CHAPTER 2

INTRODUCTION & REVIEW OF LITERATURE

INTRODUCTION AND REVIEW OF LITERATURE

2.1. Breast Cancer

Nowadays, breast cancer is very common in female and it is the second leading cause of death after lung cancer and making up to 25% of all cancer cases. In 2012, with an estimate of 1.67 million people diagnosed globally [16]. The incidence of breast cancer in less developed regions (883 000 cases) is slightly more as compared to more developed (794 000) regions. [17] In India, incidence rate of breast cancer is lower (25.8 per 100 000) than United Kingdom (95 per 100 000) but mortality rate is 12.7 versus 17.1 per 100 000 with United Kingdom [18]. Breast cancer incidence rates within India exhibit a 3–4-fold variation across the country, and in the Northeast, breast cancer rate was observed to be the highest and also in major metropolitan cities such as Mumbai and New Delhi [19]. Even with the increase in incidence rates, the breast cancer patient's survival rates have been improved due to early diagnosis, advancement in surgery practice and the introductions of various therapies, such as radiation-therapy, systemic chemotherapy, endocrine therapy and other targeted therapies [20-22]. However, breast cancer with metastasis, still remains the leading cause of death Worldwide. Therefore, it is essential to understand the mechanisms underlying carcinogenesis, invasion, metastasis, and resistance to therapies as well as identifying new targets could help us to develop more efficient and effective therapies that could help to improve patient's prognosis.

2.1.1 Classification of breast cancer

Based on the presence of receptor status, including hormone receptors (HR): estrogen receptors (ER) and progesterone receptors (PR), human epidermal growth factor receptor 2 (HER2), breast cancer has been classified into the following 4 groups:

- Luminal A (HR+/HER2-): These cancers account for 71 % and have a tendency to grow slowly and less aggressive than other subtypes. Luminal A cancers have the most favorable short term prognosis, due to high responsive to anti-hormone therapy [23].

- Triple negative (HR-/HER2-): These cancers account for 12 %, and in these cancers, ER, PR, and HER2 do not express. In premenopausal women these cancers are observed to be more common and with a *BRCA1* gene mutation [24]. At present there are no targeted therapies for these tumors, due to which triple negative breast cancers have a poorer short-term prognosis than other subtypes [25].

- Luminal B (HR+/HER2+): These cancers account for 12 %, and like luminal A cancers, luminal B cancers are ER+ and/or PR+ and are highly positive for Ki67 (indicator of actively dividing cells in a huge manner) or HER2. Luminal B breast cancers are aggressive and have poorer survival rate than luminal A cancers [23].

- HER2-enriched (HR-/HER2+) These cancers account for 5%, and have a tendency to grow and spread more aggressively than other subtypes and are associated with poorer short-term prognosis as compared to HR+ breast cancers [23]. However, the recent extensive use of targeted therapies have better outcomes for HER2+ cancer patients.

Among the three markers, ER has been the target, and benefited the therapy strategies for a long period. In 1960s, ERα was identified and the gene was cloned in 1985. Early in the 1970s, on the basis of the presence of ERα, breast cancer was divided into two disease subclasses, namely ERα negative and ERα positive breast cancer. In the early stage of breast cancer, most of the tumours are ERα positive, while in advanced stage, the proportion of ERα negative cancer increases [26]. In 1996, second estrogen receptor gene, encoding ERβ was identified. ERα is necessary for the development of mammary glands and is expressed in about 70-80% of breast tumours. A clear role for ERβ has not properly established yet, but some groups have described its expression in breast cancer [27,28] .

The ERα signaling pathway is a complex biological pathway that controls various cancer cell functions, such as cell proliferation, anti-apoptosis, invasion and angiogenesis.

ERα is activated when estrogen binds to the ligand binding domain of ERα, the

activated ERα then translocate to the nucleus and binds to DNA and modulate numerous genes expression [29,30].

ERα and ERβ are members of the nuclear receptor superfamily of transcription factors. The human ERα protein contains 595 amino acids. The C-terminal domain of ERα is highly conserved and structured upon ligand binding to its domain, whereas the N-terminal domain is poorly conserved and is less structured. N-terminal domain gets activated by both ligand dependent and ligand-independent mechanism [31-33]. ERα consists of six structural domains shown in **Figure 2.1**, that contain a number of functional domains: Activation function-1 (AF-1); DNA-binding domain (DBD) which mediates binding to promoters of estrogen responsive genes; hinge region; C-terminal, ligand binding domain (LBD) where estrogen binds; and activation function-2 (AF-2) domain. Estrogen binding results in ERα activation, allowing ERα recruitment to gene promoters and subsequent regulation of gene expression. Transcriptional regulation by ERα upon estrogen binding involves the recruitment of transcriptional co-regulator proteins to the LBD, this is called ligand dependent activation of ERα. The Mechanisms of ER action in breast cancer is shown in **Figure 2.2**. The ERα function is also known to be regulated by ligand-independent mechanism through various kinases, including mitogen activated protein kinase (ERK1/2 MAP kinase) [34], CDK2/CyclinA [35], and protein kinase A (PKA) [36]. These kinases phosphorylate ERα mainly at N-terminal region (AF-1), and can activate ERα (**Figure 2.2**).

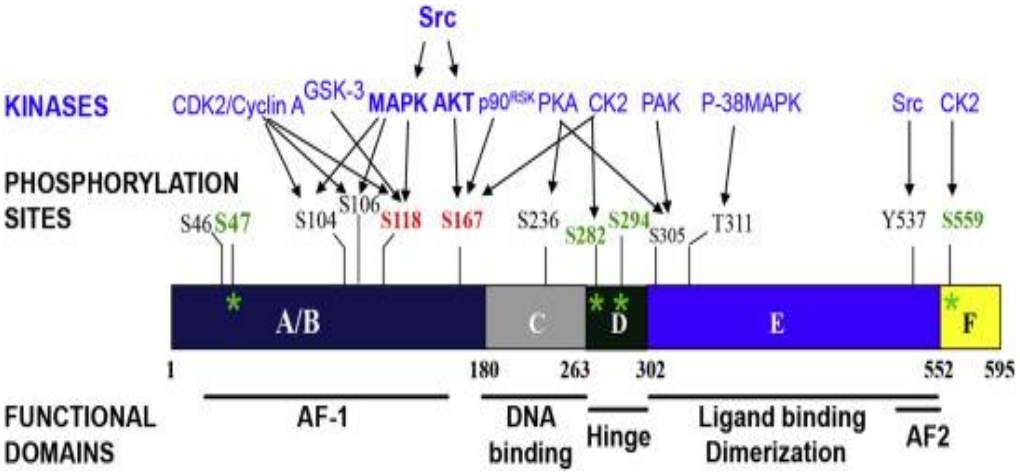

Figure 2.1. *ERα structure and phosphorylation sites. ERα consists of six structural domains that contain a number of functional domains: Activation function-1 (AF-1), DNA-binding domain (DBD), hinge, ligand binding domain (LBD), dimerization, activation function-2 (AF-2), and the F domain considered to repress ERα function. Taken from* [37]

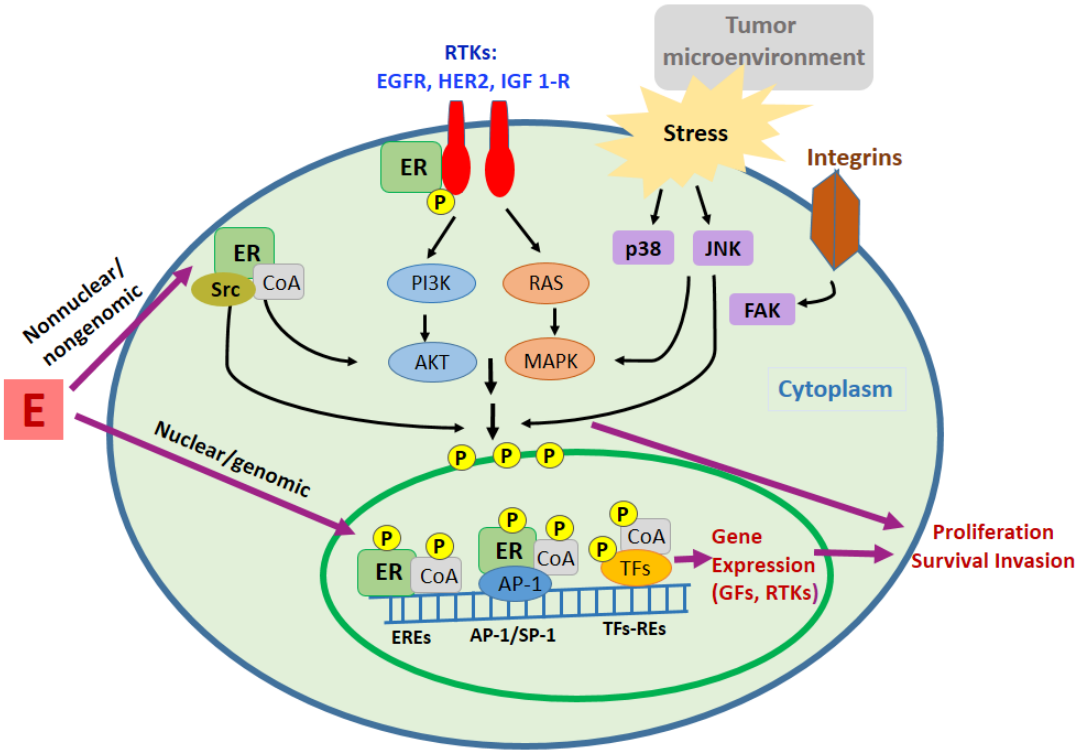

Figure 2.2. *Mechanisms of estrogen receptor (ER) action in breast cancer. Modified figure from* [38]

2.1.2. Therapies

2.1.2.1. Endocrine therapies

(i) Since ERα is a breast cancer classification marker, it is the key therapeutic target in breast cancer [39]. Estrogen is responsible for the growth of ER positive cancer cells, therefore, in therapies, anti-estrogen, primarily tamoxifen is used to target ERα, tamoxifen compete with estrogen for binding to ERα, and prevents its activation [40].

(ii) Other strategies for inhibiting ERα action include aromatase inhibitors which suppress the amount of estrogen. The aromatase enzyme plays a critical role in the synthesis of estrogens from testosterone and androstenedione. After 2 to 3 years of tamoxifen therapy, switching to an aromatase inhibitor gives better responses than tamoxifen alone [41, 42]. The current generation of aromatase inhibitors such as Anastrazole and letrozole, provide an improved disease-free survival and overall survival compared to tamoxifen alone.

(iii) Additionally, Falsodex is an ER antagonist that binds to the receptor and promotes its degradation, leads to the loss of cellular ER [43] and is used as a second line agent following recurrence after tamoxifen treatment [44]. Generally, ER+ breast cancers have a better prognosis. On the other hand resistance to these therapies are commonly occurring due to the ligand-independent activation of ERα by various kinases [34-36].

2.1.2.2. Drugs targeting Receptor Tyrosine Kinases (RTKs)

Similar to ERα, HER2 is also a key target, especially in HER2 enriched breast cancers. HER2 is a receptor tyrosine kinase (RTK), which responds to the growth signals and directs normal breast cell growth. In HER2 enriched breast cancers, the amount of HER2 increases and results in uncontrolled cell growth and tumour progression. Therefore, before modern treatments, HER2 positive breast cancer had worse prognoses [27]. However, HER2 positive breast cancer cells respond to drugs such as trastuzumab, which targets HER2, suppresses HER2 activation signalling and therefore, it significantly improves the prognosis [45]. Triple negative cancers have

poorer scenarios as compared to other breast cancer subtypes, due to the lack of positive receptors thus lack of targeted therapy [46].

2.1.2.3. Chemotherapy and radiotherapy

Chemotherapy is mainly used to treat triple-negative breast cancer. Doxorubicin and docetaxel are chemotherapy drugs which, inhibit cell division and proliferation during cell replication process by inducing damage to DNA. After surgery, radiotherapy is generally used to remove remaining or escaped microscopic tumour cells from surgery that might have potential to form tumour again [47].

2.1.3. Endocrine therapy resistance

Even though endocrine therapies have shown a significant improvement in outcomes, but one-third of women treated with tamoxifen for 5 years, deteriorate after a period of improvement, within 15 years, [48]. A great effort has been made to understand the molecular mechanisms of endocrine therapy resistance in ER-positive breast cancers. The clinical indications and molecular features of the endocrine resistance [49] in breast cancer are:

- Decrease or loss of ER: Nearly 20% of patients treated with endocrine therapy loss ER. Then growth of these tumours are no longer driven by estrogen, thus anti-estrogen treatments are no longer effective [50,51].
- After endocrine therapy, upregulation of HER2 in a few patients: the increased HER2 abundance may consequently take over the role of ER in tumour progression [51, 52].
- ERα modifications: Modifications of ERα by phosphorylation and methylation are considered to be the main cause of tamoxifen resistance that is highly related to clinical outcome [53].
- ER co-regulators: The ER transcriptional activity requires the recruitment of transcriptional co-regulator proteins. Expression of a number of important co-regulators, particularly the p160 coactivators SRC1 and transcriptional coactivator Amplified in Breast Cancer 1 (AIB1) has been linked to non-response to tamoxifen [54].

- Increased RTK signaling leads to the activation of the extracellular signal–regulated kinases (ERK) or classical MAP kinases and phosphoinositide 3-kinase (PI3K) pathways: PI3K is revealed to activate numerous molecules in survival pathways and may promote ER transcriptional activity even independent of estrogen [55], whereas ERK1/2 activation may decrease ER expression [56].

Thus, many clinical approaches have focused on co-targeting these features together with ER to overcome endocrine resistance, among which, mostly kinase pathways are co-targeting and mainly investigated.

2. 2. Protein kinases

Protein kinases are the largest enzyme family involved in cell signal transduction and they are the one of the most important targets in current cancer therapy [57, 58]. Protein kinases encoded by approximately 2% of eukaryotic genes and more than 500 protein kinases are identified based on the human genome sequencing and biochemical studies [59]. Protein kinases regulate protein modifications by transferring gamma (γ) phosphate of adenosine-5'-triphosphate (ATP), or another nucleoside triphosphate, to specific amino acids with a free hydroxyl group on the other protein, and this process is called phosphorylation. Phosphorylation plays an important role in cellular signaling transduction, leading to the activation or inhibition of biological activity, increases or decreases the molecule movement and facilitates or disrupts protein-protein interactions. As most protein kinases have more than one substrate, thus they are classified based on the receiver amino acid specificity rather than protein substrate specificity. Kinases can act on either serine/threonine (Ser/Thr) or tyrosine (Tyr), as well as on all three [60]. Due to the fact that phosphorylation has deep impact on cellular signalling, so protein kinase activity is highly regulated. The activity of a kinase can be turned on by phosphorylation, which can be mediated by the kinase itself (autophosphoylation), or by other kinases. The phosphorylation status can also be regulated by other proteins such as phosphatase which removes phosphate group (dephosphoryaltion) or by small molecules which inhibit or activate protein kinases [61].

In different types of cancer, constitutive or inappropriate activation of protein kinases are seen, wherein small molecule inhibitors are designed to target/inhibit kinase signaling in such situations [62]. Since, protein kinases have been identified as one of the most important targets in the current cancer therapy, one of the ways to inhibit abnormal kinase activity is to design small molecule inhibitors to bind into the ATP binding pocket in order to block ATP binding and therefore inhibit the phosphorylation. There are some small-molecule ATP-competitive inhibitors such as imatinib (Gleevec) for the treatment of chronic myeloid leukemia (CML) and gastrointestinal stromal tumors (GIST) which are found to be effective in treating these diseases [63]. Despite of the success of Gleevec, the designing of selective (specific) ATP-competitive inhibitors for a particular kinase seem to be quite challenging due to the conserved nature of kinase domain structures. Therefore, understanding the inhibition and phosphorylation of protein kinases are significant to guide cancer therapy. In this thesis, we are mainly focusing on Human Lemur Tyrosine Kinase 3 (LMTK3) and its implications in breast cancer.

2.2.1. General architecture of protein kinase

Studies based on the first X-ray structures of cAMP dependent protein kinase (PKA) [64], a conserved core of a typical protein kinase consists of two lobes, the small N-terminal lobe (N-lobe) and large C-terminal lobe (C-lobe) (**Figure. 2.3 A**). The two lobes form a deep cleft, and forming the active site, that accommodates one ATP molecule bound to one or two divalent cations: Mg^{2+} or Mn^{2+}. These cations compensate for the strong negative charge of the phosphates of ATP [65] and provide their coordination inside the active site [66].

2.2.1.1. N-terminal lobe

The N-lobe usually includes five β-strands and an α-helix (αC) (**Figure 2.3 B**). Despite the fact that the β-strands form a relatively rigid antiparallel β sheet, the N-lobe is very dynamic and flexible. Typically inactivation of a protein kinase involves a significant swing motion of the αC-helix directed outward from the active site (**Figure 2.3 C**). Such motion disrupts multiple interactions both within the N-lobe and between the two lobes, and devastating protein kinase activity. In some kinases a short α-helix (αB)

precedes αC. It is relatively weak and can easily unfold to accommodate the αC-helix motion. The N-lobe also contain a flexible loop between the β1 and β2 strands that usually include three conserved glycines and thus termed the glycine-rich loop (G-loop) (**Figure 2.3 B**) [67,68]. This loop covers the β and γ-phosphates of the ATP and plays an important role in both phosphoryl transfer and ATP exchange during the catalytic cycle [68].

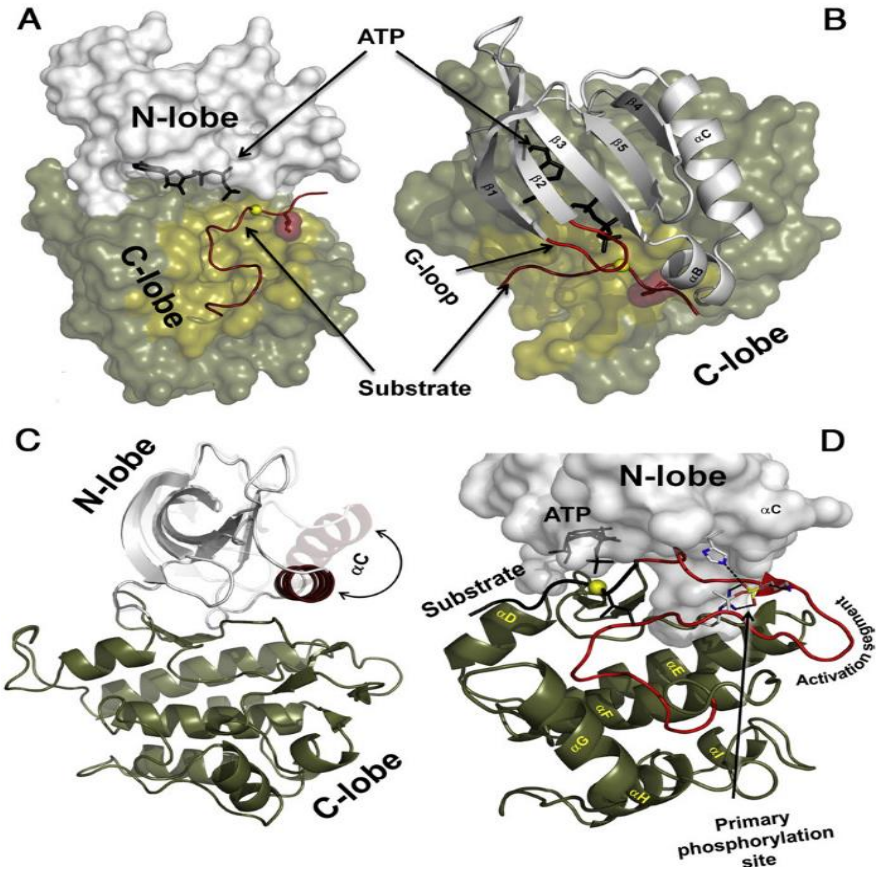

Figure 2.3. *Structure of conserved protein kinase core. Taken from* [52]

2.2.1.2. C-terminal lobe

The C-lobe is much more stable and mostly α-helical. It contains seven helices (αD through αI) and four very short β-strands (β6 through β9) (**Figure 2.3 D**). The C-lobe serves as a docking site for substrate peptides/proteins. The C-terminus residue, right after the phosphorylation site is buried in a pocket formed by the so-called P+1 loop. This loop is located in the C-terminal part of an extended activation segment. In protein kinases this activation segment is the most important regulatory element. Its

conformation can influence both substrate binding and catalytic efficiency [69-71]. The activation segment contains a magnesium binding loop at its N-terminal and the activation loop in the middle. The activation loop ranges from 20–35 residues, located between a conserved Asp-Phe-Gly (DFG) motif, and the APE or SPE motif (**Figure 2.4.B**) [71,72]. The conserved DFG motif is critically important for catalysis. The DFG-aspartate directly binds to divalent cations (Mg^{2+} or Mn^{2+}) in the active site. Usually the activation loop contains a serine/threonine or tyrosine residue that can be phosphorylated by other protein kinases or autophosphorylated via a *trans* mechanism. *Cis* mechanisms for activation loop autophosphorylation are relatively rare [73].

The negatively charged phosphate makes a set of strong hydrogen bonds and serves as an organizer both for the active site and the substrate binding surface. Some kinases have second or even third phosphorylation sites in the activation loop and perform functions specific for these kinases. There are also some kinases that do not have the phosphorylation site and these usually are constitutively active [70].

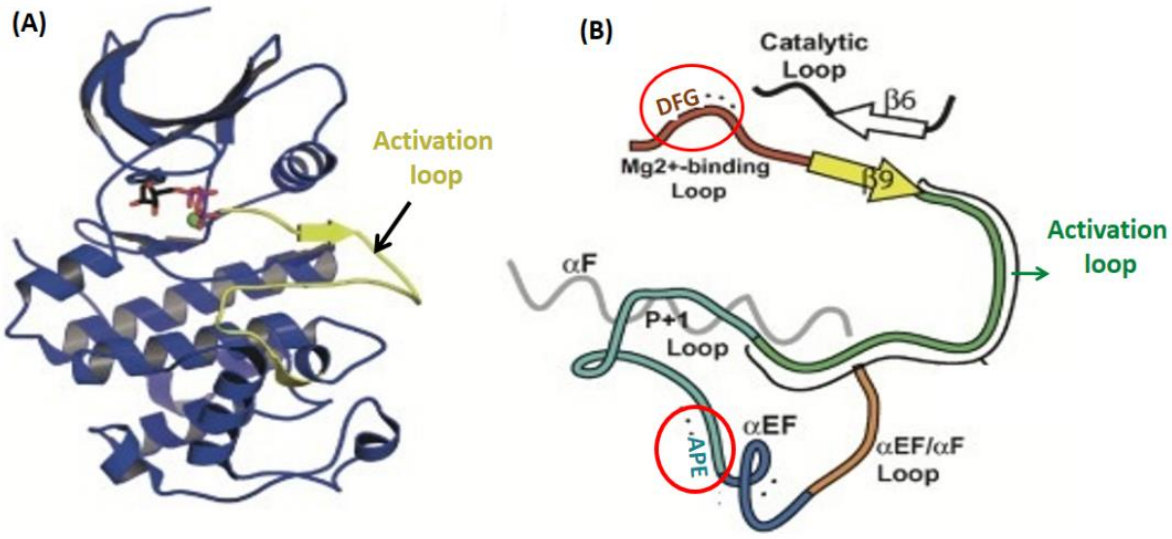

Figure 2.4. *(A) Protein kinase A, representing activation loop in yellow colour; (B) Representation of activation loop showing DFG and APE motif. Taken from* [71]

2.2.2 Active and inactive structures of protein kinases

In the kinase domain the "activation loop" is highly flexible (**Figure 2.4 B**), which adopts an extended conformation to function as a binding platform for the peptide substrate [74]. In the catalytically active kinase the activation loop is displaced and allows the protein substrates and ATP to bind. The movement of the activation loop is linked to the rotation of the DFG motif from inactive DFG-out to active DFG-in conformations. As a result, the activation loop is ''open" in the active kinase state and often ''closed" in the inactive kinase state [71] **(Figure 2.5)**.

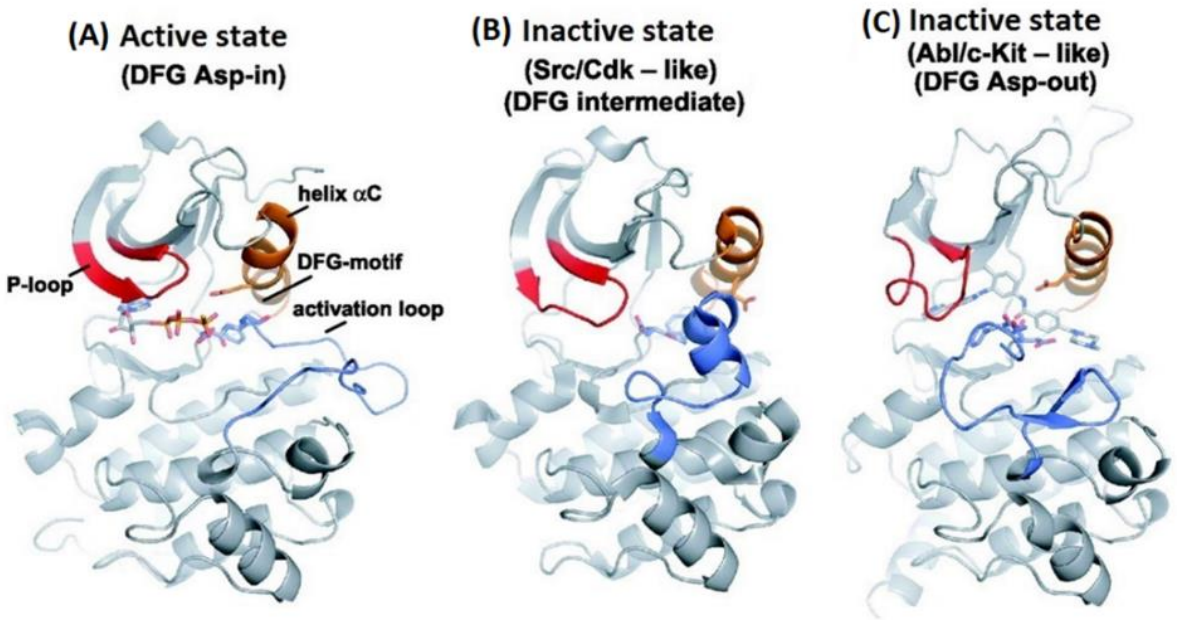

Figure 2.5. *Active and inactive conformation of kinase. Taken from* [75]

2.3. Lemur Tyrosine kinase-3 (LMTK3)

2.3.1. The LMTK family

LMTK (LMR or AATYK) family belongs to the receptor tyrosine kinases (RTKs). It consists of three members:

- LMTK1 (also named LMR1 or AATK),
- LMTK2 (also named LMR2, BREK, KPI2 orAATYK2) and
- LMTK3 (also named LMR3, AATYK3).

LMTK1 and LMTK2 genes are relatively evolutionary conserved [8], while LMTK3 has been positively selected from chimpanzees to human, except its catalytic core domain [5]. LMTK1 has an alternative splicing variant (LMTK1B) that does not contain a transmembrane domain (TM), while the rest of the LMTK families including LMTK1A, LMTK2 and LMTK3 contain the N-terminal transmembrane domains (TMs) (**Figure 2.6**). All the three LMTK members are highly sequence conserved in their N-terminal kinase domains as well as in short fragmentary homologous stretches in the N-terminal region, kinase domain-flanking regions, and C-terminal region (**Figure 2.7**). LMTK family has several amino-acid substitutions in their conserved kinase domain compared with other members of the protein tyrosine kinase (PTK) family. In addition, the C-terminal PP1 C-interacting motif of LMTK2 [76] is conserved in all three members. These proteins contain several binding motifs such as src homology 2 and 3 (SH2 and SH3) domains, suggesting that it could interact with a variety of adapter proteins that can modulate intracellular signaling [77, 78].

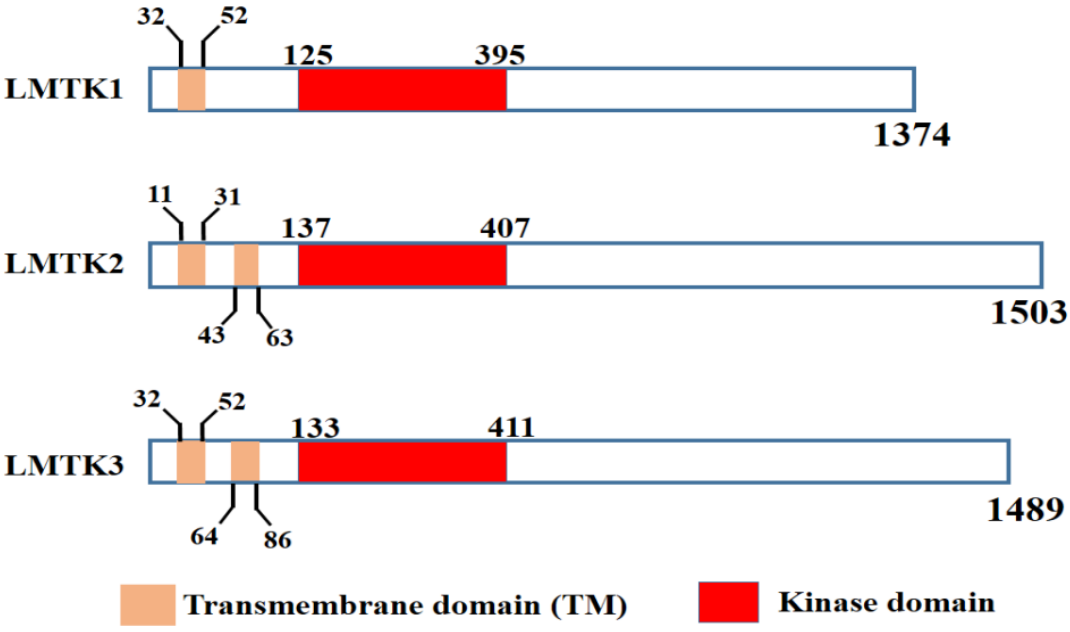

Figure 2.6. *The structure of LMTK family: The LMTK family members (LMTK1A, LMTK2 and LMTK3) contain relatively conserved transmembrane domains and catalytic/kinase domains. Taken from* [79]

```
CLUSTAL O(1.2.4) multiple sequence alignment

SP|Q6ZMQ8|LMTK1_HUMAN ------------MSSSFFNPSFAFSSHFDPDGA--PLSELSWPSSLAVVAVSFSGLFAV 45
SP|Q8IWU2|LMTK2_HUMAN MPGPPALRRRLLLLLLVLLIAGSAGAAPLPQTGAGEAPPAAEV-SSSFVILCVCS-LIIL 58
SP|Q96Q04|LMTK3_HUMAN MPAPGAL------ILLAAVSASGCLASPAHPDGF--ALGRAPLAPPYAVVLISCSGLLAF 52
                              :    .  . . ::     *              *:    * *: .

SP|Q6ZMQ8|LMTK1_HUMAN IVLMLAC-LCCKKGGIGFKEFENAEGDEYAADLAQGS-PATAAQNGPDVYVLPLTEVSLP 103
SP|Q8IWU2|LMTK2_HUMAN IVLIANCVSCCKDPEIDFKEFEDNFDDE--IDFTPPAEDTPSVQSPAEVFTLSVPNISLP 116
SP|Q96Q04|LMTK3_HUMAN IFLLLTC-LCCKRGDVGFKEFENPEGEDCSGEYTPPAEETSSSQSLPDVYILPLAEVSLP 111
                      *.*:  *  ***   :.*****:  .::   : :  :  : : *.  :*: * : ::***

SP|Q6ZMQ8|LMTK1_HUMAN MAKQPGRSVQLLKSTDVGRHSLLYLKEIGRGWFGKVFLGEVNSGISSAQVVVKELQASAS 163
SP|Q8IWU2|LMTK2_HUMAN APSQFQPSVEGLK-SQVARHSLNYIQEIGNGWFGKVLLGEIYTGTSVARVIVKELKASAN 175
SP|Q96Q04|LMTK3_HUMAN MPAPQPSHSDMTTPLGLSRQHLSYLQEIGSGWFGKVILGEIFSDYTPAQVVVKELRASAG 171
                           :  .   :.*: * *::*** ******:***: :. : *:*:****:***.

SP|Q6ZMQ8|LMTK1_HUMAN VQEQMQFLEEVQPYRALKHSNLLQCLAQCAEVTPYLLVMEFCPLGDLKGYLRSCRVAESM 223
SP|Q8IWU2|LMTK2_HUMAN PKEQDTFLKNGEPYYILQHPNILQCVGQCVEAIPYLLVFEFCDLGDLKAYLRSEQEHMRG 235
SP|Q96Q04|LMTK3_HUMAN PLEQRKFISEAQPYRSLQHPNVLQCLGLCVETLPFLLIMEFCQLGDLKRYLRAQRPPEGL 231
                         **  *:.: :**  *:* *:***:. *.*. *:**::*** ***** ***: :

SP|Q6ZMQ8|LMTK1_HUMAN AP-----DPRTLQRMACEVACGVLHLHRNNFVHSDLALRNCLLTADLTVKIGDYGLAHCK 278
SP|Q8IWU2|LMTK2_HUMAN D-----SQTMLLQRMACEVAAGLAAMHKLHFLHSDLALRNCFLTSDLNVKVGDYGIGFSR 290
SP|Q96Q04|LMTK3_HUMAN SPELPPRDLRTLQRMGLEIARGLAHLHSHNYVHSDLALRNCLLTSDLTVRIGDYGLAHSN 291
                            :   ****. *:* *:  :*  :::*********:**:**.*::****:....

SP|Q6ZMQ8|LMTK1_HUMAN YREDYFVTADQLWVPLRWIAPELVDEVHSNLLVVDQTKSGNVWSLGVTIWELFELGTQPY 338
SP|Q8IWU2|LMTK2_HUMAN YKEDYIETDDKKVFPLRWTAPELVTSFQDRLLTADQTKYSNIWSLGVTLWELFDNAAQPY 350
SP|Q96Q04|LMTK3_HUMAN YKEDYYLTPERLWIPLRWAAPELLGELHGTFMVVDQSRESNIWSLGVTLWELFEFGAQPY 351
                      *:***  * ::  .**** ****: ..:. ::..**:: .*:******:****: .:***

SP|Q6ZMQ8|LMTK1_HUMAN PQHSDQQVLAYTVREQQLKLPKPQLQLTLSDRWYEVMQFCWLQPEQRPTAEEVHLLLSYL 398
SP|Q8IWU2|LMTK2_HUMAN SNLSNLDVLNQVIRERDTKLPKPQLEQPYSDRWYEVLQFCWLSPEKRPAAEDVHRLLTYL 410
SP|Q96Q04|LMTK3_HUMAN RHLSDEEVLAFVVRQQHVKLARPRLKLPYADYWYDILQSCWRPPAQRPSASDLQLQLTYL 411
                       : *: :**  .:*::. ** :*:*:   :* **:::* **  * :**:*.:::  *:**

SP|Q6ZMQ8|LMTK1_HUMAN CAKGATEA--------EEEFERRWRSLRPGGGGVGPGPGAAGPMLGGVVELAAASSFPLL 450
SP|Q8IWU2|LMTK2_HUMAN RLQSQRDS--------EVDFEQQWNALKPNTN---------------SRDSSNNAAFPIL 447
SP|Q96Q04|LMTK3_HUMAN LSERPPRPPPPPPPPRDGPFPWPWPPAH-----------------SAPRPGTLSSPFPLL 454
                        :             :  *   *   :                     :  : **:*

SP|Q6ZMQ8|LMTK1_HUMAN EQFAGDGFHADGDDVLTVTETSRGLNFEYKWEAGRGAEA---------------------- 489
SP|Q8IWU2|LMTK2_HUMAN DHFARDRLGREMEEVLTVTETSQGLSFEYVWEAAKHDHFDERSRGHLDEGLSYTSIFYPV 507
SP|Q96Q04|LMTK3_HUMAN DGFPG----ADPDDVLTVTESSRGLNLECLWEKARRGAGRGGGA---------------- 494
                      : *       : ::******:*:**.:*  ** .:

SP|Q6ZMQ8|LMTK1_HUMAN --F-PATLSPGRTARLQELCAPDGAPPGVVPVLSAHSPSLGSEYFIRLEEAAPA-AGHD- 544
SP|Q8IWU2|LMTK2_HUMAN EVFESSLSDPGPGKQDDSGQDVPLRVPGVVPVFDAHNLSVGSDYYIQLEEKSGSNLELDY 567
SP|Q96Q04|LMTK3_HUMAN PAWQPASAPPAPHANPSNPFYEALSTPSVLPVISARSPSVSSEYYIRLEEHGSP-PEPLF 553
                        :  :   *.   . ..        *.*:**:.*:. *:.*:*:*:*** .

SP|Q6ZMQ8|LMTK1_HUMAN -PDCAGCAPSPPATADQDDDSDGSTAASLAMEPLLG-HGPPVDVPWGRGDHYPRRS-LAR 601
SP|Q8IWU2|LMTK2_HUMAN PPALLTTDMDNPERT-------GPELSQLTALRSVELEESSTDEDFFQSSTDPKDSSLPG 620
SP|Q96Q04|LMTK3_HUMAN PNDWDPLDPGVPAPQAPQAPS---EVPQ-----------LVSETWAS------------ 586
                           . *              .            ..  :

SP|Q6ZMQ8|LMTK1_HUMAN D------PLCPSRSPS---------PSAG-PLSLAEGGAEDADWGVAAFCP--AFFEDPL 643
SP|Q8IWU2|LMTK2_HUMAN DLHVTSGPESPFNNIFNDVDKSEDLPSHQKIFDLMELNGVQADFKPATLSSS---LDNPK 677
SP|Q96Q04|LMTK3_HUMAN -------PLFPAPRPF---------PAQS----SASGSFLLSGWDPEGRGAGETLAGDP- 625
                             *  *             *:       . .   :.:            :*

SP|Q6ZMQ8|LMTK1_HUMAN GTSPLGSSGAPPLPLTGEDELEEVGARRA----------AQRGHWRSNVSANNNSG---- 689
SP|Q8IWU2|LMTK2_HUMAN E-----------SVITGHFEKE--KP--------------RKIFDSEPLCLSDNLMHQDN 710
```

Figure 2.7. *Sequence alignment of human LMTK1A, LMTK2 and LMTK3. The sequence similarity alignment was carried using Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/) Similarity scores are represented by* asterisk mark

2.3.2. LMTK3 and its structure

LMTK3 gene is located at 19q13.33 on chromosome 19. LMTK3 encodes 1460 amino acids. Similar to the other two members, LMTK3 is also classified as a tyrosine kinase, but found to be able to phosphorylate Ser/Thr residues. LMTK3 is also composed of an N-terminal TM domain, a relatively conserved kinase domain, and a proline-rich C-terminal domain, LMTK3 has also showed an aberrantly high molecular weight on SDS-PAGE [9], which might be due to some post-translational modifications or may be due to its low electrophoretic mobility because of the proline-rich domains. Mass spectrometry analysis showed that both the bands of LMTK3 on SDS-PAGE were full-length LMTK3 [9]. Till now there are no crystal structures yet for these kinases, which are partly due to the fact that all these kinases comprise of large intrinsically disordered regions, whose structures are unlikely to be determined. The disordered regions of LMTK3 are depicted in **Figure 2.8**. Intrinsically disordered regions (IDRs), are devoid of stable secondary and/or tertiary structures under physiological conditions and therefore cannot be assigned in X-ray datasets. Most proteins are neither fully ordered nor fully disordered but contain ordered and disordered regions at different ratios, as we can see in case of LMTK3 also (**Figure 2.8)**. More precisely, disordered regions: 1) are frequently subjected to posttranslational modifications, which are of great importance for cellular signaling transduction; 2) function in scaffolding and the recruitment of binding partners; 3) are responsible for conformational variability and adaptability [80]. IDPs are often enriched in proline-rich repeats, glycine and alanine, which are called “disordered promoting amino acids”, while lacking “order promoting amino acids” like Asparagine and Cysteine. This rule is widely used to predict IDRs of proteins and there are several online software programs available using this approach. In literature there are some studies on the 3-dimensional (3-D) structure of LMTK3 [81, 82].

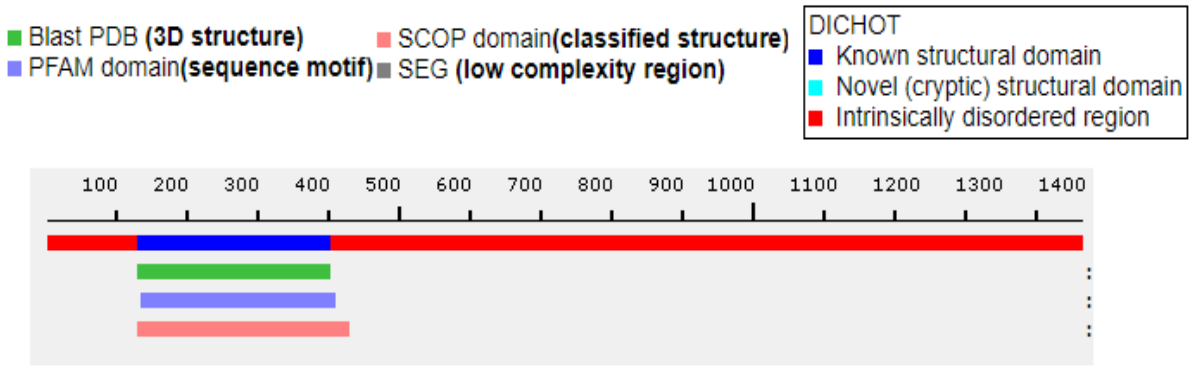

Figure 2.8 *Schematic representation of the order and disorder score in LMTK3. LMTK3 is composed of ordered domain that is a kinase domain, and variable intrinsically disordered domains depicted in red colour. Data obtained using DICHOT (http://spock.genes.nig.ac.jp/~genome/DICHOT/).*

2.3.3. LMTK3 and its association with different types of cancers

Till date the studies of LMTK3 functions are mostly highlighted on cancer. For the first time the involvement of LMTK3 in cancer has been reported in two different screening experiments [83, 84] performed by Jeffrey and colleagues. They carried out an RNAi screening and targeted the entire tyrosine kinases in acute myeloid leukemia (AML) cell lines and observed the cell viability. They have shown that LMTK3 is associated with leukemic cell survival [83]. Another study, using RNAi targeting 691 human kinases in Wnt3a-stimulated human cells has suggested a putative role of LMTK3 in the Wnt/β-catenin pathway [84].

Wakatsuki and colleagues examined the potential implication of LMTK3 in gastric cancer. They predicted that Single nucleotide polymorphism (SNP) rs9989661 T/T of LMTK3 improves breast cancer outcomes, but associated with poorer overall survival (OS) rates in gastric cancer. Another SNP of LMTK3, rs8108419 G/G, was also predictive for poor OS in gastric cancer patients [85].

Surprisingly, another group found the LMTK3 staining at the adjacent non-cancerous gastric mucosa, they noticed LMTK3 abundance in the adjacent regions were notably lower as compared to cancerous regions. By using the samples from gastric patients, they suggested that LMTK3 abundance is significantly associated with

invasion depth and the stage of gastric cancer, as well as poor prognosis. Their work revealed that instead of being protective, LMTK3 could be a negative prognostic marker and a potential therapeutic target in gastric cancer [86].

Another study was carried out to investigate the association of abundance of serum LMTK3 with colorectal cancer (CRC) patient's survival. The patients with higher level of serum LMTK3 were observed to have poorer overall survival as compared to the lower serum LMTK3. They also suggested that for predicting tumour progression in patients with CRC, LMTK3 could be a biomarker, on the basis of the fact that the serum LMTK3 abundance correlates with the histological differentiation, the depth of tumour invasion, and the tumour-node-metastasis (TNM) stage [87,88].

Recently one more study have highlighted the LMTK3 protein and mRNA expression to be higher in thyroid cancer patients as compared to the patients with benign thyroid tumors [89]. They observed LMTK3 knockdown retarded proliferation, invasion and migration in SW579 cells. In addition, LMTK3 downregulation promoted apoptosis in SW579 cells. Their findings indicated that LMTK3 knockdown retards the growth of thyroid cancer cells partly through inhibiting proliferation, invasion, migration and inducing apoptosis in SW579 cells. This study also indicates that LMTK3 could be a useful diagnostic biomarker and a novel therapeutic target for thyroid cancer patients [89]. Very recently the role of LMTK3 is well appreciated as a new therapeutic target for breast cancer [5].

2.3.4. Role of LMTK3 in breast cancer

2.3.4.1. Implication of LMTK3 in ERα positive breast cancer

ERα is an estrogen-stimulated nuclear transcription factor that is expressed more than 70 % [90] of breast tumours thus named as ERα positive breast cancer. The endocrine therapies discussed in section 2.1.2.1, have been used to treat ERα positive breast cancer for the past 30 years [91]. However, intrinsic or acquired resistance to these therapies has always been the main problem in breast cancer. The resistance is stated to be developed due to the activation of RTK and their downstream kinase pathways, which phosphorylate ERα, leading to estrogen-independent activation of ERα and

results in breast cancer progression [34-36]. In 2011 Giamas et al, identified LMTK3 as an oncogenic kinase which regulates ERα activity and plays significant role in breast cancer progression.

2.3.4.2 LMTK3 a regulator of ERα

Giamas et al in 2011 carried out kinome siRNA screening and identified kinases whose silencing alters the estrogen response including those kinases, previously implicated in regulating ERα activity (such as mitogen-activated protein kinase and AKT). LMTK3 was found to be the most potent regulators among all the kinases. In contrast to other modulators of ERα activity, LMTK3 seems to have been subject to Darwinian positive selection, a worthwhile result that gives the unique susceptibility of humans to ERα positive breast cancer [92-96].

Using some biochemical studies Giamas et al, observed that, inhibition of LMTK3 resulted in a decrease of ERα mRNA and protein level, suggesting, LMTK3 regulates ERα at transcriptional as well as translational level. An *in vitro* kinase assay reported the ability of LMTK3 to interact and phosphorylate ERα and protecting it from ubiquitin-mediated proteasomal degradation. Additionally, LMTK3 also affects ERα transcription by positively regulating Forkhead box protein O3 (FOXO3), a known transcriptional activator of *ESR1* gene, possibly via inhibition of PKC-mediated AKT phosphorylation, which leads to FOXO3 degradation [97] (**Figure 2.9**).

Moreover, increased LMTK3 expression associates with more aggressive cancers and poor overall and poor disease-free survival, revealing a potential involvement of LMTK3 in breast cancer progression and suggests LMTK3 expression may be a reliable new biomarker in breast cancer [6-8]. In summary, these investigations revealed that LMTK3 is able to regulate ERα transcriptional activity and stability, which play a role in the development of tamoxifen resistance. These reports suggest that LMTK3 could be a new potential target in ERα positive breast cancer. Now there is a great demand to stop LMTK3 activity by designing potential inhibitors. So, in order to design potential inhibitors of LMTK3, the structure and functioning of LMTK3 need to be understood well.

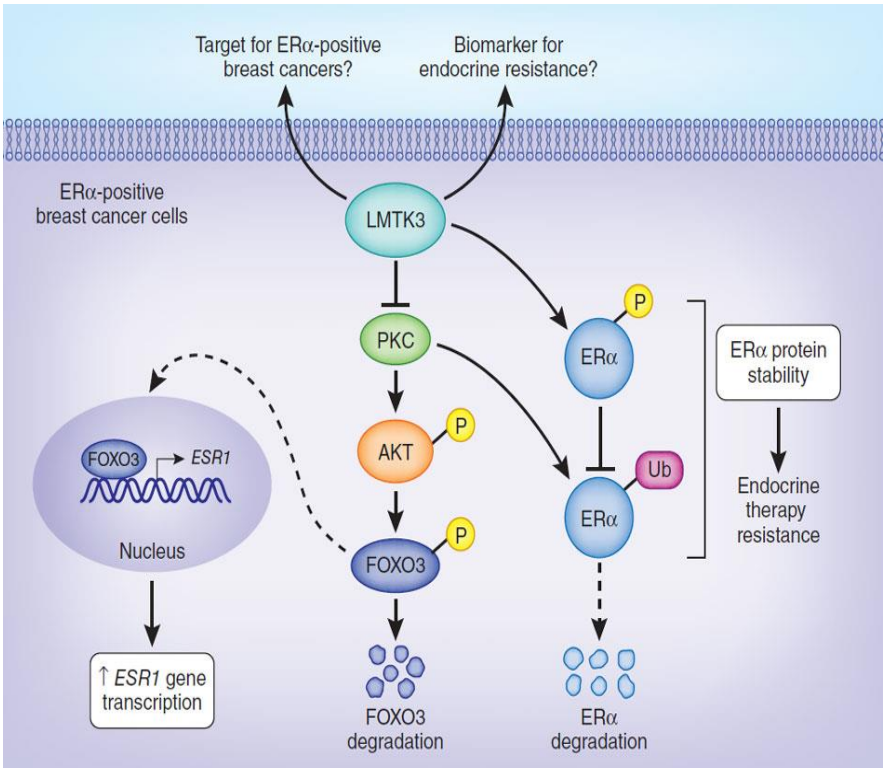

Figure 2.9. *LMTK3 positively controls ESR1 gene transcription, protein levels and phosphorylation, developing resistance against endocrine therapies. LMTK3 increases the transcriptional activation of ESR1 gene by inhibiting PKC activity, whose activity leads to the phosphorylation of AKT, which in turn phosphorylates FOXO3 and degrade it. LMTK3 also directly phosphorylates ERα protein, which protects ERα from ubiquitin (Ub)-mediated proteasomal degradation. Taken from* [98]

Scope of the present work

Since LMTK3 is considered to be the new therapeutic target in breast cancer, so, there is a need to design inhibitors against LMTK3. Before that, we need to understand the salient structural features of LMTK3 functional domain in detail. In the work described in **Chapter 4**, we have modelled 3-D structure of LMTK3 domain using Iterative Threading ASSembly Refinement (I-TASSER) [99] and studied conformational dynamics using molecular dynamics (MD) simulation. And using computational tools and software we carried out comprehensive investigation on the cavities, hydrophobicity, electrostatic potential, secondary structure topology, intra-molecular interactions and also identified the ATP binding pocket in LMTK3.

As described in section 2.3.4.2, LMTK3 is able to interact and phosphorylate ERα, leading to the breast cancer progression and endocrine therapy resistance. To understand the role of LMTK3 in ERα regulation, it is critical to study the molecular interactions between them. So in **Chapter 5** we demonstrated the transient interactions between ERα and LMTK3 using computational techniques. We modelled ERα-LMTK3 complex structure, and analyzed the molecular interactions between them. The C-terminal region of LMTK3 displayed non-bonding interactions with the N-terminal region of ERα. With this interaction study we identified the probable phosphorylation sites in ERα at N-terminal region. From conformational dynamics study ERα-LMTK3 complex found to be stable. The outcomes of this study enhance the understanding of interactions between ERα and LMTK3 which are thought to be critical in signaling pathway in ERα positive breast cancers.

2.3.5. Implication of LMTK3 in triple negative breast cancer promotes tumour invasion and metastasis through GRB2 mediated induction of integrin β_1

In addition to ERα regulation, LMTK3 is involved in other mechanism in breast cancer progression. Xu et al [9] highlighted the LMTK3-regulatory pathways and its involvement in invasion and metastasis. They revealed that elevated cytoplasmic LMTK3 abundance in triple-negative breast cancer promotes tumour invasion and metastasis through GRB2 mediated induction of integrin β_1, provided an example of ERα-independent action of LMTK3 [9] **(Figure 2.10)**. Integrins are composed of two chains α and β, wherein the binding of integrin (α_5, β_1) to the collagen and fibronectin, the components of extracellular matrix (ECM) causes the motility and invasion of cancer cells. Numerous studies have suggested the abundance of altered α_5 and β_1 integrins subunits to be often associated with tumorigenesis and metastasis which are accountable for the increase in disease progression and decrease the patient survival rate [100-104]. Growth factor receptor bound protein 2 (GRB2) is an adaptor protein which is expressed ubiquitously in healthy conditions and over expressed in tissue samples and breast cancer cell lines [105, 106].

GRB2 has been reported to directly associate with receptor tyrosine kinase (RTK), activating downstream RAS guanosine triphosphosphatase (GTPase) and other

extracellular signals that regulate kinase and mitogen-activated protein kinase (ERK/MAPK) [106-108]. Recent study reveals the interaction between LMTK3 with GRB2, enhances the binding of GRB2 with son of sevenless (SOS) that subsequently leads to the activation of RAS and Cell division control protein 42 homolog (CDC42). CDC42, in turn, stimulates the serum response factors (SRF) activity which then binds to integrin promoter, thus inducing the transcriptional activation of integrin [9]. Considering these events, understanding the interaction between GRB2 and LMTK3 at the molecular level considered to be helpful in providing the insights to control this signaling pathway.

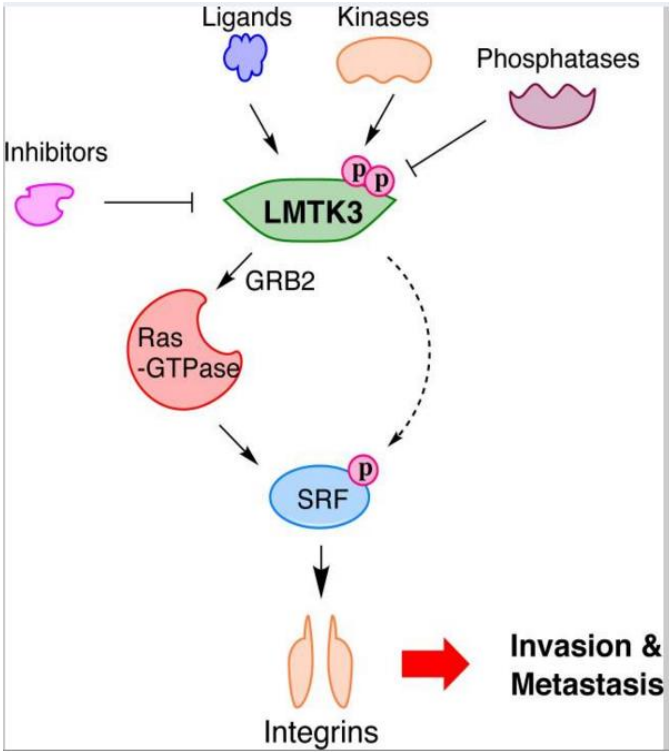

Figure 2.10. *LMTK3 promotes invasion and metastasis in Breast cancer through GRB2-Mediated induction of Integrin β1. Taken from* [109]

GRB2 structure: GRB2 contains one SRC Homology 2 (SH2) domain and two SH3 domains (**Figure 2.11**). The majority of the secondary structure of GRB2 is comprised mainly of beta sheets; all three domains of the protein contain beta sheets. Only two alpha helices are present per subunit. Random coils are intermittent throughout the secondary structure. Turns are less prevalent, again only two turns occur per subunit

[110]. The SH2 domain of Src exhibits a high degree of structural similarity with the SH2 domain of Grb2. Both domains contain two alpha helices surrounded by beta sheets. Functional similarities are also present, as the SH2 domains of both Grb2 and Src are influenced by tyrosine phosphorylation, and both proteins play a role in protein kinase pathways [110].

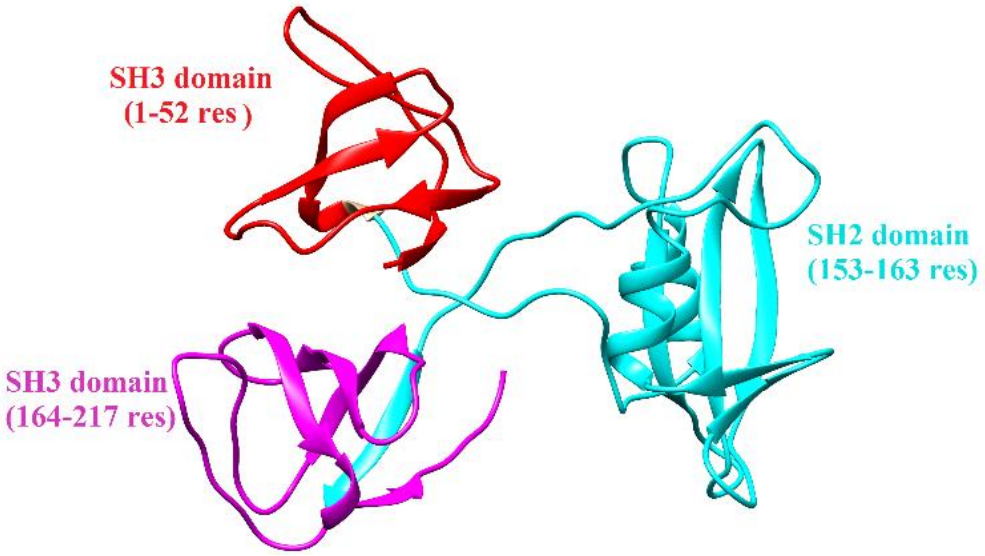

Figure 2.11. *Structure of GRB2 (PDB ID: 1gri) visualized in Chimera*

Scope of the present work

In **Chapter 6** we studied the protein-protein interaction between GRB2 and LMTK3, wherein we identified the probable interacting interface area, interacting residues between them. In our study we observed that the SH3 domain of GRB2 is interacting with N-lobe of LMTK3 and SH2 domain interacting with C-lobe of LMTK3. Similar kind of interaction we can see in src kinase family [111]**.** Understanding the interaction between GRB2 and LMTK3 at the molecular level may be helpful in providing the insights to control this signaling pathway.

2.3.6. Role of phosphorylation in LMTK3 activation by CDK5 and its contribution in breast cancer progression

Cyclin dependent kinase 5 (CDK5) is a cytoplasmic proline-directed serine/threonine kinase which is commonly overexpressed in many tumours [112]. Recently, *in vitro*

studies have identified the ability of CDK5 to phosphorylate LMTK3 and results in breast cancer progression [10].

Phosphorylation is a common posttranslational modification of proteins in eukaryotic cells [113,114,115] that regulate several important cellular processes. Before a kinase to phosphorylate its substrate, it has to activate themselves first through phosphorylation by other kinases and that phosphorylation event occurs at kinases activation loop, which is part of the activation segment [11,12]. The activation loop phosphorylation is a major mechanism that induces the dynamic changes in activation process 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 [13-15]**.** The active form of kinase is able to phosphorylate its substrate. Therefore, identification of probable phosphorylation site in LMTK3 domain and intermolecular interactions between CDK5 and LMTK3 during phosphorylation are yet to be understood.

Scope of the present work

In **Chapter 7a**, we have carried out CDK5 and LMTK3 interaction by protein-protein docking study, to identify the probable interactions that may involve during the phosphorylation of LMTK3 by CDK5. We identified probable interface residues across CDK5-LMTK3 domain. CDK5 displayed non-bonding interactions with the probable phosphorylation sites of LMTK3 at activation loop. 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 tumour progression. In **Chapter 7b** we carried out *in silico* phosphorylation of LMTK3 at activation loop, and studied the effect of phosphorylation in the dynamic system.

2.3.7. LMTK3 inhibitors

It can be seen from literature, LMTK3 appears to be a promising new target against breast cancer progression and metastasis. A drug screening program is required to facilitate the identification of specific LMTK3 inhibitors, which could eventually aid to tackle breast cancer metastasis and endocrine resistance.

Scope of the work

In **Chapter 8** and **9** we made an attempt to design potential inhibitors against LMTK3 using ATP competitive and E-pharmacophore modelling approach.

2.4. Main objectives of this thesis

[1] Since the experimental 3-D structure of LMTK3 domain (133-411 amino acids) is not available, inhibition of LMTK3 activity becomes elusive. So we set our first objective to understand the characteristic structural features and dynamics of the Human LMTK3 domain.

[2] It is known that LMTK3 regulates ERα activity through phosphorylation which leads to breast cancer progression and endocrine therapy resistance, it is critical to understand the Protein-Protein interactions between ERα and LMTK3 at molecular level. This we have set as our second objective.

[3] Apart from ERα regulation, LMTK3 is able to interact with GRB2 and promotes breast cancer motility and invasion in triple negative breast cancer. So we set our third objective to understand the Protein-Protein interactions of GRB2 and LMTK3 at molecular level.

[4] It has been identified that LMTK3 can be activated through phosphorylation by CDK5 which promotes tumour progression in breast cancer. So we set our fourth objective to understand the transient protein-protein interactions that phosphorylates LMTK3 by CDK5 and also the effect of phosphorylation in LMTK3 using MD simulation.

[5] Since LMTK3 is considered as a novel therapeutic target, its activity can be modulated by potential inhibitors. So we set our fifth objective to identify high affinity potential inhibitors against LMTK3 using virtual screening and molecular docking. We used two different approaches to identify potential inhibitors against LMTK3: **(i)** ATP competitive approach **(ii)** E-pharmacophore modelling approach.