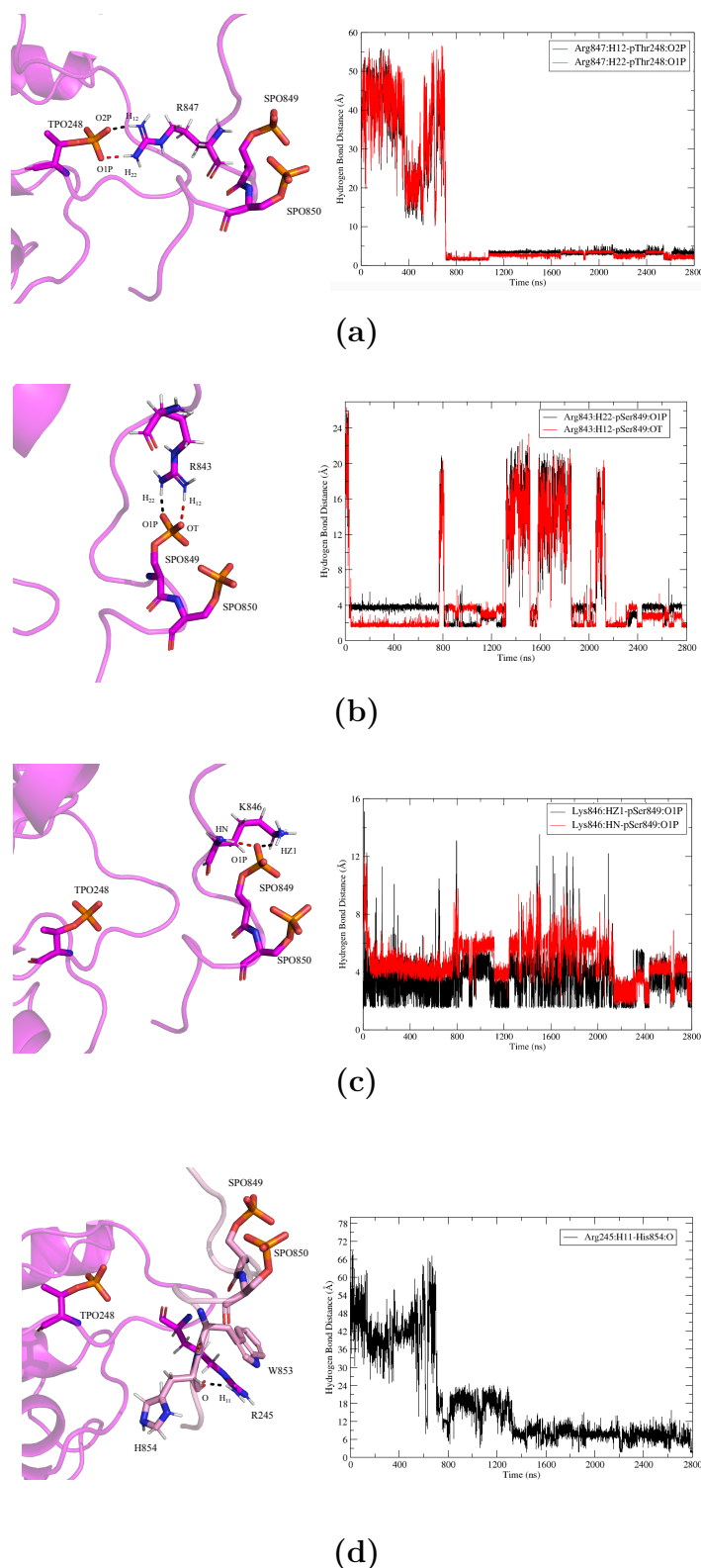
In the analysis of INCENP-activation loop interactions in AURKB<sup>full-P</sup>, we decided to focus on four different interactions (see Figure 3.4.14):

1. Aurora B<sup>TPO248</sup> – INCENP<sup>Arg847</sup>
2. INCENP<sup>Arg843</sup> – INCENP<sup>SPO849</sup>
3. INCENP<sup>Lys846</sup> – INCENP<sup>SPO849</sup>
4. Aurora B<sup>Arg245</sup> – INCENP<sup>His854</sup>



**Figure 3.4.14:** Interactions between Aurora B<sup>TPO248</sup> – INCENP<sup>Arg847</sup>, INCENP<sup>Arg843</sup> – INCENP<sup>SPO849</sup>, INCENP<sup>Lys846</sup> – INCENP<sup>SPO849</sup> and Aurora B<sup>Arg245</sup> – INCENP<sup>His854</sup> in AURKB<sup>full-P</sup> with formed hydrogen bonds and respective distances. INCENP is represented in yellow. Distances are shown in Å.

These four distances were monitored over the course of the simulation of 2.8  $\mu$ s (Figure 3.4.15) and they are considered below.



**Figure 3.4.15:** Monitored hydrogen bonds in red and black with corresponding colors in plot of the distance ( $\text{\AA}$ ) over time (ns) in  $AURKB^{\text{full-P}}$ . Also to be noticed are the different values on the y-axis ( $\text{\AA}$ ). (a): Hydrogen bonds between Aurora B<sup>TPO248</sup> – INCENP<sup>Arg847</sup>, (b): Hydrogen bonds between INCENP<sup>Arg843</sup> – INCENP<sup>SPO849</sup>, (c): Hydrogen bonds between INCENP<sup>Lys846</sup> – INCENP<sup>SPO849</sup> and (d): Hydrogen bond between Aurora B<sup>Arg245</sup> – INCENP<sup>His854</sup>.

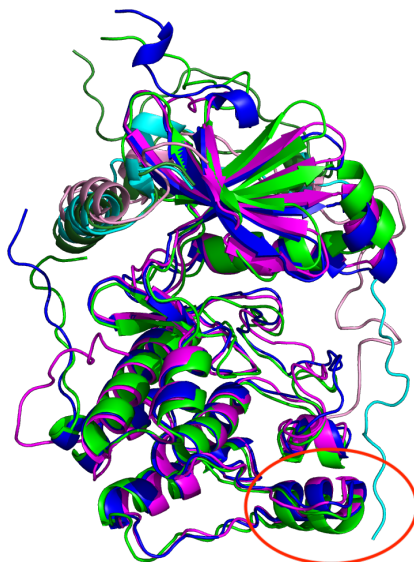
- **Aurora B<sup>TPO248</sup> – INCENP<sup>Arg847</sup> Figure 3.4.15(a)** – INCENP<sup>Arg847</sup> orients into the loop region and forms two stable hydrogen bonds with O2P and O1P of Aurora B<sup>TPO248</sup> around 700 ns into simulation and this interactions stays stable for the rest of the simulation. This indicates that Aurora B<sup>TPO248</sup> is kept in a firm orientation which can also be noted in the RMSF analysis (Figure 3.1.2(a)), indicating that this is a stable and important interaction.
- **INCENP<sup>Arg843</sup> – INCENP<sup>SPO849</sup> Figure 3.4.15(b)** – INCENP<sup>Arg843</sup> in the AURKB<sup>loop-P</sup> is involved in stabilization of the Aurora B<sup>TPO248</sup> in the loop region but in AURKB<sup>full-P</sup>, this Arg residue stabilizes phosphorylated TSS-motif. Again, bidentate interaction between INCENP<sup>Arg843</sup> – INCENP<sup>SPO849</sup> is observed in the final structure. Graph in Figure 3.4.15(a) indicates that distance between these 2 residues is established at the very beginning of the simulation but that it has some periods of instability in the middle of the simulation. However, overall, for most of the simulation at least one hydrogen bond with the distance below 2 Å is being maintained.
- **INCENP<sup>Lys846</sup> – INCENP<sup>SPO849</sup> Figure 3.4.15(c)** – INCENP<sup>Lys846</sup> is somewhat conserved in Aurora B in more developed vertebrates (see Figure 1.1.6) so we reasoned that it could be important in keeping INCENP<sup>SPO849</sup> in a stable orientation towards a structure of enhanced activity. This residue was not observed forming hydrogen bonds in AURKB<sup>loop-P</sup> (Figure 3.4.10) but in AURKB<sup>full-P</sup> it makes hydrogen bonds with INCENP<sup>SPO849</sup> through its main chain and side chain amino groups (NH and NZ1) (Figure 3.4.15(c)). After a closer look at the distance between INCENP<sup>Lys846</sup> – INCENP<sup>SPO849</sup>, it was revealed that both interactions are transient although they got somewhat stabilized towards the end of the simulation.
- **Aurora B<sup>Arg245</sup> – INCENP<sup>His854</sup> Figure 3.4.15(d)** – Aurora B<sup>Arg245</sup> located in the loop in AURKB<sup>full-P</sup> final structure is involved in  $\pi$ -stacking interactions with INCENP<sup>Trp853</sup>. The carbonyl group in the main chain of INCENP<sup>His854</sup> is stabilizing Aurora B<sup>Arg245</sup> in a productive orientation for hydrophobic stacking with INCENP<sup>Trp853</sup>. INCENP<sup>Trp853</sup> is highly conserved in Aurora B and this interaction at the end of INCENP could be important for locking in the INCENP in a conformation where it interacts with the activation loop. Aurora B<sup>Arg245</sup> and INCENP<sup>His854</sup> begin pretty far away but stabilize at 6 Å or shorter in the second half of simulation.

It can be observed from Figure 3.4.10 that the C-terminal end of INCENP interacts in a tight, coiled up conformation in AURKB<sup>full-P</sup>, while in AURKB<sup>loop-P</sup> the end is extended down to the  $\alpha$ G-helix. This raised a question about why the C-terminal end of AURKB<sup>loop-P</sup> interacts with this helix instead of the loop region as in the fully phosphorylated state. The RMSF-plot also presented high fluctuations in this region (Figure 3.1.2). Therefore, we investigated how the  $\alpha$ G-helix is affected by allosteric regulation.

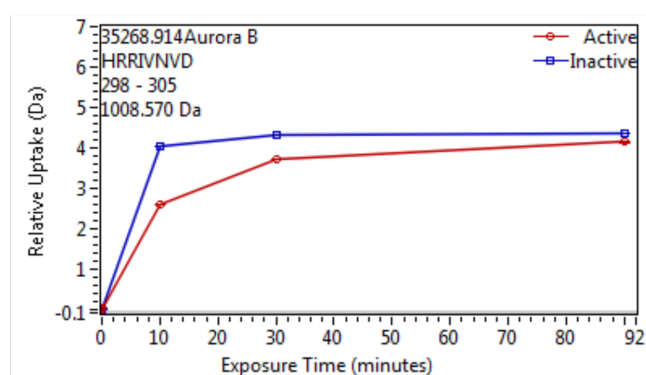
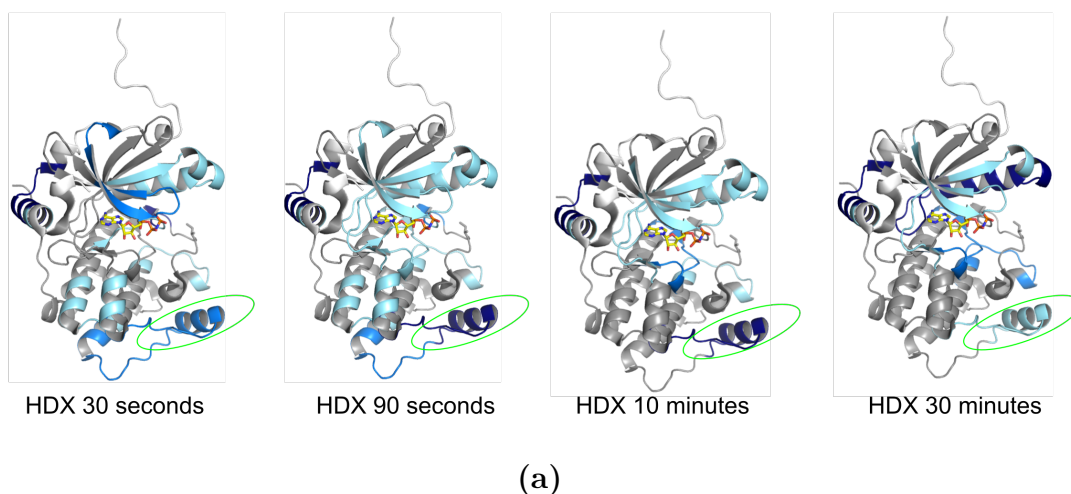


### 3.4.3 $\alpha$ G-Helix

The  $\alpha$ G-helix encompasses Aurora B<sup>Ser293-Val304</sup> and is exposed to the solvent (encircled in red in Figure 3.4.16). The RMSF-plot of Aurora B revealed elevated fluctuations in this region in all states, with AURKB<sup>full-P</sup> being the highest (Figure 3.1.2). The  $\alpha$ G-helix becoming more flexible upon phosphorylation has also been reported by McClendon et. al [75] in other kinases such as protein kinase A (PKA) that has obtained more closed and active conformations. In addition, the flexibility of the  $\alpha$ G-helix of PKA was reported to increase with a ligand present in the active site [75]. On the other hand, hydrogen-deuterium exchange (HDX) experiments indicate that upon phosphorylation of Aurora B<sup>Thr248</sup> in the activation loop,  $\alpha$ G-helix becomes more structured, since the rate of HDX is lower for the fully phosphorylated form as well as the AURKB<sup>loop-P</sup> than for the non-phosphorylated, inhibited form (Figure 3.4.17).



**Figure 3.4.16:** Aligned structures at end of simulations of all states: AURKB<sup>no-P</sup> (green), AURKB<sup>loop-P</sup> (blue) and AURKB<sup>full-P</sup> (magenta). The lighter shade represents their respective bound INCENP. The  $\alpha$ G-helix (residues 293-304) is encircled in red.



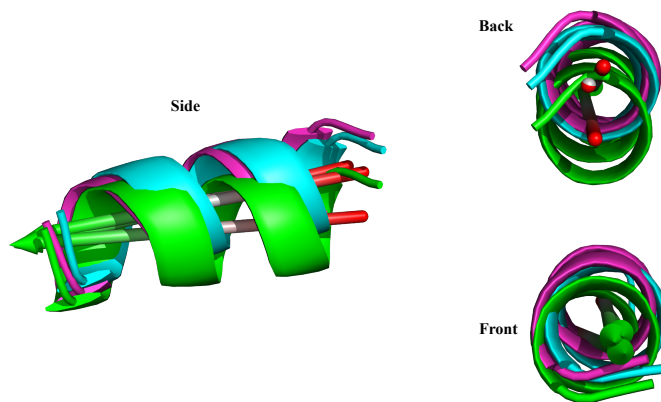
**Figure 3.4.17:**  $AURKB^{full-P}$  is more structured than  $AURKB^{no-P}$ . (a): The difference in deuteration between  $AURKB^{full-P}$  and  $AURKB^{no-P}$  are plotted in ribbon diagrams representing the crystal structure used as starting point for the simulation studies. The regions colored in blue represent parts of the protein where HDX is faster for  $AURKB^{no-P}$  and slower for  $AURKB^{full-P}$ . The darker blue color represents larger difference in deuterium uptake. The  $\alpha$ G-helix is enclosed by the green ellipse. (b): Deuteration uptake plot for a specific peptide of Aurora B covering most of the  $\alpha$ G-helix region (residues 298-305). The red trace represents the uptake for  $AURKB^{full-P}$  and blue trace for  $AURKB^{no-P}$ .

To better understand how our simulations align with literature and experimental data, we analyzed the movements in the  $\alpha$ G-helix throughout simulations for different phosphorylation states. First, we were interested to know if the helix adopts a different position throughout the simulations. Therefore, we measured the angle displacement of the helix between phosphorylated states and  $AURKB^{no-P}$ . The shift was measured by drawing an arrow through the center of the  $\alpha$ G-helix and

then determining the angle formed between the arrows in different phosphorylation forms. The average difference angle between the helices of the last  $\mu$ s was tabulated in Table 3.4.2 and the different shifts shown in Figure 3.4.18.

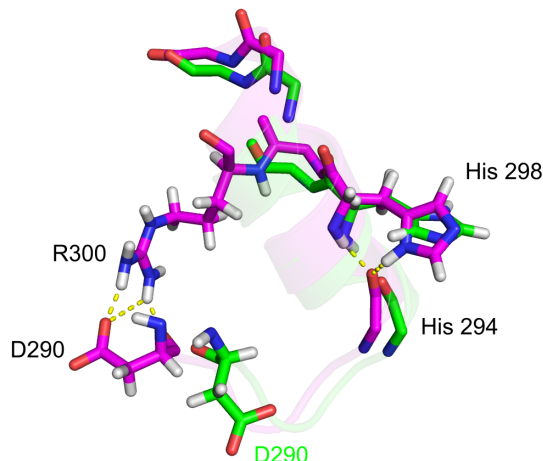
**Table 3.4.2:** Angles between  $\alpha$ G-helix of the states measured for the last  $\mu$ s of simulation. The average angle between the helices and variance are presented.

Compared States	Avg. Angle Between Helices	Variance
AURKB <sup>loop-P</sup> -AURKB <sup>no-P</sup>	$7.526 \pm 3.636$	13.224
AURKB <sup>full-P</sup> -AURKB <sup>no-P</sup>	$8.467 \pm 4.515$	20.389



**Figure 3.4.18:** The  $\alpha$ G-helix in the final structures (AURKB<sup>no-P</sup> (green), AURKB<sup>loop-P</sup> (blue) and AURKB<sup>full-P</sup> (magenta)) are aligned and arrows are drawn through the center of each helix. Helix is shown in front, back and side views to illustrate helix displacement. Notice bigger displacement between unphosphorylated forms of the enzyme in comparison with the phosphorylated enzyme complex.

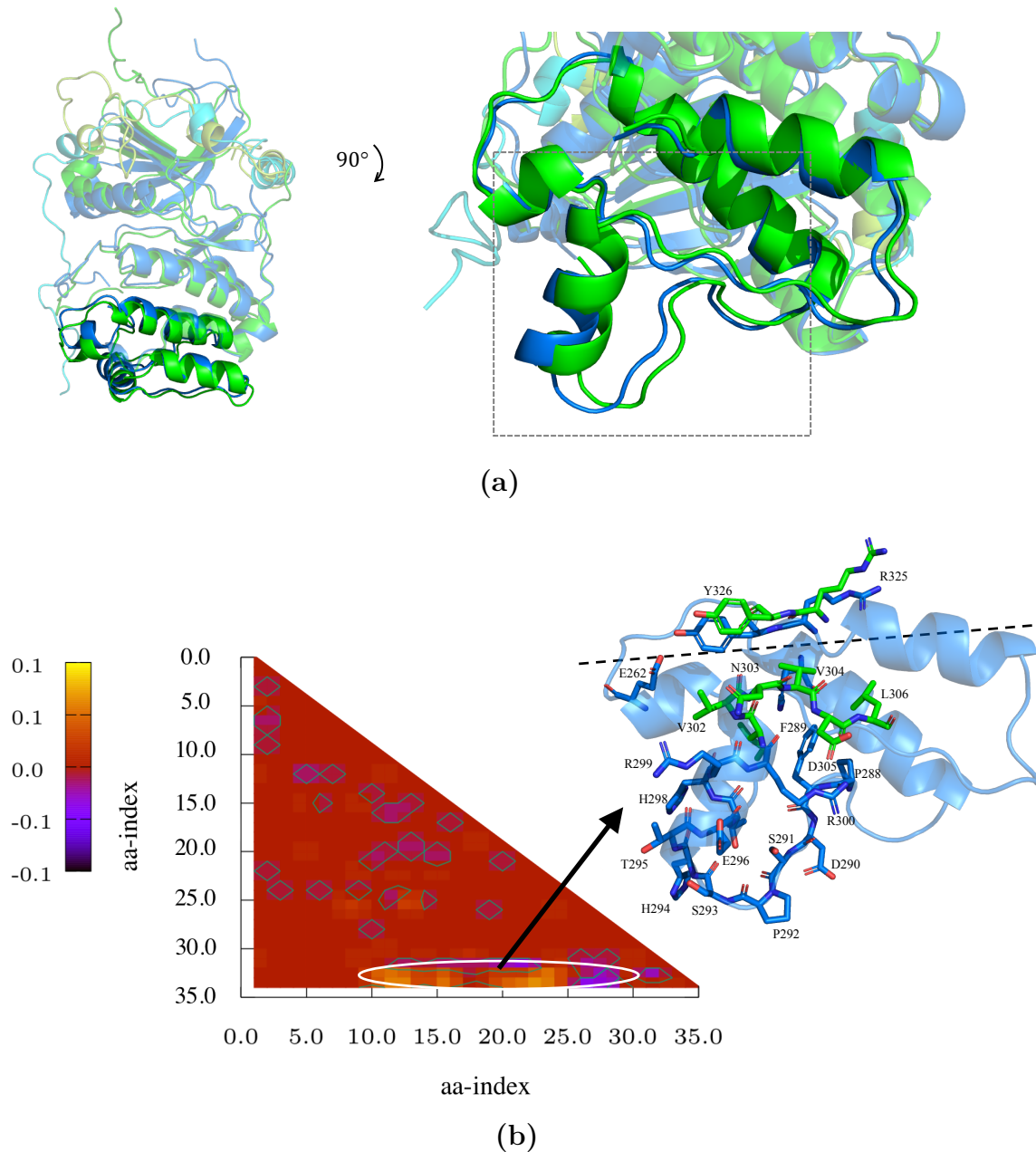
From Table 3.4.2, the largest angle difference and thus the largest shift on average is between AURKB<sup>full-P</sup> and AURKB<sup>no-P</sup>. This difference in helix position could be through stabilization by surrounding residues. We notice that in AURKB<sup>full-P</sup>, the helix is stabilized through hydrogen bonds with the neighboring protein region (Figure 3.4.19). This is consistent with the hydrogen-deuterium exchange results reporting more structured  $\alpha$ G-helix (Figure 3.4.17).



**Figure 3.4.19:** Hydrogen bond stabilizing the  $\alpha$ G-helix in the phosphorylated forms of the enzyme. The helix for  $AURKB^{\text{full-P}}$  and  $AURKB^{\text{no-P}}$  are showed in magenta and green, respectively. The hydrogen bonds showed are only present in the phosphorylated form of the enzyme, probably contributing towards helix stability observed in HDX experiments.

It can be observed in Figure 3.4.16 that for phosphorylated states, the  $\alpha$ G-helix is clearly shifted towards the core of the protein, while in  $AURKB^{\text{no-P}}$ , the helix is further away. Interestingly, the shift might be caused by INCENP being attracted towards this region upon phosphorylation in the activation loop, as observed in the simulations for  $AURKB^{\text{loop-P}}$  and  $AURKB^{\text{full-P}}$  since the  $\alpha$ G-helix is mobile and it is usual for kinases to become more compact upon activation. The shift occurs after around  $1 \mu\text{s}$  in  $AURKB^{\text{loop-P}}$  and  $AURKB^{\text{full-P}}$ , and remains in this orientation for the rest of the simulation. The shift in the  $\alpha$ G-helix therefore occurs almost simultaneously with the C-terminal end of INCENP interacting with the activation loop region.

To further understand the movement of the  $\alpha$ G-helix upon phosphorylation, we performed a contact analysis between unphosphorylated and partially phosphorylated states. First, we identified 35 residues that are in the radius of  $4 \text{ \AA}$  from the  $\alpha$ G-helix (indicated in Figure 3.4.20(a)) and computed residue distances between them (matrix  $35 \times 35$  shown in supplemental data section C.1, Figure C.1.1). To facilitate comparison between the two matrices, we generated a *difference* contact map (Figure 3.4.20). The 35 residues used in calculations are not in order, and their indices are tabulated in Table 3.4.3 for an overview.



**Figure 3.4.20:** (a): Globally aligned proteins of  $AURKB^{loop-P}$  and  $AURKB^{no-P}$  with residues located around  $4 \text{ \AA}$  of the  $\alpha G$ -helix included. On the right, the  $\alpha G$ -helix is further focused in the frame. (b): Difference contact map for the region in (a) with encircled region of gained and lost contacts visualized in the structures on the right. Yellow color indicates that  $AURKB^{loop-P}$  has gained contacts, purple color that  $AURKB^{no-P}$  has gained contacts and red that neither has gained contacts.

**Table 3.4.3:** The 35 residues indices with corresponding residue number in close contact (4 Å) with the  $\alpha$ G-helix.

Residue Index	Residue Protein	Residue Index	Residue Protein	Residue Index	Residue Protein
1	253	13	289	25	301
2	254	14	290	26	302
3	255	15	291	27	303
4	256	16	292	28	304
5	257	17	293	29	305
6	258	18	294	30	306
7	259	19	295	31	307
8	260	20	296	32	308
9	261	21	297	33	324
10	262	22	298	34	325
11	273	23	299	35	326
12	288	24	300		

Most of the distances have not changed between the two states of the protein (red in the Figure 3.4.20). However, the encircled region on the difference distance map indicates that residues Aurora B<sup>Arg325</sup> and Aurora B<sup>Tyr326</sup> (y-axis) in AURKB<sup>loop-P</sup> become closer to residues within the range of Aurora B<sup>262-300</sup> (yellow in Figure 3.4.20(b)) and further away from residues Aurora B<sup>301-306</sup> (purple in Figure 3.4.20(b)). The model next to the graph illustrates where imaginary hinge between the  $\alpha$ G-helix and the rest of the protein is based on this analysis.

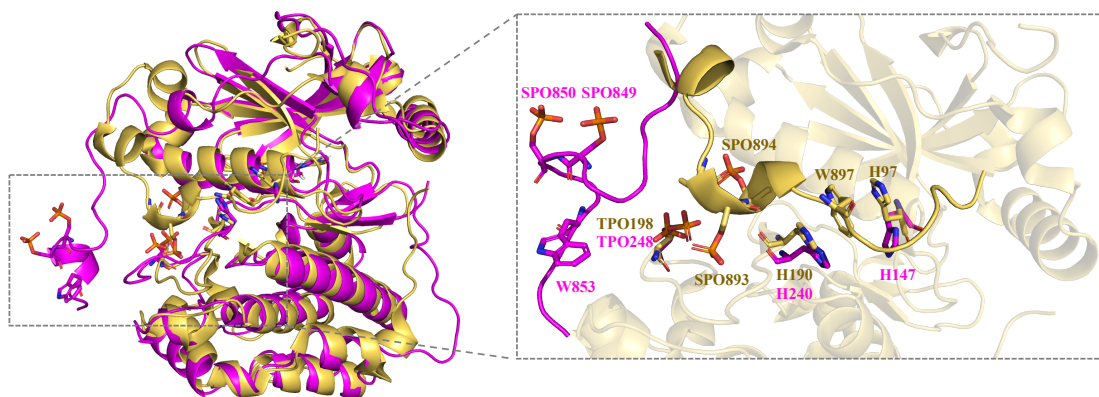
Overall, the analysis show that in AURKB<sup>loop-P</sup>,  $\alpha$ G-helix moves towards the protein core. This movement is probably stabilizing the helix, which is consistent with experimental observations that the helix becomes more structured. The difference in  $\alpha$ G-helix mobility observed in the simulations could be explained through a rigid body motion, where this structural element moves together with other elements of the protein as explained for mode 1 motion deduced from the PCA.

### 3.5 Comparison of fully Phosphorylated Aurora C/INCENP and AURKB<sup>full-P</sup>

During the work of this thesis, a crystal structure of highly similar *H.sapiens* Aurora kinase C in complex with INCENP in a fully phosphorylated state was published (PDB entry: 6GR8<sup>3</sup>[20]), and we have compared it to the final structure generated in my simulation. The sequences have 84% identity so we expected structural similarities. Nevertheless, we still noticed differences in the C-terminal end of INCENP between the two structures. While the C-terminal tail of INCENP

<sup>3</sup><http://www.rcsb.org/structure/6GR8>

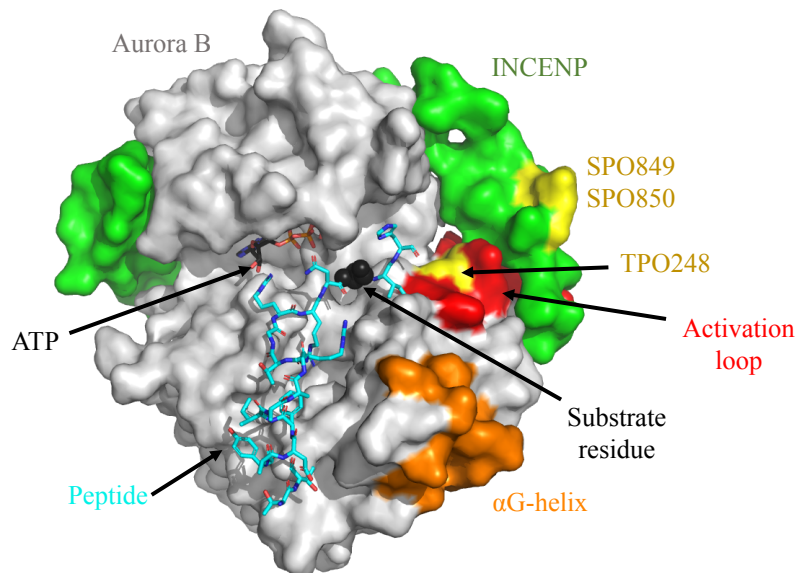
in  $AURKB^{\text{full-P}}$  interacts in the activation loop region near Aurora B<sup>Thr248</sup> (as considered in subsection 3.4.2), in the Aurora C/INCENP structure, C-terminal end of INCENP interacts with the  $\alpha$ C-helix of Aurora C instead. This can be observed in the alignment in Figure 3.5.1.



**Figure 3.5.1:**  $AURKB^{\text{full-P}}$  (magenta) at end of simulation globally aligned to fully phosphorylated Aurora C and INCENP (gold). The region of the C-terminal end of INCENP Aurora B and C are highlighted on the right.

The interaction in Aurora C/INCENP-complex is an aromatic stacking interaction where  $INCENP^{\text{Trp897}}$  is sandwiched between  $INCENP^{\text{His97}}$  and  $INCENP^{\text{His190}}$  (Figure 3.5.1). This interaction, combined with the oriented  $INCENP^{\text{SPO893-SPO894}}$  towards Aurora C<sup>TPO198</sup>, was reported by the authors to be important for keeping the active conformation stable on both sides of  $\alpha$ C-helix and the activation loop. In their work,  $INCENP^{\text{Trp897}}$  of Aurora C was mutated which resulted in destabilization of active conformation of INCENP. In our simulations, this orientation of C-terminal end of INCENP and thus, the stacking interaction, could not be detected. The corresponding Trp in  $AURKB^{\text{full-P}}$ ,  $INCENP^{\text{Trp853}}$ , was observed to form cation- $\pi$  interaction with Arg245 in Figure 3.4.15(d), which could further contribute to keep the orientation of C-terminus of INCENP stable. One of the reasons for the differences between Aurora C/INCENP and our simulation could be because  $AURKB^{\text{full-P}}$  did not reach equilibrium state. In our PCA analysis, the lobes are attempting to close (coupled with loop motion) but with water and  $\text{Na}^+$  ions in the active site, they can not properly do so. It can be speculated that if a substrate peptide had been bound in the complex, we could have overcome an activation barrier and potentially observe a more closed structure similar to the structure of Aurora C/INCENP (Figure 3.1.1). We have modelled a peptide substrate on our final model for  $AURKB^{\text{full-P}}$  based on the structure of inhibitor

peptide observed in *M. musculus* protein kinase A (PKA) (PDB entry: 1ATP<sup>4</sup>)[10]. PKA has similar substrate preference as Aurora B, [R/K]<sub>x</sub>[T/S][I/L/V] [25], and key amino acids recognizing the sequence are conserved between the two (in Aurora B those are Glu177, Glu220, and Glu281).

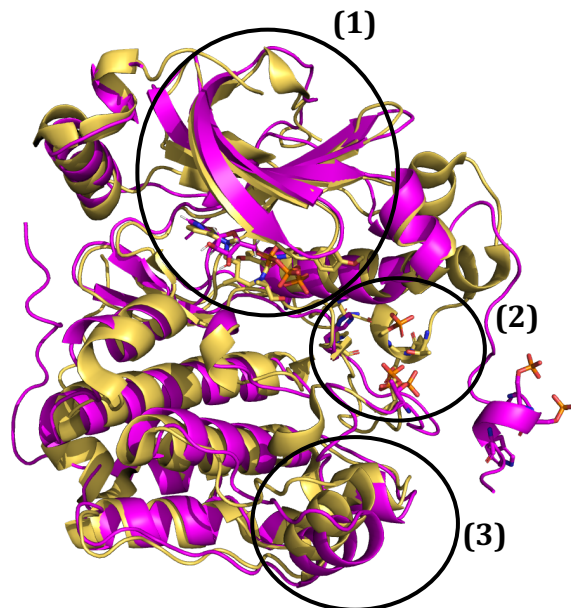


**Figure 3.5.2:** *AURKB<sup>full-P</sup>*, with Aurora B (grey) and INCENP (green) with structural regions  $\alpha$ G-helix (orange), activation loop (red) and phosphorylated residues (TPO248, SPO849 and SPO850) highlighted in yellow on the figure. The peptide modelled from PKA structure (cyan) is in sticks. The Ser/Thr (in the original structure mutated to Ala) that accepts the phosphate transfer is shown in black.

If this peptide was bound in our structures, it can be speculated that a conformation similar to Aurora C could have been reached. Interestingly, the same PKA structure has two  $Mn^{2+}$  ions in the active site in similar positions as observed  $Mg^{2+}$  and  $Na^{+}$  in our structure, so perhaps we would have observed formation of the ion-pair in the fully phosphorylated state with peptide present. Even though there are differences between the structure of Aurora C/INCENP and *AURKB<sup>full-P</sup>*, there are mostly similar regions. Three of them are encircled in Figure 3.5.3.

<sup>4</sup><http://www.rcsb.org/structure/1ATP>





**Figure 3.5.3:**  $AURKB^{full-P}$  (magenta) at the end of simulation aligned globally to fully phosphorylated Aurora C (gold) with encircled regions of the active site (1), activation loop region (2) and  $\alpha G$ -helix (3).

Observed in Figure 3.5.3, circle (1) represents the  $\beta$ -hairpin  $\beta_1$ -P-loop- $\beta_2$  that was unstructured in the unphosphorylated state (see Figure 3.4.2). Here, the region is structured and align well, suggesting similar active sites keeping ATP stable as expected since ATP binding sites of Aurora kinases are 100% conserved [20]. In circle (2), phosphorylated Thr substrate residue of Aurora B and C are kept in the same position in the activation loops, but are oriented differently towards the corresponding phosphorylated Ser residues of INCENP. The orientation of the Ser residues of INCENP in Aurora C can be favored by the conserved Arg-rich region around Aurora C<sup>TPO198</sup>. The structural difference of INCENP in the two Aurora kinases can thus affect the activation loop to become structurally different. In circle (3), it can be observed that the  $\alpha G$ -helix is more shifted towards the protein in Aurora C than in Aurora B. This shift is expected since Aurora C represents a fully active state and we observed an increasing shift of the  $\alpha G$ -helix with the state of phosphorylation. Together, these structural regions support that the evaluated regions in this study were rational. The general structural differences observed between Aurora B and C in Figure 3.5.3 can also be related to the fact that they origin from different organisms (*X. laevis* for Aurora B/INCENP and *H. sapiens* for Aurora C/INCENP).

# 4

## Conclusion & Future Research

The protein kinase activity of Aurora B is critical for all its mitotic regulatory functions. The major regulatory mechanism of Aurora B enzymatic activity is through auto-phosphorylation of its activation loop and of interacting partner INCENP. The present study is aimed to provide a better understanding of the role of phosphorylation in the catalytic properties of the Aurora B/INCENP enzymatic complex. Several studies [25] [33] [20] have established the requirement of phosphorylation to achieve full Aurora B enzymatic activity, and efforts to understand the structural basis of this phenomenon resulted in several crystal structures used to propose Aurora B activation models [20] [25]. A major gap in the field has been the lack of structural models for the enzyme in the dephosphorylated state as well as in the fully phosphorylated state. At the beginning of this thesis, all the available structures were representative of a partially phosphorylated form of the enzyme, with the phospho-groups present only in the Aurora B activation loop but absent in INCENP's TSS-motif. A structure of the functionally similar Aurora C/INCENP was published [20] during this thesis.

To understand how dynamics of the enzymatic complex depend on phosphorylation we have used the computational approach – molecular dynamics (MD) simulations. MD simulations were performed on three states of the enzyme: the non-phosphorylated, the partially phosphorylated (with only the activation loop bearing the phospho-mark) and the fully phosphorylated form. These different forms of the enzyme are referred to as  $\text{AURKB}^{\text{no-P}}$ ,  $\text{AURKB}^{\text{loop-P}}$  and  $\text{AURKB}^{\text{full-P}}$  throughout this thesis, respectively. The MD simulations have produced a set of structural models, for each of the mentioned phosphorylation states, during a simulation of 2.8  $\mu\text{s}$  that we have further analyzed.

The most dramatic conformational change between the three states is the C-terminal region of INCENP. In AURKB<sup>no-P</sup>, the C-terminal region of INCENP interacts with the solvent and the N-terminus of Aurora B. Interestingly, phosphorylation of Aurora B activation loop leads to a conformational change in the C-terminal region of INCENP bringing it into close proximity with phospho-threonine in the activation loop where it stays throughout the simulation. This observation is supported by the recent crystal structures of Aurora C/INCENP and explains how INCENP helps shape the active site of the enzyme complex [20].

A hallmark of the protein kinase fold are N- and C-terminal lobes with the active site located at the interface between the two. The two lobes are moving synchronously, generating open and closed conformations [72]. In our principal component analysis (PCA), we observe opening and closing of the active site for all three phosphorylation states of the enzyme, but this motion becomes very pronounced and also synchronized with the motion in the activation loop only in the fully phosphorylated form. This also provides an explanation for the high RMSF of AURKB<sup>full-P</sup> being due to global motions instead of increased local flexibility. Coupling activation loop conformation with open-close motion of the active site is thus an important feature that might contribute towards increased activity in the AURKB<sup>full-P</sup> form of the enzyme complex.

The structural alignment of AURKB<sup>no-P</sup>, AURKB<sup>loop-P</sup> and AURKB<sup>full-P</sup> during the last stages of the simulations and hydrogen-deuterium exchange (HDX)-analysis, reveal generally more structured phosphorylated forms of the enzyme. Even though MD simulations and HDX-analysis are not directly comparable methods due to the difference in time scales (HDX in seconds and MD in ns), the analysis can still be complemented to provide insights into the activation mechanism of Aurora B. With RMSF as point of departure, we identified regions with structural changes taking place. Therefore, we further analyzed the ATP binding site, the activation loop and the  $\alpha$ G-helix.

The ATP molecule in AURKB<sup>no-P</sup> adopts multiple conformations, most of them non-productive for catalysis, indicating a non-structured binding pocket. Phosphorylation of Aurora B<sup>Thr248</sup> in the activation loop results in stabilization of the ATP binding site, so in AURKB<sup>loop-P</sup> and AURKB<sup>full-P</sup> enzyme complexes ATP binds in a more productive conformation.

We do not observe the formation of the ion-pair Lys-Glu in the active site reported to be associated with the active conformation in similar protein kinases in any of the phosphorylation states. Although, we observe reduced distance between these residues in the AURKB<sup>loop-P</sup>. At the same time, in the AURKB<sup>full-P</sup> we observe Na<sup>+</sup> ion coordinating Aurora B<sup>Glu141</sup> and  $\gamma$ -phosphate of ATP. The systematic presence of Na<sup>+</sup> in the binding cavity seems to be a physical feature of the substrate-unbound state. Nonetheless, we cannot rule out that it may be also an artifact of the force field, with an overestimation of the electrostatic attraction forces on the free electrolyte. Further investigation is necessary to resolve if this observation has any physiological relevance for Aurora B/INCENP-complex.

The activation loop of Aurora B becomes more structured upon phosphorylation of Aurora B<sup>Thr248</sup>. The loop gets reorganized, so that positively charged arginines surround and shield phosphate on Aurora B<sup>Thr248</sup>. In both AURKB<sup>loop-P</sup> and AURKB<sup>full-P</sup>, INCENP<sup>Arg843</sup> and INCENP<sup>Arg847</sup>, contribute to this shielding, respectively. The rigidification in the loop region is consistent with the reported crystal structures of several kinases, including Aurora B, as well as with the unpublished HDX studies.

We analyzed movements of the  $\alpha$ G-helix in the enzyme complex as a consequence of phosphorylation and because we observe a pronounced difference in stability of this helix in HDX when activation loop is phosphorylated (helix becomes more stable). MD analysis confirmed that  $\alpha$ G-helix in unphosphorylated structure resides further away from the rest of the Aurora B.

We have pointed out regions showing properties characteristic of a more catalytically competent kinase and by comparison with fully phosphorylated Aurora C/INCENP-structure, we noticed a different C-terminal end from AURKB<sup>full-P</sup>. It can be speculated that if the phosphorylated structures had been simulated with a bound peptide, the fully phosphorylated structure could have been more closed with a higher resemblance towards Aurora C and perhaps formed the ion-pair in the active site.

In summary, we can conclude that phosphorylation have a profound impact both in the structure and the dynamics of the Aurora B/INCENP enzymatic complex, creating a clear allosteric communication between INCENP and Aurora B. The phosphorylation of both Aurora B and INCENP results in a positive synergistic effect in the enzymatic activity as have been showed with enzyme kinetics experiments (unpublished data Sekulic lab).

High RMSF in the AURKB<sup>full-P</sup> indicates that the system might still not have reached a state of equilibrium. To obtain an even more accurate representation of the molecular motions for the enzyme complex in the fully phosphorylated state, representing the highest activity, future studies could take advantage of generating a starting model from recently reported fully phosphorylated structure of Aurora C/INCENP, including a substrate peptide in the active site and/or conducting longer simulation. A further study could also set out to simulate the catalytic mechanism of Aurora B and attempt to observe the  $\gamma$ -phosphate transfer from ATP to the substrate Thr/Ser residues. In order to study this chemical process, a quantum mechanics/molecular mechanics (QM/MM) hybrid approach can be applied.

The results derived from this thesis will help experimental scientists rationalize results and design new experiments aimed at understanding the function of this critical mitotic kinase. Together, theoretical and experimental studies will generate new opportunities to find a better and more specific inhibitor of Aurora B, that might help in future cancer treatment.



# Abbreviations and Units

## A.1 List of Abbreviations

<b>AURKA</b>	Aurora Kinase A
<b>AURKB</b>	Aurora Kinase B
<b>AURKC</b>	Aurora Kinase C
<b>CDK</b>	Cyclin-Dependent Kinase
<b>CHARMM</b>	Chemistry at HARvard Macromolecular Mechanics
<b>CML</b>	Chronic Myelogenous Leukemia
<b>CPC</b>	Chromosomal Passenger Complex
<b>DOF</b>	Degrees of Freedom
<b>EGFR</b>	Epidermal Growth Factor Receptor
<b>GROMACS</b>	GRoningen MACHine for Chemical Simulations
<b>HDX</b>	Hydrogen-Deuterium Exchange
<b>INCENP</b>	Inner Centromere Protein
<b>MD</b>	Molecular Dynamics
<b>MS</b>	Mass Spectrometry
<b>NMR</b>	Nuclear Magnetic Resonance

<b>PBC</b>	Periodic Boundary Conditions
<b>PCA</b>	Principal Component Analysis
<b>PK</b>	Protein Kinase
<b>PKA</b>	Protein Kinase A
<b>RMSD</b>	Root-Mean-Square Deviation
<b>RMSF</b>	Root-Mean-Square Fluctuation
<b>VMD</b>	Visual Molecular Dynamics

## A.2 Units

---

M	molar (1 mol/L)
$\mu$ s	microsecond ( $10^{-6}$ s)
ns	nanosecond ( $10^{-9}$ s)
ps	picosecond ( $10^{-12}$ s)
Å	Ångström ( $10^{-10}$ m)

---

### A.3 Overview of Amino Acids

An overview of the 20 different amino acids with their respective three- and one-letter codes.

Amino Acid	3-letter code	1-letter code
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartate	Asp	D
Cysteine	Cys	C
Glutamate	Glu	E
Glutamine	Gln	Q
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

# B

## GROMACS Commands

### B.1 Overview of GROMACS Commands

GROMACS commands used to analyze MD simulations with GROMACS are listed below and can also be found in the GROMACS documentation<sup>1</sup>. For commands related to the execution of simulations, see GROMACS manual [41] and article by Justin A. Lemkul [47]. In the latter article, a range of tutorials for simulations in GROMACS are presented where *Tutorial 1: Lysozyme in Water* is similar to the workflow in our MD simulations [47].


**anaeig** Analyzes a selection of eigenvectors generated with *covar*. The output is the extreme projections of the trajectory and the eigenvectors written to respective .pdb-files. Used to create Figure 3.3.2 and movies of the PCA motion available from GitHub: *odasho/Master-thesis/movies\_PCA\_states*<sup>2</sup>.

**covar** Generates and diagonalizes the covariance matrix of  $\sim 753 \times \sim 753$  from the simulation. The output is a matrix consisting of eigenvectors, the sum of eigenvalues, and trajectory file. Used in PCA to make Figure 3.3.1 representing the eigenvalues over eigenvector index.

**hbond** To calculate the average number of hydrogen bonds between two groups in a trajectory, in this thesis, between ATP and surrounding residues and water molecules. The output is the average number of hydrogen bonds

---

<sup>1</sup>GROMACS commands: <http://manual.gromacs.org/documentation/5.1/user-guide/cmdline.html>

<sup>2</sup>  [odasho/Master-thesis](https://github.com/odasho/Master-thesis)



between the groups with standard deviation. The command was used to create Table 3.4.1 for average hydrogen bonds in each state between ATP and residues/water.

**make\_ndx** Creates special index groups from simulation structure files. Used widely in the study for selecting specific residue groups or atoms for further study.

**mdmat** Creates a distance matrix of the smallest distance between residue pairs. Used to create a distance matrix contact map for the  $\alpha$ G-helix region to be processed with *xpm2ps*.

**rms** Computes the RMSD between a reference and a simulated structure. The output is a plot with the RMSD-values over simulation time. The command was used to create Figure 3.1.1.

**rmsf** Computes the RMSF of positions in the trajectory after fitting to a reference. The output is a plot with the RMSF-values over positions. Used to create Figure 3.1.2 with RMSF-values over  $C_\alpha$  of residues.

**trjconv** Primarily modifies simulation trajectory files. *trjconv* can convert trajectory files into other formats, center atoms in simulation box, select a subset of atoms, change time steps and select specific frames from trajectory. Used in the study for generating trajectory files to be viewed in VMD and pdb-files of specific frames to be viewed in PyMOL.

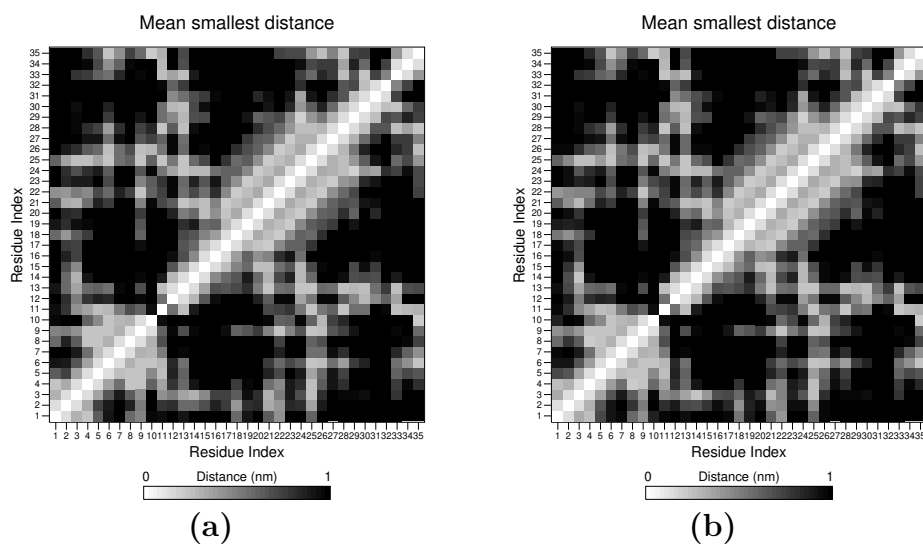
**xpm2ps** Creates a plot from .xpm-file generated by *mdmat*. Used to create plots Figure C.1.1.

# C

## Supplemental Figures

### C.1 Contact Maps

Here we present the contact maps made from distance matrices used for making the difference contact map for the  $\alpha$ G-helix shown in Figure 3.4.20 for states  $AURKB^{\text{no-P}}$  and  $AURKB^{\text{loop-P}}$ .



**Figure C.1.1:**  $35 \times 35$  distance matrices between nearby residues (see Table 3.4.3) for residue index in state  $AURKB^{\text{no-P}}$  in (a), and state  $AURKB^{\text{loop-P}}$  in (b) with truncation cut-off 1 nm. The range of white to black corresponds to shortest to longest distances between residues.

# References

- [1] Ann H. West Alla O. Kaserer. *Histidine Kinase*. URL: <https://www.sciencedirect.com/topics/biochemistry-genetics-and-molecular-biology/histidine-kinase> (visited on 06/04/2020) (cit. on p. 1).
- [2] Jane Endicott, Martin Noble, and Louise Johnson. “The Structural Basis for Control of Eukaryotic Protein Kinases.” In: *Annual review of biochemistry* 81 (Apr. 2012), pp. 587–613. DOI: 10.1146/annurev-biochem-052410-090317 (cit. on pp. 1, 2, 4, 5).
- [3] Manning GD et al. “The Protein Kinase Complement of the Human Genome.” In: *Science (New York, N.Y.)* 298 (Jan. 2003), pp. 1912–34. DOI: 10.1126/science.1075762 (cit. on pp. 1, 2).
- [4] Bostjan Kobe and Bruce Kemp. “Chapter 74. Principles of Kinase Regulation.” In: *Handbook of Cell Signaling, 2/e 2* (Dec. 2010), pp. 559–562. DOI: 10.1016/B978-0-12-374145-5.00074-7 (cit. on pp. 1, 2, 10).
- [5] Lauren Pecorino. *The Molecular Biology of Cancer: Mechanisms, Targets, and Therapeutics*. Jan. 2012, pp. 78–81, 80–83, 98–100 (cit. on pp. 1, 2, 4, 5).
- [6] Daisuke Kitagawa et al. “Activity-based kinase profiling of approved tyrosine kinase inhibitors.” In: *Genes to cells : devoted to molecular & cellular mechanisms* 18 (Dec. 2012). DOI: 10.1111/gtc.12022 (cit. on p. 2).
- [7] KinHub. *Human Kinome Tree*. URL: <http://www.kinhub.org/kinmap/> (visited on 05/29/2020) (cit. on p. 4).
- [8] kinase.com. *Introduction to Kinases*. URL: [http://kinase.com/wiki/index.php/Introduction\\_to\\_Kinases](http://kinase.com/wiki/index.php/Introduction_to_Kinases) (visited on 04/30/2020) (cit. on p. 4).
- [9] Brad Nolen, Susan Taylor, and Gourisankar Ghosh. “Regulation of protein kinases: Controlling activity through activation segment conformation.” In: *Molecular cell* 15 (Oct. 2004), pp. 661–75. DOI: 10.1016/j.molcel.2004.08.024 (cit. on pp. 5, 45).
- [10] J. Zheng et al. “2.2 A refined crystal structure of the catalytic subunit of cAMP-dependent protein kinase complexed with MnATP and a peptide inhibitor.” In: *Acta crystallographica. Section D, Biological crystallography* 49 (June 1993), pp. 362–5. DOI: 10.1107/S0907444993000423 (cit. on pp. 5, 39, 45, 61).
- [11] Hendrik L. De Bondt et al. “Crystal structure of cyclin-dependent kinase 2.” In: *Nature* 363.6430 (1993), p. 595. ISSN: 0028-0836 (cit. on p. 5).

- [12] Dorian Fabbro, Sandra Cowan-Jacob, and Henrik Möbitz. “Ten things you should know about protein kinases: IUPHAR Review 14.” In: *British journal of pharmacology* 172 (Jan. 2015). DOI: 10.1111/bph.13096 (cit. on p. 5).
- [13] Robert Roskoski. “Properties of FDA-approved small molecule protein kinase inhibitors: A 2020 update.” In: *Pharmacological Research* 152 (Dec. 2019), p. 104609. DOI: 10.1016/j.phrs.2019.104609 (cit. on p. 5).
- [14] Brian Druker et al. “Efficacy and Safety of a Specific Inhibitor of the BCR-ABL Tyrosine Kinase in Chronic Myeloid Leukemia.” In: *The New England journal of medicine* 344 (May 2001), pp. 1031–7. DOI: 10.1056/NEJM200104053441401 (cit. on p. 5).
- [15] K. B. et al. “Kinase-targeted cancer therapies: Progress, challenges and future directions.” In: *Molecular Cancer* 17 (Dec. 2017). DOI: 10.1186/s12943-018-0804-2 (cit. on p. 5).
- [16] Georgina Long et al. “Adjuvant Dabrafenib plus Trametinib in Stage III BRAF -Mutated Melanoma.” In: *New England Journal of Medicine* 377 (Sept. 2017). DOI: 10.1056/NEJMoa1708539 (cit. on p. 5).
- [17] Mar Carmena, William Earnshaw, and David Glover. “The Dawn of Aurora Kinase Research: From Fly Genetics to the Clinic.” In: *Frontiers in Cell and Developmental Biology* 3 (Nov. 2015). DOI: 10.3389/fcell.2015.00073 (cit. on pp. 5–7).
- [18] Xin Zhang. “Aurora kinases.” eng. In: *Current Biology* 18.4 (2008), R146–R148. ISSN: 0960-9822 (cit. on p. 6).
- [19] Richard Bayliss et al. “Structural Basis of Aurora-A Activation by TPX2 at the Mitotic Spindle.” In: *Molecular cell* 12 (Nov. 2003), pp. 851–62. DOI: 10.1016/S1097-2765(03)00392-7 (cit. on p. 6).
- [20] Kamal Abdul Azeez et al. “Structural mechanism of synergistic activation of Aurora kinase B/C by phosphorylated INCENP.” In: *Nature Communications* 10 (Dec. 2019). DOI: 10.1038/s41467-019-11085-0 (cit. on pp. 6, 10, 11, 31, 59, 62–64).
- [21] UniProt. *Q96GD4 (AURKB\_HUMAN)*. URL: <https://www.uniprot.org/uniprot/Q96GD4> (visited on 05/31/2020) (cit. on p. 6).
- [22] UniProt. *Q9UQB9 (AURKC\_HUMAN)*. URL: <https://www.uniprot.org/uniprot/Q9UQB9> (visited on 05/31/2020) (cit. on p. 6).
- [23] Sandrine Ruchaud, Mar Carmena, and William Earnshaw. “Chromosomal passengers: Conducting cell division.” In: *Nature reviews. Molecular cell biology* 8 (Nov. 2007), pp. 798–812. DOI: 10.1038/nrm2257 (cit. on pp. 6, 7).
- [24] Veronica Krenn and Andrea Musacchio. “The Aurora B Kinase in Chromosome Bi-Oriented and Spindle Checkpoint Signaling.” In: *Frontiers in oncology* 5 (Nov. 2015), p. 225. DOI: 10.3389/fonc.2015.00225 (cit. on pp. 6, 7).
- [25] Fabio Sessa et al. “Mechanism of Aurora B Activation by INCENP and Inhibition by Hesperadin.” In: *Molecular cell* 18 (May 2005), pp. 379–91. DOI: 10.1016/j.molcel.2005.03.031 (cit. on pp. 6, 7, 10, 11, 23, 30, 35, 40, 61, 63).
- [26] Dr. Arockia Jeyaprakash Arulanandam et al. “Structure of a Survivin–Borealin–INCENP Core Complex Reveals How Chromosomal Passengers Travel To-

- gether.” In: *Cell* 131 (Nov. 2007), pp. 271–85. DOI: 10.1016/j.cell.2007.07.045 (cit. on p. 7).
- [27] Fabio Sessa and Fabrizio Villa. “Structure of Aurora B–INCENP in complex with barasertib reveals a potential transinhibitory mechanism.” In: *Acta crystallographica. Section F, Structural biology communications* 70 (Mar. 2014), pp. 294–8. DOI: 10.1107/S2053230X14002118 (cit. on pp. 8, 9, 23, 31).
- [28] Morgan Huse and John Kuriyan. “The Conformational Plasticity of Protein Kinases.” In: *Cell* 109 (June 2002), pp. 275–82. DOI: 10.1016/S0092-8674(02)00741-9 (cit. on pp. 10, 40).
- [29] Jonah Beenstock, Navit Mooshayef, and David Engelberg. “How Do Protein Kinases Take a Selfie (Autophosphorylate)?” In: *Trends in biochemical sciences* 41 (Sept. 2016). DOI: 10.1016/j.tibs.2016.08.006 (cit. on pp. 10, 39).
- [30] M Kimura et al. “Cell cycle-dependent expression and centrosome localization of a third human aurora/Ipl1-related protein kinase, AIK3.” eng. In: *The Journal of biological chemistry* 274.11 (1999), pp. 7334–7340. ISSN: 0021-9258. URL: <http://search.proquest.com/docview/69607796/> (cit. on p. 11).
- [31] Gerben Vader and Susanne M.A Lens. “The Aurora kinase family in cell division and cancer.” eng. In: *BBA - Reviews on Cancer* 1786.1 (2008), pp. 60–72. ISSN: 0304-419X (cit. on p. 11).
- [32] Estelle Willems et al. “The functional diversity of Aurora kinases: A comprehensive review.” In: *Cell Division* 13 (Dec. 2018). DOI: 10.1186/s13008-018-0040-6 (cit. on p. 11).
- [33] Av Zaytsev et al. “Bistability of a coupled Aurora B kinase-phosphatase system in cell division.” English. In: *Molecular Biology Of The Cell* 26 (2015). ISSN: 1059-1524 (cit. on pp. 11, 35, 63).
- [34] Giuseppe Portella, Carmela Passaro, and Paolo Chieffi. “Aurora B: A New Prognostic Marker and Therapeutic Target in Cancer.” In: *Current medicinal chemistry* 18 (Feb. 2011), pp. 482–96. DOI: 10.2174/092986711794480203 (cit. on p. 12).
- [35] Ankit Borisa and Hardik G. Bhatt. “A comprehensive review on Aurora kinase: Small molecule inhibitors and clinical trial studies.” In: *European journal of medicinal chemistry* 140 (2017), pp. 1–19 (cit. on p. 12).
- [36] Marco D’Abramo et al. “Modeling conformational transitions in kinases by molecular dynamics simulations: Achievements, difficulties, and open challenges.” In: *Frontiers in genetics* 5 (May 2014), p. 128. DOI: 10.3389/fgene.2014.00128 (cit. on pp. 14, 15).
- [37] Martin Karplus. “Molecular Dynamics Simulations of Biomolecules.” In: *Accounts of chemical research* 35 (July 2002), pp. 321–3. DOI: 10.1021/ar020082r (cit. on p. 14).
- [38] M. Karplus and J. Kuriyan. “Molecular dynamics and protein function.” In: *Proceedings of the National Academy of Sciences* 102.19 (2005), pp. 6679–6685. ISSN: 0027-8424. DOI: 10.1073/pnas.0408930102. eprint: <https://www.pnas.org/content/102/19/6679.full.pdf>. URL: <https://www.pnas.org/content/102/19/6679> (cit. on p. 15).

- [39] Daan Frenkel and Berend Smit. *Understanding molecular simulation : from algorithms to applications. 2nd ed.* Vol. 50. Jan. 1996, pp. 63–71. DOI: 10.1063/1.881812 (cit. on pp. 15, 19).
- [40] Furio Ercolessi. “A molecular dynamics primer.” In: *Spring College in Computational Physics, ICTP, Trieste* (Jan. 1997) (cit. on pp. 16, 20, 21).
- [41] Mark Abraham et al. *GROMACS User Manual version 2018*. www.gromacs.org. GROMACS. 2018, pp. 26–27 (cit. on pp. 16, 19, 22, 69).
- [42] Peter J. Steinbach. *The Empirical Potential Energy Function*. URL: [https://cmm.cit.nih.gov/intro\\_simulation/node15.html](https://cmm.cit.nih.gov/intro_simulation/node15.html) (visited on 05/23/2020) (cit. on p. 17).
- [43] Wendy D Cornell et al. “A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules.” eng. In: *Journal of the American Chemical Society* 117.19 (1995), pp. 5179–5197. ISSN: 0002-7863 (cit. on p. 19).
- [44] WF Van Gunsteren et al. “vdf Hochschulverlag AG an der ETH Zürich and BIOMOS bv: Zürich, Groningen.” In: (1996), pp. 1–1042 (cit. on p. 19).
- [45] W Jorgensen, David Maxwell, and Julian Tirado-Rives. “Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids.” In: *J. Am. Chem. Soc.* 118 (Jan. 1996), pp. 11225–11236 (cit. on p. 19).
- [46] Kenno Vanommeslaeghe et al. “CHARMM General Force Field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force field.” In: *Journal of computational chemistry* 31 (Nov. 2009), pp. 671–90. DOI: 10.1002/jcc.21367 (cit. on p. 19).
- [47] Justin Lemkul. “From Proteins to Perturbed Hamiltonians: A Suite of Tutorials for the GROMACS-2018 Molecular Simulation Package [Article v1.0].” In: *Living Journal of Computational Molecular Science* 1 (Oct. 2018). DOI: 10.33011/livecoms.1.1.5068 (cit. on pp. 20, 69).
- [48] Anders Malthe-Sørenssen and Dag Kristian Dysthe. *Statistical and Thermal Physics Using Python*. Department of Physics, University of Oslo, 2017. URL: [https://www.uio.no/studier/emner/matnat/fys/FYS2160/h18/pensumliste%5C%5C%20/stat%5C\\_thermal%5C\\_phys%5C\\_python.pdf](https://www.uio.no/studier/emner/matnat/fys/FYS2160/h18/pensumliste%5C%5C%20/stat%5C_thermal%5C_phys%5C_python.pdf) (cit. on pp. 20, 21).
- [49] Norbert Attig et al. *Computational Soft Matter: from Synthetic Polymers to Proteins*. NIC series 2. NIC, 2004. ISBN: 9783000126413. URL: <https://books.google.no/books?id=IG8rAwAACA AJ> (cit. on p. 20).
- [50] Ramu Anandakrishnan et al. “Speed of Conformational Change: Comparing Explicit and Implicit Solvent Molecular Dynamics Simulations.” In: *Biophysical journal* 108 (Mar. 2015), pp. 1153–64. DOI: 10.1016/j.bpj.2014.12.047 (cit. on p. 21).
- [51] Atanu Das and Chaitali Mukhopadhyay. “Application of principal component analysis in protein unfolding: An all-atom molecular dynamics simulation study.” In: *The Journal of chemical physics* 127 (Nov. 2007), p. 165103. DOI: 10.1063/1.2796165 (cit. on p. 22).
- [52] Charles David and Donald Jacobs. “Principal Component Analysis: A Method for Determining the Essential Dynamics of Proteins.” In: *Methods in molecular*

- biology (Clifton, N.J.)* 1084 (Jan. 2014), pp. 193–226. DOI: 10.1007/978-1-62703-658-0\\_11 (cit. on p. 22).
- [53] Matej Janeček et al. “Allosteric modulation of AURKA kinase activity by a small-molecule inhibitor of its protein-protein interaction with TPX2.” In: *Scientific Reports* 6 (June 2016), p. 28528. DOI: 10.1038/srep28528 (cit. on pp. 23, 36, 39).
- [54] Jing Huang et al. “CHARMM36m: An Improved Force Field for Folded and Intrinsically Disordered Proteins.” In: *Biophysical Journal* 112 (Feb. 2017), 175a–176a. DOI: 10.1016/j.bpj.2016.11.971 (cit. on p. 23).
- [55] Schrödinger, LLC. “The PyMOL Molecular Graphics System, Version 1.8.” Nov. 2015 (cit. on pp. 23, 25).
- [56] Herman Berendsen et al. “Molecular-Dynamics with Coupling to An External Bath.” In: *The Journal of Chemical Physics* 81 (Oct. 1984), p. 3684. DOI: 10.1063/1.448118 (cit. on p. 24).
- [57] Ulrich Essmann et al. “A Smooth Particle Mesh Ewald Method.” In: *J. Chem. Phys.* 103 (Nov. 1995), p. 8577. DOI: 10.1063/1.470117 (cit. on p. 24).
- [58] Berk Hess et al. “LINCS: A Linear Constraint Solver for molecular simulations.” In: *Journal of Computational Chemistry* 18 (Apr. 1998). DOI: 10.1002/(SICI)1096-987X(199709)18:123.0.CO;2-H (cit. on p. 24).
- [59] S. Nose. “A molecular dynamics method for simulations in the canonical ensemble.” In: *Mol. Phys.* 100 (Jan. 2002), pp. 191–198. DOI: 10.1080/00268970110089108 (cit. on p. 25).
- [60] William Hoover. “Canonical Dynamics: Equilibrium Phase-Space Distributions.” In: *Phys. Rev. A: At., Mol., Opt. Phys.* 31 (Apr. 1985), p. 1695. DOI: 10.1103/PhysRevA.31.1695 (cit. on p. 25).
- [61] Glenn Martyna, Michael Klein, and Mark Tuckerman. “Nos??-Hoover chains: The canonical ensemble via continuous dynamics.” In: *The Journal of Chemical Physics* 97 (Aug. 1992), pp. 2635–2643. DOI: 10.1063/1.463940 (cit. on p. 25).
- [62] M Parrinello and A Rahman. “Polymorphic transitions in single crystals: A new molecular dynamics method.” In: *Journal of applied physics*. 52.12 (1981), pp. 7182–7190. ISSN: 0021-8979 (cit. on p. 25).
- [63] David van der Spoel et al. “GROMACS: fast, flexible, and free.” In: *Journal of computational chemistry* 26 (Dec. 2005), pp. 1701–18. DOI: 10.1002/jcc.20291 (cit. on p. 25).
- [64] Herman Berendsen, David van der Spoel, and Rudi Drunen. “GROMACS: A message-passing parallel molecular dynamics implementation.” In: *Computer Physics Communications* 91 (Sept. 1995). DOI: 10.1016/0010-4655(95)00042-E (cit. on p. 25).
- [65] Mark Abraham et al. “GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers.” In: *SoftwareX* 1 (July 2015). DOI: 10.1016/j.softx.2015.06.001 (cit. on p. 25).
- [66] Berk Hess et al. “GROMACS 4: algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation.” In: *J Chem Theory Comput* 4 (Mar. 2008), pp. 435–447. DOI: 10.1021/ct700301q (cit. on p. 25).

- [67] Sander Pronk et al. “GROMACS 4.5: A high-throughput and highly parallel open source molecular simulation toolkit.” In: *Bioinformatics (Oxford, England)* 29 (Feb. 2013). DOI: 10.1093/bioinformatics/btt055 (cit. on p. 25).
- [68] Uninett Sigma2. *Fram*. URL: [https://documentation.sigma2.no/hpc%5C\\_machines/fram.html](https://documentation.sigma2.no/hpc%5C_machines/fram.html) (visited on 02/24/2020) (cit. on p. 25).
- [69] William Humphrey, Andrew Dalke, and Klaus Schulten. “VMD – Visual Molecular Dynamics.” In: *Journal of Molecular Graphics* 14 (1996), pp. 33–38 (cit. on p. 25).
- [70] Xavier Robert and Patrice Gouet. “Deciphering key features in protein structures with the new ENDscript server.” In: *Nucleic Acids Research* 42.W1 (Apr. 2014), W320–W324. ISSN: 0305-1048. DOI: 10.1093/nar/gku316. eprint: <https://academic.oup.com/nar/article-pdf/42/W1/W320/17422987/gku316.pdf>. URL: <https://doi.org/10.1093/nar/gku316> (cit. on p. 25).
- [71] Ac Wallace, Ra Laskowski, and Jm Thornton. “LIGPLOT - A PROGRAM TO GENERATE SCHEMATIC DIAGRAMS OF PROTEIN LIGAND INTERACTIONS.” English. In: *Protein Engineering* 8.2 (1995), pp. 127–134. ISSN: 0269-2139 (cit. on p. 25).
- [72] Atul Srivastava et al. “Synchronous Opening and Closing Motions Are Essential for cAMP-Dependent Protein Kinase A Signaling.” In: *Structure (London, England : 1993)* 22 (Dec. 2014), pp. 1735–43. DOI: 10.1016/j.str.2014.09.010 (cit. on pp. 33, 64).
- [73] Diwakar Shukla et al. “Activation pathway of Src kinase reveals intermediate states as targets for drug design.” In: *Nature communications* 5 (Jan. 2014), p. 3397. DOI: 10.1038/ncomms4397 (cit. on p. 40).
- [74] Nicholas Levinson. “The multifaceted allosteric regulation of Aurora kinase A.” In: *Biochemical Journal* 475 (June 2018), pp. 2025–2042. DOI: 10.1042/BCJ20170771 (cit. on p. 40).
- [75] Christopher McClendon et al. “Dynamic architecture of a protein kinase.” In: *Proceedings of the National Academy of Sciences of the United States of America* 111 (Oct. 2014). DOI: 10.1073/pnas.1418402111 (cit. on p. 54).