CHAPTER 7 (a)

Unveiling the Transient Protein-Protein Interactions that Regulate the Activity of Human Lemur Tyrosine Kinase-3 (LMTK3) Domain by Cyclin Dependent Kinase 5 (CDK5) in Breast Cancer: An *in silico* Study

Unveiling the Transient Protein-Protein Interactions that Regulate the Activity of Human Lemur Tyrosine Kinase-3 (LMTK3) Domain by Cyclin Dependent Kinase 5 (CDK5) in Breast Cancer: An *in silico* Study

7a.1. Abstract

In many human diseases protein kinases are known to play a central role. Protein kinases phosphorylate its substrates and they themselves regulated through phosphorylation of their activation loop and become catalytically active. Recent report says phosphorylation of LMTK3 by CDK5 results in breast cancer tumor progression. Thereby information about interface residues and probable phosphorylation site on LMTK3 is critical. Our objective is to understand the transient protein – protein interactions between CDK5 and LMTK3 using computational techniques. LMTK3 structure was superimposed with known kinases to determine the probable activation segment and phosphorylation sites in LMTK3. PatchDock was used to obtain CDK5-LMTK3 complex structure. PDBsum server was used to identify the interface residues between CDK5 and LMTK3. The stability of CDK5-LMTK3 complex was studied using Molecular dynamics (MD) simulation. From PatchDock, interface area between CDK5-LMTK3 complex was found to be 2081 $Å^2$ with atomic contact energy of -228.80 kcal/mol. PDBsum result reveals that, CDK5 interacts and displayed non-bonding interactions with the probable phosphorylation sites of LMTK3. Total number of interface residues across CDK5-LMTK3 was found to be around 50 and the interface area was found to be 1274 Å2 (in CDK5) and 1224 Å2 (in LMTK3). From MD simulation, CDK5-LMTK3 complex was found to be stable. This study enhances the understanding of interactions between CDK5 and LMTK3 that may be helpful in understanding the LMTK3 phosphorylation by CDK5 which is considered to be a new cellular pathway in breast cancer tumor progression

7a.2. Introduction

In the recent past LMTK3 was found to be involved in various types of cancer including breast [5], lungs [367] and colorectal cancer [87]. Recently, in vitro studies have

identified the ability of CDK5 to phosphorylate LMTK3 and results in breast cancer progression [10]. Phosphorylation is a common post-translational modification of proteins in eukaryotic cells [113] that regulate several important cellular processes, such as cell growth and differentiation [114, 115]. Eukaryotic protein kinases (EPKs) are the family of enzymes that catalyzes the regulatory phosphorylation reaction, and they themselves regulated and activated through phosphorylation. The regulatory phosphorylation event occurs at kinases activation loop, which is a part of the activation segment [11, 12]. The C-lobe of protein kinase contains an activation loop which is 20-35 residues stretch located between a conserved DFG and APE motif which is conformationaly very flexible, its conformation affects on both substrate binding and catalytic efficiency [71,72]. The activation loop phosphorylation is a major mechanism that induces the dynamic changes in activation process and undergoes structural and conformational changes that leads to stabilization of the active conformation and catalyses the phosphoryl transfer reaction of γ-phosphate of an ATP molecule to its substrate. Hence activation-loop phosphorylation is critical because it is required for the interconversion from an inactive to an active conformation of kinase [368, 13-15]. In addition, the protein kinase domain is a structurally conserved and contains the catalytic function of all protein kinases [369, 370, 291]. Cyclin dependent kinase-5 (CDK5) is a cytoplasmic proline-directed serine/threonine kinase which is commonly over expressed in many tumors [112]. However intermolecular interactions between CDK5 and LMTK3 during phosphorylation are yet to be understood. Due to the unavailability of 3D structure of LMTK3 domain we modeled the 3D structure of LMTK3 structure from I-TASSER [99]. In the present study, we predicted the activation segment in LMTK3 because protein kinases activated through phosphorylation of their activation loop in the activation segment. We compared the LMTK3 structure with Protein kinase A (PKA, PDB ID: 1ATA) and mitogen activated protein kinase-1 (MAPK1) or ERK2 (PDB ID: 2ERK) and based on two conserved motifs (DFG and APE) [11,12] we have determined the probable activation segment in LMTK3. It is known that phosphorylation of these kinases in activation loop at threonine regulates the kinase activity [25-28]. Based on our comparative study we have determined the probable phosphorylation sites to be Thr167 and Thr189 within the activation segment of LMTK3. In order to determine the

probable interacting residues and the interface area during LMTK3 phosphorylation by CDK5, we performed molecular docking of CDK5 (PDB ID: 3O0g) [371] with LMTK3 and studied its interaction profile. Using PatchDock [229] we carried out molecular docking and obtained a CDK5-LMTK3 complex structure. The protein-protein interaction was then studied using PDBsum server [252]. With the CDK5-LMTK3 interaction study we noticed that near the C terminal residues of CDK5 forming non-bonded contacts with LMTK3 at Thr167 and Thr189 which we have considered to be the probable phosphorylation sites. Conformational dynamics of CDK5-LMTK3 complex was studied from molecular dynamics simulation. Our computational study provides significant insights into the probable interacting residues, binding regions, phosphorylation sites, interface area, bonded and non-bonded interactions across CDK5-LMTK3 complex. Thus, with the knowledge of interface residues and phosphorylation sites, we can have control over protein-protein interactions which may be helpful for the discovery of new drug targets.

7a.3. Materials and Methods

7a.3.1. Prediction of activation segment and probable phosphorylation sites in LMTK3 domain

The 3D structure of LMTK3 domain was modeled from I-TASSER. In order to determine probable activation segment and phosphorylation site in LMTK3, the equilibrated structure of LMTK3 was superimposed with X-ray crystallographic structure of PKA (PDB ID: 1ATA) and ERK2 (PDB ID: 2ERK).

7a.3.2. Protein – protein docking

7a.3.2.1. Rigid docking

PatchDock [229] a molecular docking tool was used, to generate the rigid docking complex of CDK5 (PDB ID: 3O0g) with LMTK3. In PatchDock, the docking is carried out by computing the 3D transformations of one protein with respect to the other protein by maximizing surface shape complementarity while reducing the number of steric clashes. The docked complexes rank based on geometric shape complementary score.

Geometric scoring means, good molecular shape complementarity between the docked structures due to optimal fit with wide interface area and lesser steric clashes [229].

7a.3.2.2. Refinement of complex structure

The rigid docking model structures of CDK5-LMTK3 obtained from PatchDock then, the structures were refined by using FireDock algorithm [330]. The FireDock algorithm refines the complex structures by optimizing the side chain conformation and reduces steric hindrances. The resultant refined model structures were then given score based on an

energy function. The score includes atomic contact energy [242], softened van der Waals interaction, electrostatics, hydrogen bonding and additional estimations of the binding

free energy.

7a.3.3. Prediction of interface residues between CDK5 and LMTK3 domain

Information about the interface area and residues across CDK5-LMTK3 complex has been determined using PDBsum online server. From PatchDock we have selected the best docking solution based on highest geometric surface area with lowest atomic contact energy (ACE) and then submitted to PDBsum [252]. PDBsum server summarizes the information about bonded, non-bonded contacts and the interacting residues involved between CDK5 and LMTK3.

7a.3.4. Molecular dynamics simulation of CDK5-LMTK3 Complex

From the MD simulations study of protein-protein complex, protein-protein interactions and flexibility can be ensured and characterized [347, 349]. Due to the dynamic nature of protein-protein interfaces, MD simulation allows the transient pockets and buried binding hot spots to appear on the protein surfaces and binding of small molecules to these transient areas can be targeted [345, 348]. Since the pioneer paper entitled 'The Biological Functions of Low-Frequency Phonons' [372] was published in 1977, a series of investigations into bio macromolecules from dynamic point of view have been stimulated. These studies have suggested that low frequency (or terahertz frequency) collective motions do exist in proteins and DNA (see, *e.g.*, [373-375] as well as a long

list of papers cited in two comprehensive review papers [376,377]). Furthermore, many important biological functions in proteins and DNA and their dynamic mechanisms, such as switch between active and inactive states [378], cooperative effects [379], allosteric transition [373], intercalation of drugs into DNA [374], and assembly of microtubules [380] , can be revealed by studying the low-frequency internal motions as summarized in a recent review [381]. Some scientists even applied this kind of low-frequency internal motion for medical treatments [382-384]. Actually, investigation into the internal motion in biomacromolecules and its biological functions is deemed as a "genuinely new frontier in biological physics". In view of this, to really understand the action mechanisms of biomacromolecules, we should consider not only the static structural information but also the dynamical information acquired by studying their internal motions. To realize this, the MD simulation is one of the feasible tools. In the order to study the conformational dynamics of CDK5-LMTK3 complex we have performed molecular dynamics simulation. The refined complex structure of CDK5-LMTK3 obtained from Firedock with lowest energy has been used as input structure for the Molecular Dynamics (MD) simulation. The complex was then subjected to MD simulations using Particle Mesh Ewald Molecular Dynamics (PMEMD) [305] module of AMBER12 [122] software and AMBER ff99SB force field [306] protein parameters were used on the CDK5-LMTK3 complex. The topology and coordinate files for CDK5-LMTK3 complex were prepared using implicit solvation in xleap module of AMBER package. The resultant structure was then used for energy minimization by using 500 steps of steepest descent and another 500 steps of conjugate gradient. We did not fix any restraint during energy minimization, to hold the protein system. The minimized structure was then subjected to 100 ps of MD using 2 fs time step for integration. During the MD, the system was gradually heated from 0 to 300 K and ensured slow relaxation of the built initial structure. In addition, shake [165] constraints were imposed with a geometric tolerance of 5 x 10-4 Å on all the covalent bonds involving hydrogen atoms. Subsequently, MD was performed under constant pressure temperature conditions (NPT) with temperature regulation achieved using Berendsen weak coupling method [166] (0.5 ps time constant for heat bath coupling and 0.2 ps pressure relaxation time). This was followed by another 20 ps of equilibration step.

Finally, for the analysis of structures and properties we carried out MD for 14 ns at NPT conditions using a heat bath coupling time constant of 1 ps.

7a.4. Results and Discussions

7a.4.1. Activation segment and phosphorylation sites determined in LMTK3

By comparing the LMTK3 structure with MAPK1 and PKA protein kinase, we determined the probable activation segment and phosphorylation site in LMTK3. As discussed the above activation segment of protein kinases contains two conserved motifs (DFG and APE motif) [11,12]. From **Figure 7a.1** and **7a.2** we noticed that LMTK3 has an APE motif, as that of ERK2 and PKA. Based on this conserved APE motif we predicted the activation segment ranges from Tyr164 to Ser196 in LMTK3 consists of 32 residues stretch. This activation segment may be important for catalytic activation of LMTK3.With the visual inspection we assumed, Thr167 and Thr189 to be the probable phosphorylation sites within activation loop where CDK5 may phosphorylate. From NetPhos3.1 server we cross checked our predicted phosphorylation site, and we observed that Thr167 and Thr189 to be the probable phosphorylation site in LMTK3.

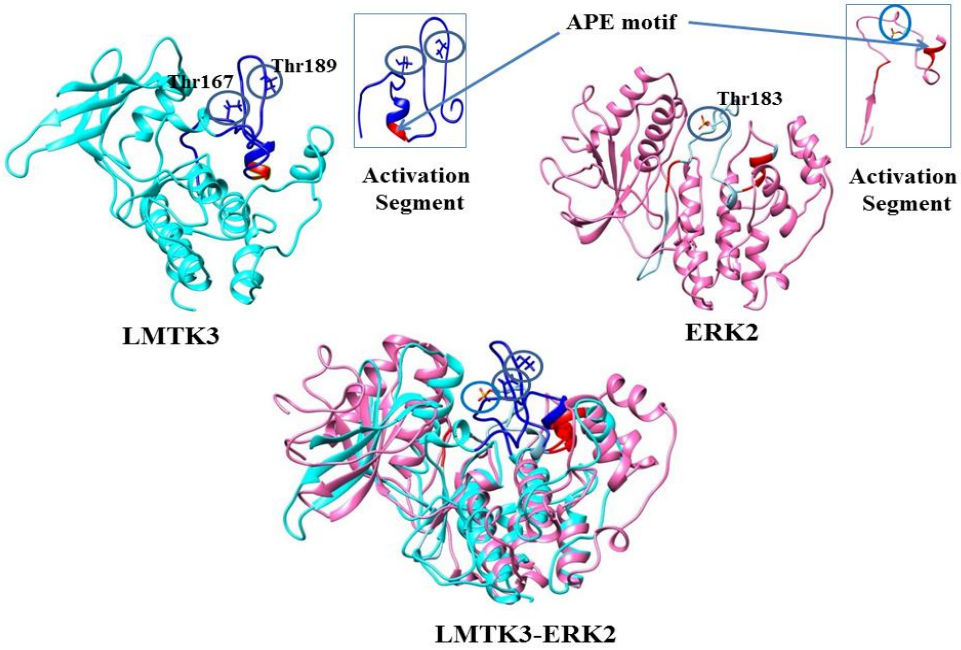

Figure 7a.1. *Superimposed structure of LMTK3 domain with ERK2*

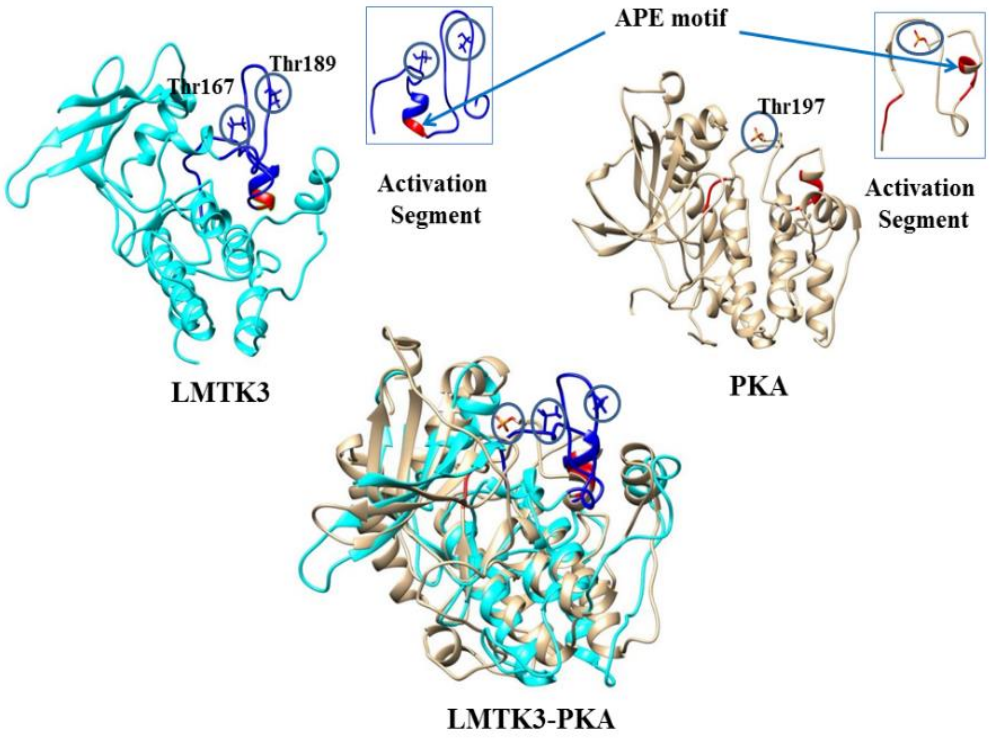

Figure 7a.2. *Superimposed structure of LMTK3 domain with PKA.*

7a.4.2. Protein-protein interaction study

Recently, molecular/cellular and biochemical experiments have confirmed the ability of CDK5 to phosphorylate LMTK3 in vitro and leads to breast cancer tumor progression [10]. However molecular interaction during phosphorylation of LMTK3 by CDK5 is not studied yet. In the present study, we docked LMTK3 and CDK5 (PDB ID: 3O0g) using PatchDock (**Figure 7a.3)**, a molecular docking algorithm.

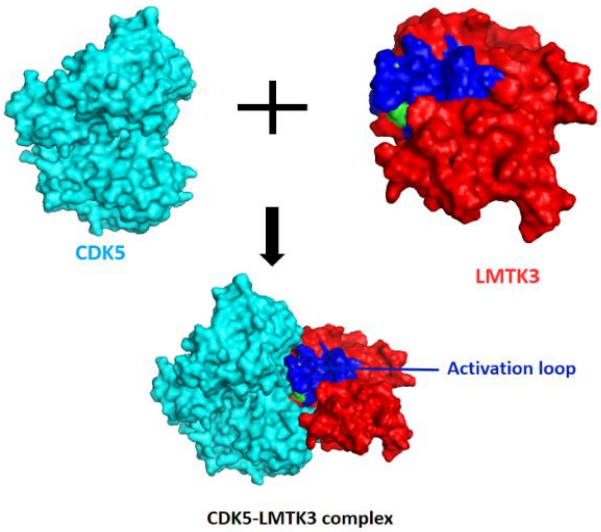

Figure 7a.3 *CDK5-LMTK3 docked complex from PatchDock server*

Several docked structures for CDK5-LMTK3 complex were obtained from PatchDock and they were ranked based on geometric shape complementary score, interface area and atomic contact energy as shown in **Figure 7a.4**.

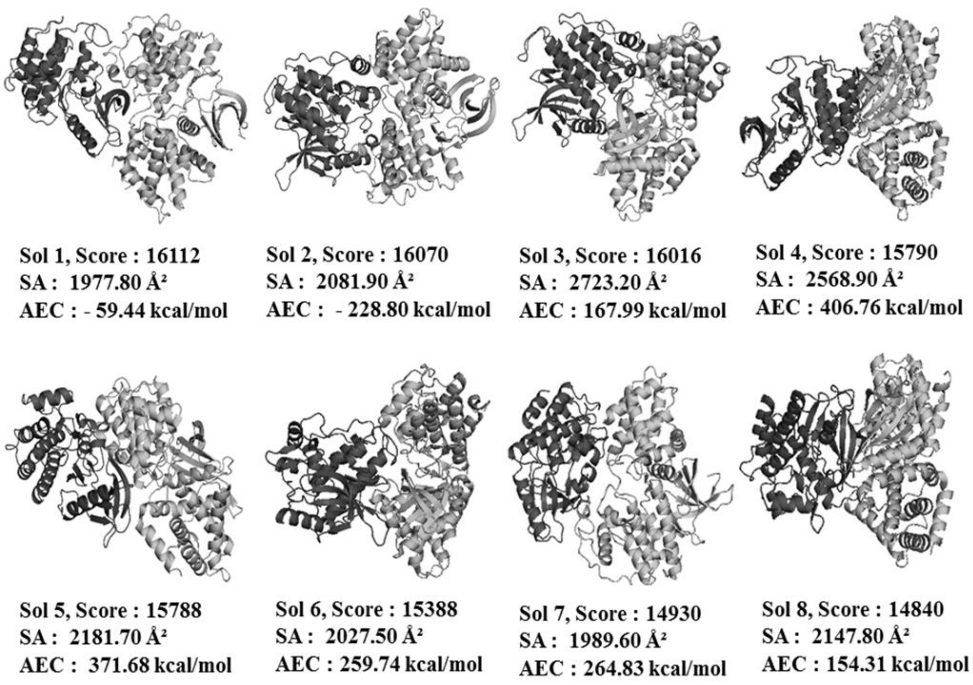

Figure 7a.4. *CDK5-LMTK3 complex structures obtained from PatchDock, ranked based on: high geometric shape complementary score, surface area and atomic contact energy*

The best ranked structure was observed to have a geometric shape complementarity score of 16070, approximate interface area of 2081.90 Å2 and highest atomic contact energy of (ACE) of -228.80 kcal/mol as shown in **Figure 7a.3**

Using PDBsum server we determined the interacting residues and the interface area across CDK5-LMTK3 complex. We submitted the modeled structure of CDK5-LMTK3 complex having high geometric shape complimentary score and lowest atomic contact energy to the PDBsum server. The total number of probable interface residues in CDK5 and LMTK3 complex was found to be 26 and 24, respectively (**Figure 7a.5**) and the interface area for each protein involved in the interaction was observed to be 1272 Å2 (in CDK5) and 1224 Å2 (in LMTK3). The interface statistics is summarized in **Table 7a.1**. The docked complex was stabilized by molecular interactions involving number of hydrogen bonds, non- bonded contacts and one salt bridge.

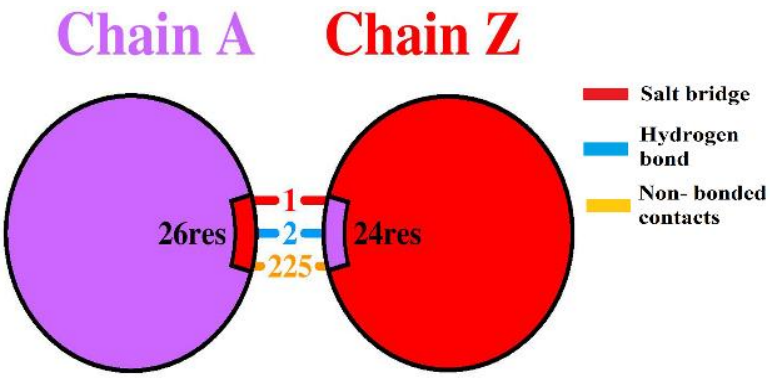

Figure 7a.5. *Summary of the total number of residues involved in different types of interactions in CDK5-LMTK3 complex obtained from PatchDock based on high geometric shape complementary score.*

Table 7a.1. Interface statistics for CDK5-LMTK3 complex

Chain	Number of interface residues	Interface area (Å^2)	Number of salt bridges	Number of disulphide bonds	Number of hydrogen bonds	Number non-bonded contacts
A (CDK5)	26	1272	1	-	2	225
Z (LMTK3)	24	1224				

Figure 7a.6 (A) represents the number of interacting residues across the CDK5-LMTK3 interface. Some interface residues (Tyr179, Leu178 and Tyr158, Arg156) in CDK5 involved in non-bonded interactions with Thr167 and Thr189 (probable phosphorylation sites) within the activation segment of LMTK3. From **Figure 7a.6 (B)** we see CDK5 interacting with the activation segment of LMTK3. The probable phosphorylation sites (Thr167 and Thr189) denoted as sphere in **Figure 7a.7** located across the CDK5-LMTK3 interface. Our results represented the probable interface area and the interface residues involved in CDK5-LMTK3 complex that may play a significant role in breast cancer tumorogenesis.

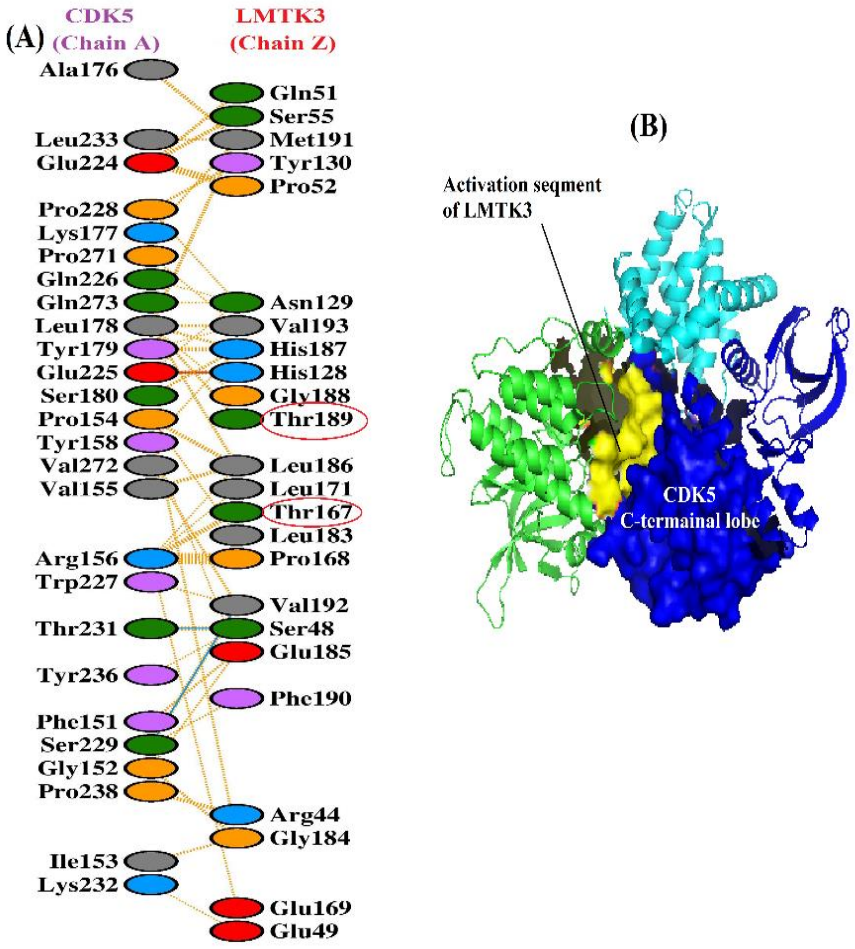

Figure 7a.6. *A) Number of interacting residues that are involved in bonded and non-bonded interactions between CDK5 and LMTK3 interface. (B) Representation of CDK5-LMTK3 complex: C terminal residues of CDK5 interact with activation segment of LMTK3 domain.*

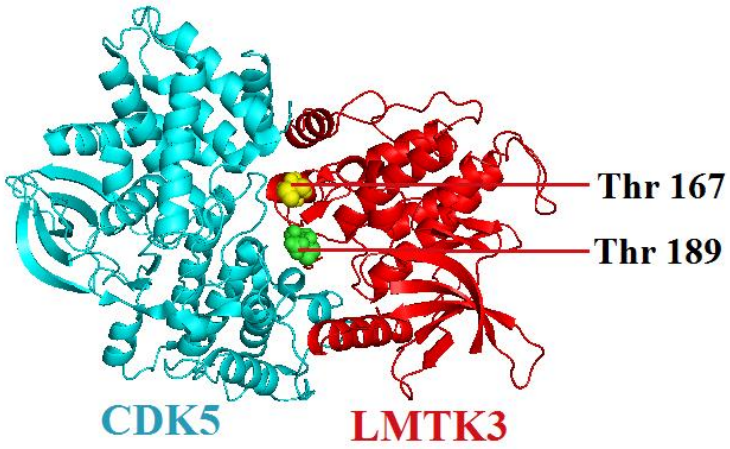

Figure 7a.7. *CDK5-LMTK3 complex representing probable phosphorylation sites in LMTK3 activation loop*

Our results represented the probable interface area and the interface residues involved in CDK5-LMTK3 complex that may play a significant role in breast cancer tumorogenesis. The information of a binding pocket of a receptor for its ligand is very important for drug design, particularly for conducting mutagenesis studies [385]. In the literature, the binding pocket of a protein receptor to a ligand is usually defined by those residues that have at least one heavy atom (i.e., an atom other than hydrogen) within a distance of 5Å from a heavy atom of the ligand. Such a criterion was originally used to define the binding pocket of ATP in the Cdk5- Nck5a* complex [386] that has later proved quite useful in identifying functional domains and stimulating the relevant truncation experiments [387]. The similar approach has also been used to define the binding pockets of many other receptor-ligand interactions important for drug design [388-395]. In our present study we observed that, the residues of N and C terminal lobe of CDK5 are involved in inter-molecular interaction with LMTK3 and between these two lobes a binding pocket is observed for ATP analog as shown in **Figure 7a.8**. This binding pocket is observed to be almost at the center of CDK5-LMTK3 interface. And this binding pocket may be targeted to disrupt the probable transient protein-protein interaction of CDK5 and LMTK3. In the literature, there are studies targeting protein-protein interface that have opened a wide field for drug design [396,397]. Therefore, the information about the probable binding interface residues of N and C terminal lobe that

are involved in interaction with LMTK3 may be utilized for drug target in order to disrupt the binding surfaces.

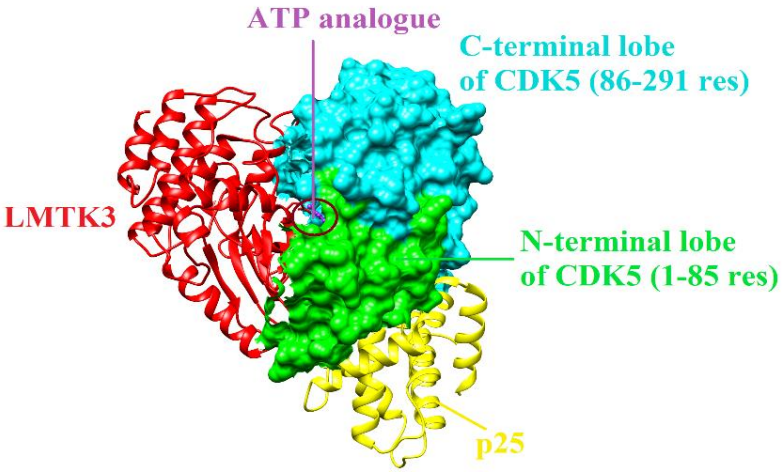

Figure 7a.8. *CDK5-LMTK3 complex showing N and C terminal lobe of CDK5 interacting with LMTK3. The binding pocket for ATP analogue is present between these two lobes which is almost at the center of CDK5 and LMTK3 interface.*

In addition, the information acquired *via* molecular docking not only can provide useful insights for in-depth understanding some subtle action mechanisms such as the marvelous allosteric mechanism revealed recently by the NMR observations on the M2 proton channel of influenza A virus [395,398] but also can stimulate new strategies for drug development as demonstrated by a series of previous studies [388, 389, 392-394,399-404]. As pointed out in [405] and demonstrated in a series of recent publications (see, *e.g.*, [381, 406-425], user friendly and publicly accessible web-servers represent the future direction for developing practically more useful computational tools that will significantly enhance their impacts [426], we shall make efforts in our future work to provide a web-server to display the findings reported in this paper

7a.4.3. MD simulation study on the CDK5-LMTK3 complex

The analysis of stability and convergence of the model structure of CDK5-LMTK3 complex has been studied from the molecular dynamics, as a function of time. From the temperature and energy plots (**Figure 7a.9**), we see that our modeled complex structure has reached equilibration.

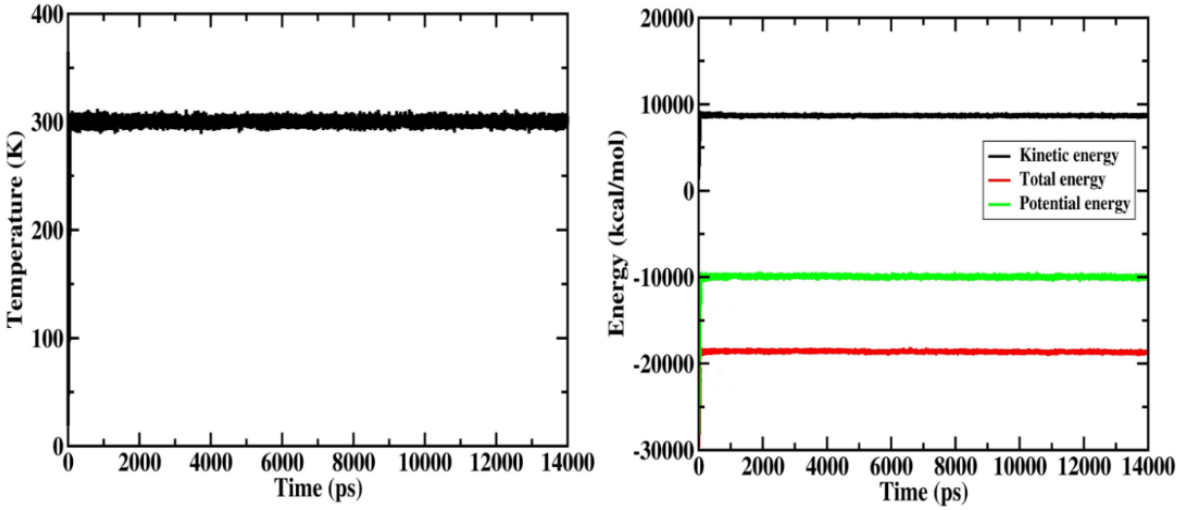

Figure 7a.9. *Temperature and energy plot of CDK5-LMTK3 complex as a function of simulation time*

For RMSD analysis the equilibrated complex structure was used as a reference structure. The RMSD profile for the CDK5- LMTK3 complex is shown in **Figure 7a.10**. In the complex structure, the RMSD value settles well around 3 Å after 6 ns in the case of LMTK3, while in the case of CDK5 the RMSD value settles around 4 Å after 8 ns. The stability of the complex structure can be seen from RMSD analysis. The snapshots of CDK5-LMTK3 complex at different intervals of simulation time are shown in **Figure. 7a.11**.

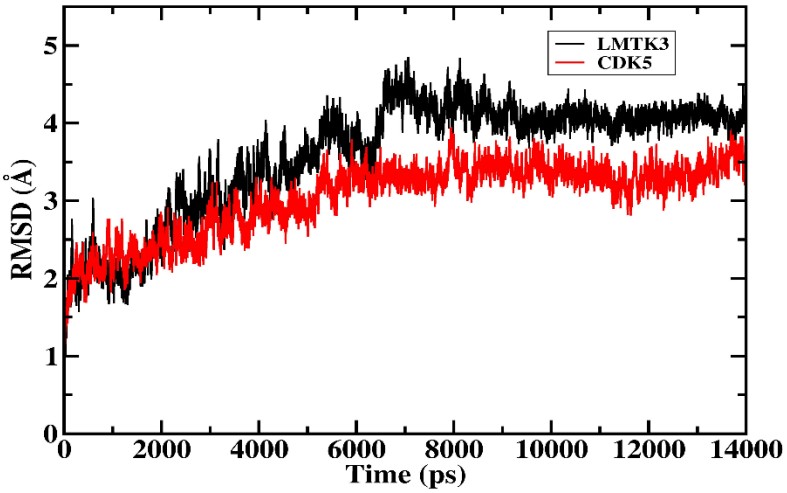

Figure 7a.10. *RMSD of CDK5 and LMTK3 as a function of simulation time.*

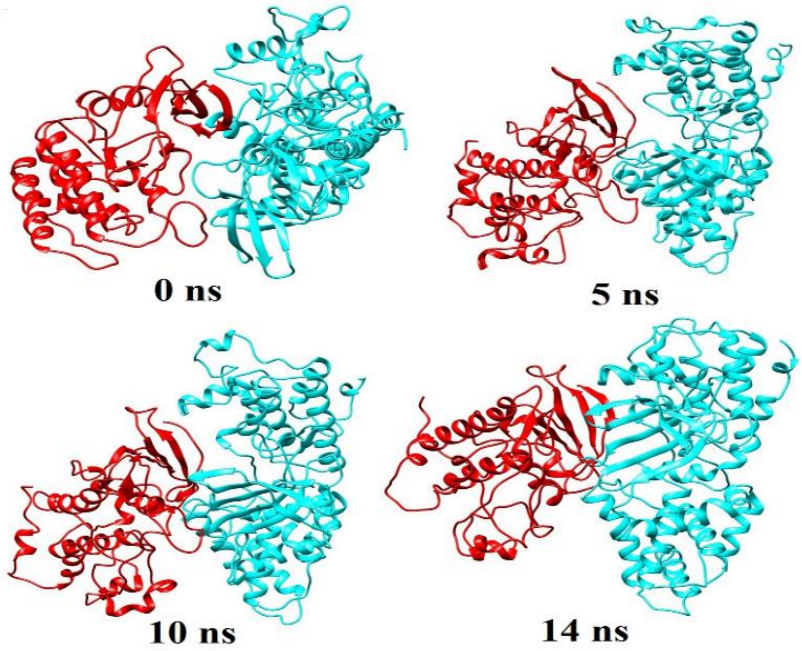

Figure 7a.11. *Snapshots of CDK5-LMTK3 complex at different time intervals of MD simulation*

We also analyzed the number of intermolecular hydrogen bonds **Figure 7a.12** between CDK5 and LMTK3 as a function of time. For calculating the hydrogen bond, the cut-off for angle and distance was set to 120° and 3.5 Å, respectively. From **Figure 7a.12** we see the number of intermolecular hydrogen bonds in the complex to be oscillating around 8-12 during simulation. Thus we can infer the intermolecular hydrogen bonds play the significant role in stabilizing the CDK5-LMTK3 complex. The CDK5-LMTK3 complex model structure that we have predicted in this study can also be tested and analyzed experimentally using biochemical techniques, such as co-immunoprecipitation or pull down assay and Surface plasmon resonance.

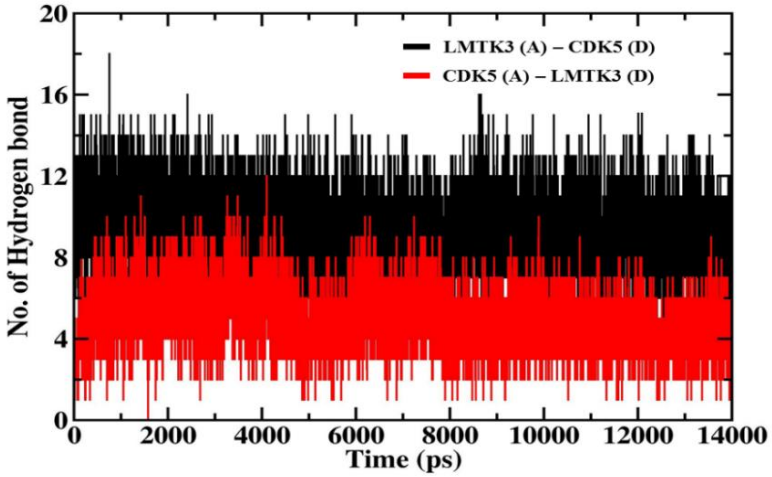

Figure 7a.12. *Number of inter-molecular hydrogen bonds in CDK5-LMTK3 complex model structure as a function of simulation time period. Here we have considered inter-molecular hydrogen bonds involved in two cases: LMTK3 as acceptor and CDK5 as donor (black line) and the other one, LMTK3 as donor and CDK5 as acceptor (red line)*

7a.5. Conclusions

In this work, we determined the probable activation segment and phosphorylation sites in LMTK3. The activation segment ranges from Tyr164 to Ser196 in LMTK3 and probable phosphorylation sites found to be at Thr167 and Thr189 in activation segment. From the CDK5 and LMTK3 interaction study we observed that the residues near the C-terminal region of CDK5 (Tyr179, Leu178 and Tyr158, Arg 156) involved in long range interactions with predicted phosphorylation sites (Thr167 and Thr189) in LMTK3. Thus, we see that CDK5 may phosphorylate LMTK3 at Thr167/Thr189 and leads to breast cancer progression. In addition, our findings in this work also highlighted the probable interactions and interface area between CDK5 and LMTK3 that may play an important role in LMTK3 phosphorylation in breast cancer tumorigenesis. Thus, we suggest that the probable interacting interface regions or binding sites may be targeted to control LMTK3 phosphorylation. With the knowledge of interacting residues in LMTK3 and CDK5, it may be possible to have control over the role of CDK5 in LMTK3 phosphorylation. Therefore, targeting the CDK5 and LMTK3 interacting interface residues may be helpful for the discovery of new drug targets.