

Introduction

Centrosomes are membrane-less and self-replicating organelles encompassing a pair of orthogonally arranged cylindrical structures called centrioles embedded in an electron-dense proteinaceous matrix referred to as pericentriolar material (PCM) [Mattison and Winey, 2006; von Theodor Boveri, 1900]. Centrioles are composed of microtubule triplets exhibiting a nine-fold radial symmetric arrangement twisted in an anti-clockwise manner. Among the centriole pairs, one centriole is more mature, having the distal and sub-distal appendages, referred to as the mother centriole, and the other is called the daughter centriole. In dividing cells, centrosomes act as major microtubule (MT) nucleating centers of the cell that form the bipolar spindles assembly during the M-phase of the cell cycle and equally divide the genetic material between the two newly formed daughter cells [Bornens, 2002; Jaiswal *et al.*, 2021]. To create two poles for a bipolar spindle generation at the M-phase, the centrosome duplicates in a regulated manner at the S-phase of the cell cycle [Alberts *et al.*, 2002; Cooper, 2000; Ong and Torres, 2019]. During the duplication process, at the proximal end of the parent centrioles, centrosome proteins CEP192 and CEP152 act as a molecular scaffold to recruit Polo-Like Kinase 4 (PLK4) in a ring form at the late G1-phase of the cell cycle [Kim *et al.*, 2013; Sonnen *et al.*, 2013]. At the beginning of the S-phase, PLK4 undergoes ring-to-dot transitioning and recruits downstream proteins to initiate cartwheel assembly and daughter centriole formation [Arquint and Nigg, 2016; Moyer and Holland, 2019].

PLK4, the master regulator of the centrosome duplication cascade, is a unique member of the Polo-Like Kinases (PLKs) family [Habedanck *et al.*, 2005; Nigg, 2007]. PLKs are serine/threonine kinases having a kinase domain (KD) at their amino-terminal and a distinctive polo-box domain at their carboxyl-terminal [Barr *et al.*, 2004]. The cyclin-dependent kinases (CDKs) and PLKs-mediated phosphorylations are the primary regulators of cell division [Ong and Torres, 2019]. The most widely studied PLK family member, PLK1, has two characteristic polo-boxes (PBs) separated by a linker region in the polo-box domain (PBD). However, PLK4

has a distinctive PBD with three PBs. The two PBs (PB1 and PB2) in tandem are referred to as the cryptic-polo-box (CPB) region, followed by a linker region and a third PB (PB3) region [Sillibourne and Bornens, 2010; Swallow *et al.*, 2005]. PLK4 homodimerizes using its CPB region [Park *et al.*, 2014] and promotes the trans-autophosphorylation of the serine and threonine residues in the PEST motif. Phosphorylated residues provide the SCF/ β -TrCP E3 ubiquitin ligase binding site and target PLK4 for degradation [Cunha-Ferreira *et al.*, 2013; Holland *et al.*, 2012; Lopes *et al.*, 2015]. Regulated levels of PLK4 are essential, as their overexpression results in centrosome amplification, causing the formation of rosette-like structures around the parent centriole [Kleylein-Sohn *et al.*, 2007]. Centrosome amplification is a hallmark of cancer, and levels of PLK4 are significantly deregulated in several types of cancers [Liu *et al.*, 2012; Marina and Saavedra, 2014; Shinmura *et al.*, 2014; Wang *et al.*, 2019]. It is a valuable biomarker and a potential therapeutic target for cancer treatment. Several small molecules targeting PLK4 in its kinase domain are under preclinical and clinical phase-1/2 trials for anticancer drug development [Garvey *et al.*, 2021].

In resting cells (G0 phase), the parent centriole docks up in the plasma membrane to form the basal body that emerges as hair-like protruding structures called cilia and flagella [Bornens, 2012; Jaiswal and Singh, 2021]. Abnormalities in centrosome organization and functioning also induce ciliary defects, as observed in metabolic disorders such as Type 2 Diabetes Mellitus (T2DM). It is associated with insulin resistance, resulting in hyperinsulinemia, hyperglycemia, and hyperlipidemia. Recently, centrosome amplification is also observed in the peripheral blood mononuclear cells (PBMC) of type 2 diabetes mellitus (T2DM) patients [Chae *et al.*, 2005]. Epidemiological studies indicate increased risk for colon, liver, pancreatic [Pan *et al.*, 2018], breast [Boyle *et al.*, 2012; Hardefeldt *et al.*, 2012] and endometrial [Saed *et al.*, 2019] cancer in T2DM patients. Although chronic T2DM is linked with co-morbidities like ketoacidosis, retinopathy, nephropathy, neuropathy, and coronary heart diseases [Deshpande *et al.*, 2008; Farmaki *et al.*, 2020], there is no clear molecular understanding of cancer risks in T2DM.

1.1 Purpose of the Study

We investigated the impact of PLK4 homodimerization on centrosome functioning and organization, which was found to be disrupted in cancer. We further elucidated the first full-length structure of human PLK4 using *ab initio* and threading approaches and employed it to repurpose FDA-approved drugs. This work identifies novel compounds targeting the unique homodimerization region of PLK4, thus potentially useful for cancer chemotherapeutics. Finally, we utilized gene expression profiling studies to understand the molecular link between cancer and metabolic disorder, Type 2 Diabetes Mellitus (T2DM), thereby identifying potential biomarkers that could predict the onset of cancer in T2DM patients.

Overexpression of PLK4 causes centrosome amplification [Kleylein-Sohn *et al.*, 2007], and its downregulation reduces centrosome numbers [Bettencourt-Dias *et al.*, 2005; Holland *et al.*, 2012], which has implications on spindle organization, that could translate to genomic instability found in cancers. Among the polo-like kinase family members, PLK4 has a unique C-terminal cryptic-polo box (CPB) region, essential for its functioning and protein stability. The CPB is utilized for binding with the PLK4 centrosome recruiters CEP192 and CEP152, and it is also required for homodimerization, which regulates its protein stability [Cunha-Ferreira *et al.*, 2013; Hatch *et al.*, 2010; Kim *et al.*, 2013]. However, the precise role of PLK4 homodimerization on the CEP192-CEP152 network remains unexplored. We identified a cancerous mutant that tweezes apart the role of PLK4 homodimerization in regulating temporal recruitment of CEP152 at the S phase centrosomes, thereby affecting spindle organization and cell viability.

Several small molecule inhibitors targeting the kinase domain of PLK4 are under preclinical and clinical phase-1/2 trials for anticancer drug development [Zhao and Wang, 2019]. However, the PLK4 kinase domain has structural similarities with the catalytic domains of other serine-threonine-tyrosine (STY), and aurora kinases [Mason *et al.*, 2014; Oegema *et al.*, 2018; Sredni *et al.*, 2017]. Therefore, these inhibitors show non-specificity and cross-reactivity with other kinases. Identifying novel chemical scaffolds against the unique CPB region of PLK4 involved in protein homodimerization could bring better specificity. However, the structure of full-length human PLK4 is not available. Therefore, we also predicted the

structure of full-length human PLK4 and showed the usefulness of the predicted structure by identifying a novel chemical scaffold targeting the CPB region.

Centrosomal aberrations cause ciliary defects, which have been reported in metabolic disorders, including T2DM [Gerdes *et al.*, 2014]. Accordingly, in T2DM patients, fasting blood glucose and HbA1c (glycated hemoglobin) levels correlate with an increase in the number of centrosomes per cell [Wang *et al.*, 2018b]. Mendelian randomization studies specify a positive association between T2DM and the risk of pancreatic, breast, lung, liver, and kidney cancer [Pearson-Stuttard *et al.*, 2021; Shen *et al.*, 2022]. Epidemiological studies also indicate an increased risk for liver (LC), pancreatic (PC), and breast cancer (BC) in T2DM patients in different human populations. T2DM pathophysiology factors, hyperglycemia, hyperinsulinemia, chronic inflammation, and insulin resistance can trigger mitogenic pathways and cause cancer [Shlomai *et al.*, 2016]. Despite these linkages between T2DM and cancer risk, there is no systematic understanding of cancer risks in T2DM at the molecular and genetic levels. Thus, we investigated differentially expressed genes that revealed lead biomarkers and shared functional pathways in T2DM *vs.* common cancer types (PC, LC, and BC) and have potential prognostic and therapeutic applications.

1.2 Brief Results, Scope, and Future Prospects

In the first part of the thesis work, we provide mechanistic insights into the involvement of PLK4 homodimerization in centrosome organization. We utilized publicly available pan-cancer datasets from TCGA, COSMIC, ICGC, and Broad Institute to identify non-synonymous mutations in the PLK4 cryptic polo-box region. The PB2 region was found to be highly mutated, and we narrowed it down to a protein truncation mutation occurring at the glutamic acid (E) of PLK4 at 774 residue, which disrupts the α -helix involved in PLK4 homodimerization. The PLK4 homodimerization mutant (E774*) expression in cells was found more than wild-type PLK4 due to its better protein stability, as indicated by its ubiquitination status. The homodimerization mutant encompasses residues (K685 and K711) involved in the direct binding with its centrosome recruiter CEP152. However, it shows diminished interaction with CEP152 in the GST pull-down and immunoprecipitation assays. The homodimerization mutant also reduced the centrosome recruitment of CEP152 at the late S-

phase, which eventually diminished the CEP152-dependent pericentrin recruitment, resulting in unfocussed spindles at the M-phase that affect cell viability. These findings establish a cross-talk between PLK4 and CEP152 for centrosome maintenance at the late S-phase. In the future, biophysical and structural studies that allow better distinction of monomeric and dimeric states of PLK4 could shed more light on the temporal requirement of PLK4-CEP152-CEP192 scaffolds for centrosome organization and functioning.

In the next part of the thesis, we utilized I-TASSER and Galaxy Refine web servers to predict the full-length structure of human PLK4. The predicted structure is γ (gamma) shaped, with the kinase domain (KD) forming one arm of the γ , the polo-box domain (PBD) forming another arm, and the linker-1 region forming its tail. The predicted structure has 96.42% residues falling in Ramachandran's favored region with an Errat score of 77.3543. It also shows average to good alignment with the already available XRD structures of the PLK4 KD (PDB: 4YUR), CPB (PDB: 4N7V), and PB3 (PDB: 2N19) regions from the Protein DataBank (PDB). We further utilized the predicted monomer structure to anticipate its dimeric form. The predicted dimeric structure of PLK4 showed an X-shaped organization in the CPB region, similar to the XRD structure for CPB dimer (PDB ID: 4N9J and 4N7V). We also validated the predicted structure by docking it with the widely used PLK4-specific inhibitor, centrinone. Once validated, the predicted structure was utilized to virtually screen the ChEMBL drug library to identify novel chemical scaffolds targeting PLK4. Seven ligands falling under the 10% of the distributed range with binding energy ≤ -10 kcal/mol were shortlisted. They were found to show good pharmacophore and pharmacodynamic properties using ADMET analysis. Among them, we explored the ALK tyrosine kinase inhibitor, Alectinib, which was predicted to bind in the PLK4 CPB region. FTIR analysis showed Alectinib interaction with the recombinantly purified PLK4 polo-box domain encompassing the CPB. Like the other PLK4-specific inhibitors, Alectinib also showed a bimodal mechanism of action, *i.e.*, at low concentrations, it enhanced PLK4 levels at centrosomes, which resulted in centrosome amplification. At high concentrations, it inhibited PLK4, which reduced centrosome number. Further structural and biophysical studies are required to validate the predicted structure.

The structure-activity relationship in the future could also identify possible chemical modifications of the scaffold, improving its potency against PLK4.

The last part of the thesis presents the identification of genetic markers that can predispose the risk of cancer onset in T2DM patients. This collaborative work enabled us to perform gene expression profiling and protein-protein network analysis to reveal prognostic hub biomarkers linking risk of common cancer types, namely, pancreatic (PC), liver (LC), and breast cancer (BC) in T2DM patients. We utilized three gene expression omnibus (GEO) datasets of T2DM, PC, LC, and BC patients. A three-tier filtering criterion, *i.e.*, statistical significance, biological significance, and relative distance, was employed on the common differentially expressed genes (DEGs) from peripheral blood mononuclear cells of the diseased conditions. The protein-protein interaction network of the filtered DEGs yielded 7, 6, and 5 hub genes in T2DM *vs.* PC, T2DM *vs.* LC, and T2DM *vs.* BC conditions, respectively. The functional and pathway enrichment analysis revealed involvement of cell cycle events, immune system processes, cell morphogenesis, gene expression, and metabolism pathways. We found 5 out of the 17 common hub genes, that affect centrosome functioning and organization. Since majority of centrosome-associated genes have low expression in cells, future analysis on specific cancer tissue samples with a gene-specific approach could yield a better correlation of centrosome involvement in cancer and T2DM.