*Chapter - 2*

# LITERATURE REVIEW

## 2.1 CANCER

Cancer is undoubtedly one of the most studied and unsolved non communicable human diseases. Cancer is a defined as deregulated multiplication of cells with the consequence of an abnormal increase of the cell number in particular organs or tissues (http://www.cancer.gov/). Beginning stages of the developing cancer are restricted to the organ of origin while advanced stage of the cancers grow beyond the tissue of origin .In advanced stage cancerous cells were invade the connected adjacent tissues of the primary cancer (http://www.who.int/). At a later stage, these are distributed via the hematopoietic and lymphatic systems throughout the body where they can colonize in distant tissues and cause metastasis(Butterworth and Co., 1952). The development of cancers is thought to result from the damage or mutation of the cellular genome, either due to random endogenous mechanisms or caused by environmental influences. In the attempt to build how, why and when cancer occurs, ample of regulatory cellular pathways have been found in biological system that are essential to sustain common cellular functions like cell proliferation, cell differentiation and cell migration. (Chiang and Massagué., 2008; Poste and Fidler., 1980).Any alterations or unwanted changes in this fine-tuned regulatory network and interactions either due to endogenous failure or due to exogenous factors like environmental factors that may hinder the normal functioning of any member of such

regulatory pathways. This could stimulate the death of the affected nearby cell, may mark it for cancerous development within a particular tissue. Since body natural mechanisms always amend such environmentally induced abnormalities, apoptosis or programmed cell death is the best example which can control proliferation in cancer diseases (Kerr et al.,1972).

Cancer is one of the most life threatening disease condition around the world (American Cancer Society., 2008). A total of 14 million new cases and 8.2 million cancer associated death in 2012 according to WHO report (World cancer report. 2014). It is expected that global cancer burden will rise by approximately 70% in next two decades. Annually, nearly 500,000 people die of cancer in India. The WHO said this number is expected to rise to 700,000 by 2015. Any individual can affect by this disease at any stages of their life time. However the risk of being diagnosed with this disease is accounted for to be expanding with age (Panno., 2005). About 80% of all malignancies are diagnosed in patients age 55 and more (Panno., 2005). So far more than 100 types are cancers are reported (Pecorino., 2008). Among all types of cancer lung cancer is reported to be leading cancer killer around the world (Parkin et al., 2000). About 1.59 million deaths were reported due to lung cancer itself in year 2012. Lung, collateral and breast cancer are the three main leading cancers in women living in developed countries (Garcia et al., 2007). While in developing countries the main leading cancers are lung, stomach and liver in men, and breast, cervix and stomach in women (Garcia et al., 2007). Leukemia and lymphoma are reported to be the main leading cancer in children in most countries (Ross & olshan., 2004) According to WHO report, lung, oral, lip, throat and neck cancers are the most

common types of cancer among the Indian men while women are more suffering from cervix, breast and ovarian cancers.

### *2.1.1 Global Burden of Cancer*

Cancer is a leading cause of death worldwide, accounting for 8.2 million deaths in 2012. The most common causes of cancer death are cancers of:

- lung (1.59 million deaths)
- liver (745 000 deaths)
- stomach (723 000 deaths)
- colorectal (694 000 deaths)
- breast (521 000 deaths)
- oesophageal cancer (400 000 deaths)

The usual chemotherapy is one of the leading treatment option available to the cancer patients (Gibbs., 2000). Apart of chemotherapy surgery and radiation therapy are commonly used for the treatment of localized cancer (Remers., 2004) are two other most commonly used treatment options for localized tumours (Remers., 2004).The main problem for treatment of cancer is that the drugs which are effective against one type of cancer are reported to be generally not active against cancer of other tissues (Remers., 2004). A large portion of the chemotherapeutic medications are non-specific target just all dividing cells and leave numerous dangerous side-effects (Cao., 2008), causing very poor therapeutic index. The side effect as well as high cost associated with this disease makes it dangerous disease in world wide. Designing of selective targeting drugs are most promising area in cancer drug development (Cao., 2008). Hence an

increase of resistance towards currently available drugs is one of the leading obstacle in cancer treatment (Bennett et al., 2008;Raguz and Yague., 2008 ; Szakacs., 2006).

### *2.1.2 Cell Signaling Pathways And Cancer*

Generally there is few better treatment options for majority of cancers (Jain et al., 2007). There is also a huge variation in the world wide survival rate for different types of cancer. There is remarkable improvement in the survival rate of breast, colorectal and prostate cancer in developed countries in last five years (Coleman et al., 2008). Currently, however there are no suitable drugs available for some cancers like pancreatic cancer after failure of main line treatment (Li et al., 2004). Hence it is an urgent demand to discover new compounds that could be developed as anti-cancer drugs. The newer approach in cancer treatment is to target the altered cell survival pathways which are over expressed in cancer cells (Fabian et al., 2005). This approach is based on the assumption that targeting over-expressed pathways will be more selective in destroying the cancer cells and will have fewer effects on normal cells (Fabian et al., 2005). Various biological cell signalling pathways are altered in cancer.

Cellular signaling pathways are pathways that interconnected to form complex signaling networks. In this signaling pathway cells receive information from many different growth factor receptors leading to form cell-matrix and cell-cell contacts. After this they pass this information to nearby cell to regulate diverse important cellular processes, such as protein synthesis and cell growth, motility, cell architecture and polarity, differentiation, and programmed cell death (Wong and Chen. 2014). The

same signaling molecules are used to control different processes within different signaling complexes or at different intracellular locations (Martin., 2003). Moreover, cell signal pathways are subject to cellular regulation and generate completely different outcomes in several cell types; the activation of a signal molecule might have distinct consequences, betting on the cellular context. Understanding however these terribly complicated signal networks operate in vivo and the way they're altered in cancer cells represents a serious intellectual challenge (Martin., 2003).

Now a days the search for cancer drugs has basically focused on the molecular cell signaling pathways that go altered in cancer cells. These signaling pathways depends intensely over the action of 500 protein kinases whose dysregulation has been embroiled in cancer (Manning et al., 2002; Krause and Van., 2005). Thus, developing new drug molecules that target cellular signaling pathways has become an attractive endeavour for most of pharmaceutical companies and biotechnology industries now a days. For instance, the receptor tyrosine kinases EGFR and ErbB2 (Her-2) receptor were cloned in 1983 by analysts at Genentech, Inc., and monoclonal antibodies were accordingly created to target them. These monoclonal antibodies were at that point formed into the medications cetuximab (Erbitux, C-225, BMS; Merck) and trastuzumab (Herceptin; Genentech), individually. These medications were endorsed by the FDA in 2004 for the treatment of colon malignancy. The Bcr/Abl kinase inhibitor imatinib mesylate (Gleevec; Novartis) was endorsed in 2001 for the treatment of perpetual myelogenous leukemia (CML). With the exception of Gleevec, the other single-target medications created so far have poor safety as well as

efficacy profiles and turned out to be prohibitively costly. Protein kinases are main components all signal transduction pathway(Fabian et al., 2005). There are more than 500 protein kinases are in human genome, are considered as the second largest group of drug targets after GPCR (Manning et al., 2002; Akritopoulou-Zanze and Hajduk., 2009). Kinases basically catalyze the transfer of the terminal phosphate group of adenosine triphosphate (ATP) to the specific hydroxyl group of serine, threonine, or tyrosine residues in a protein substrate (Cherry and Williams., 2004) This phosphorylation acts as a key event in signal transduction and ultimately helps to transfer the signal from extracellular to intracellular environment (Halazy., 2003). The kinase mediated signal transduction plays a key role in a number of important cellular processes like cell growth, cell differentiation, maintaining cytoskeletal integrity and apoptosis (Gill., 2004).Hence aberrant kinase signaling either from an activating mutation or over expression was reported to play key role in many diseases like cancer, inflammation, diabetes neurodegeneration and psoriasis (Bogoyevitch and Fairlie., 2007; Hameed., 2009 ).In cancer cells, the kinase signaling pathways are often altered, resulting in uncontrolled growth and increased capability to invade surrounding tissue (Bianco et al., 2006; Dancey and Sausville., 2003). Among different kinase signaling pathways, altered or constitutively activated phosphoinositide 3-kinase (PI3K) pathway is reported in many human cancers(Zhang et al., 2009 ; Luo et al.,2003 ; Hennessy et al., 2005). Such aberrant PI3K pathway is implicated in tumor development and progression and also in tumor's response to cancer treatment (Vara et al., 2004). 3-Phosphoinositide-dependent protein kinase-1 (PDK-1) is a pivotal kinase for the PI3K pathway (Toker and Newton., 2000) and is an

attractive target for developing anticancer therapeutics (Fujita and Tsuruo.,2003).Agents targeting altered kinase pathways are currently being developed as the next generation anti-cancer agents. A representative sample of small molecule inhibitors targeting kinases are listed in Table 2.1

**Table 2.1 Small molecules Kinase inhibitors in clinical development (Adopted from Pearson and Fabbro., 2004)**

| Agent | Company | Target | Indication |
|---|---|---|---|
| Gefitinib | Astra Zeneca | EGFR | NSCLC |
| Eriotinib | OSI/Genentech/Roche | EGFR | NSCLC |
| Valatanib<br>PKT 747/2K 222584 | Novartis | VEGFR | Solid Tumour |
| SU11248 | Pfizer | VEGFR<br>C-Kit<br>Flt-3 | Solid Tumour<br>GISI<br>AMC |
| Imatinib | Novartis | ABL<br>PDGFR<br>C-Kit | CML<br>CMML/HES<br>GIST |
| BMS-35485 | Bristol/Myers Squibb | Src,AB | Solid tumours |
| Rapamycin<br>(CCI-779) | Wyeth | TOR | Solid tumours |
| SU011248 | SUGEN/Pfizer | PDGF-R<br>FIt-3, C-kit<br>VRGFR | GIST |
| RAD001 | Novartis | TOR | Solid tumours |
| Midostaurm/PKC412 | Novartis | FIt-3, C-kit<br>PDGFR | AMC<br>Solid tumours |
| PD0325901 | Pfizer | MEK | Solid tumours |
| BAY-43-9006 | Bayer/Onyx | BFAR<br>VEGFR1-3 | Malignant<br>Melanoma<br>Renal cell<br>carcinoma |

## 2.2 3-Phosphoinositide-Dependent Protein Kinase-1 (PDK-1)

The 3-phosphoinositide-dependent protein kinase-1 (PDK-1) was first discovered in the insulin signal transduction research (Alessi et al., 1997; Stokoe et al., 1997).PDK 1 kinase is main component of PI3/AKT pathway, is a 63 kDa Serine/Threonine class of kinase. It is ubiquitously expressed in many human tissues (Vanhaesebroeck and Alessi.,2000). It is a 556 amino acid containing enzyme (Alessi et al., 1997) with two important domains, an N-terminal kinase domain and C-terminal pleckstrin homology (PH) domain (Alessi et al., 1997). Pleckstrin homology (PH) domain is involved in the interaction with the phosphatidyl-inositol 3,4,5 triphosphate (PtdIns(3,4,5)P3) in the membrane while the kinase domain is involved in activation of many other downstream kinases (Currie et al., 1999). PDK-1 is reported as the master regulator as it activating 23 different kinases of the cAMP-dependent, cGMP-dependent, protein kinase C (AGC) kinase family (Mora et al., 2004; Komander et al., 2005).These include protein kinase B (PKB, also known as Akt),(Alessi et al., 1997; Brazil and Hemmings., 2001) p70 ribosomal S6 kinase (S6K), (Pullen et al., 1998) serum- and glucocorticoid-induced protein kinase (SGK) (Perrotti et al., 2001) and protein kinase C (PKC) isoforms (Newton., 2003; Belham and Avruch., 1999;). Several High resolution X-ray crystal structures are available for PDK-1 binding with its inhibitors in the protein data bank (PDB) (Biondi et al., 2002; Komande et al., 2004; Komander et al., 2003; Feldman et al., 2005). The catalytic core of PDK-1 has a bilobal domain, with predominantly α-helical C-terminal lobe and an N-terminal lobe consisting mainly of β-sheets (Biondi et al., 2002). The adenosine triphosphate (ATP) binding region is positioned in the hinge region

between the N- and C-terminal lobes (Biondi et al., 2002). Similar to other kinases, the ATP binding site in PDK-1 can be divided into three regions, adenine region, sugar region, and phosphate region (Vulpetti and Bosotti., 2004). PDK-1 itself falls to the same AGC family enzyme as its substrates so it needs to be phosphorylated at the activation loop or T-loop (Mora et al., 2003). PDK-1 is reported to be constitutively active and has an intrinsic ability to phosphorylate its own T-loop at Ser241 residue (Casamayor et al., 1999). This autophosphorylation of T-loop of PDK-1 is mediated by intermolecular reaction (Wick et al., 2003). The presence of activation loop is a common feature in all kinase superfamily (Ballif et al., 2001). In PDK-1, kinase domain has hydrophobic motif (HM) pocket or PDK-1 interacting fragment (PIF) pocket (Biondi et al., 2002).It was reported that activation of AGC kinases were mediated by phosphorylation in T-loop region and hydrophobic motif region (Vanhaesebroeck and Alessi., 2000). PDK-1 does not have a separate hydrophobic motif like other AGC kinases (Mora et al., 2004).

PDK-1 function is reported to be regulated by two factors, substrate conformation and sub-cellular location (Toker and Newton., 2000).The detailed PDK-1 signaling pathway is given in Figure 2.2. First phase in PDK-1 signaling pathway is activation of upstream PI3K leads to the synthesis of phosphatidyl-inositol 3,4,5 triphosphate [PtdIns(3,4,5)P3]/ phosphatidyl-inositol 3,4 diphosphate [PtdIns(3,4)P2] at the plasma membrane (Vanhaesebroeck and Alessi.,2000). This activation results in PKB's and PDK-1's PH domain's interaction with PtdIns(3,4,5)P3/PtdIns(3,4)P2 and thus causes translocation to membrane and co-localization (Currie et al., 1999).This results autoinhibition of PKB and PDK-1 phosphorylates PKB on Thr308 (Toker and Newton.,

2000).Substrates other than PKB was reported to interact with PIF pocket of PDK-1 using their phosphorylated hydrophobic motif and then subsequently gets activated by PDK-1 (Biondi et al., 2004). Hence it was concluded that PDK-1 acts as the sensor of substrate conformation and is regulated by substrate conformation.

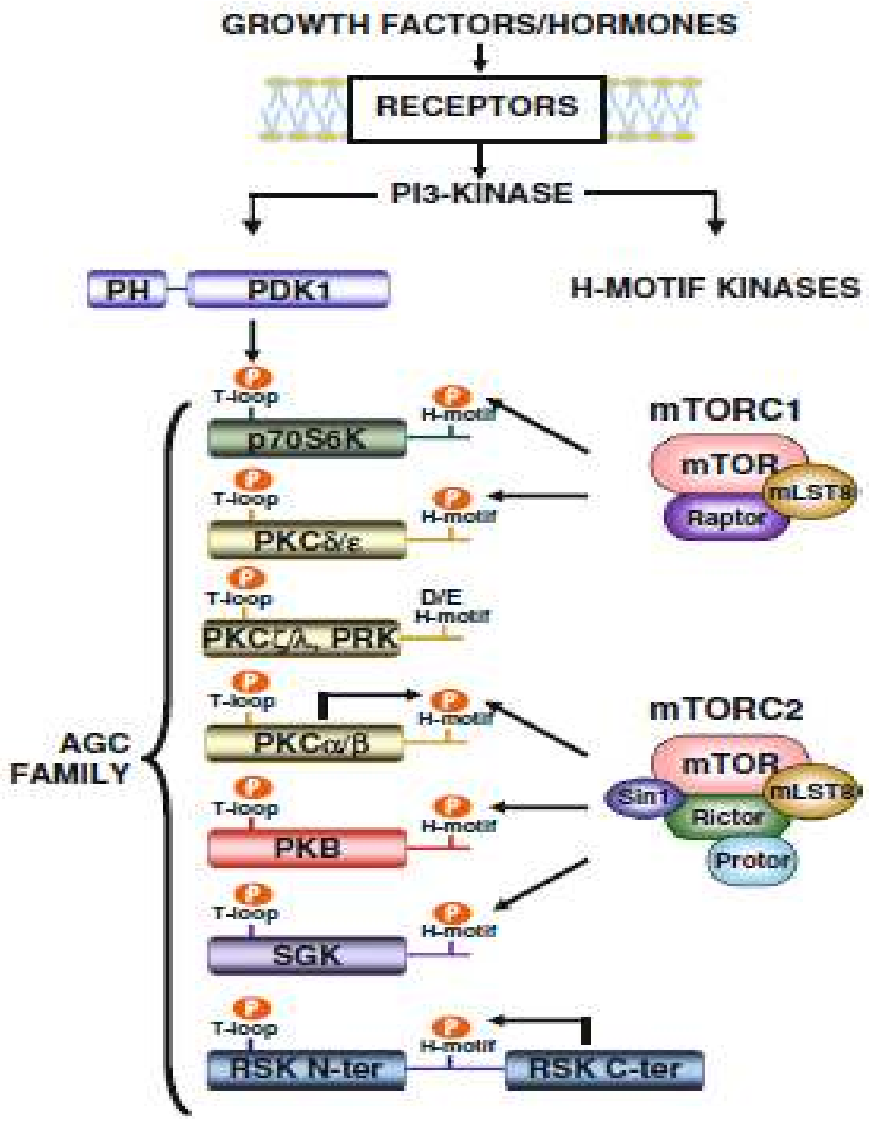


**Figure 2.1: Cell signaling pathway of PDK-1 (Adopted from Bayascas., 2010)**

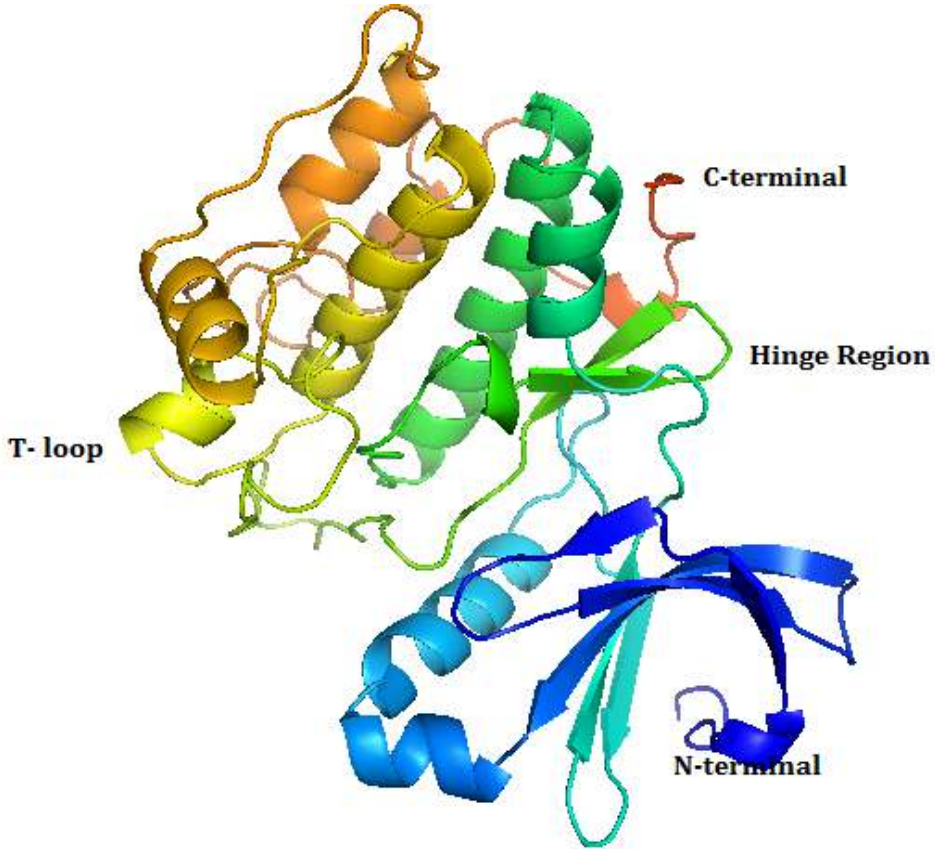


**Figure 2.2: Structure of PDK-1 kinase domain (PDB ID: 2BIY) (Komander et al., 2005)**

### 2.2.1 *Targeting PDK-1 In Cancer*

It was reported that the tumor suppressor, Phosphatase and Tensin homologue deleted on chromosome Ten (PTEN), acts as the phosphatidylinositol 3-phosphatase (Leslie and Downes., 2002) and thus down regulates the PDK-1 mediated growth and signaling pathway. Mutations in PTEN resulting in elevated levels of PtdIns(3,4,5)P3 was reported in many human cancers and in such cases inhibition of PDK-1 is expected to mimic the tumor suppressing activity of PTEN (Gao et al., 2005;) .Moreover, most of the downstream proteins of PDK-1 are also implicated in cancers (Toker and Newton., 2000; Kim and Chung., 2002;

Sahoo et al., 2005;Gao et al., 2004; Smith et al., 2005).As PDK-1 is the master regulator of such AGC kinases, it is in a unique position to control all these proteins. Accumulating pharmacologic and genetic evidences supports the potential role of PDK-1 as anticancer target (2005; Flynn et al., 2000; Zeng et al., 2002).It has been reported that over-expression of PDK-1 in mammary cells resulted in their transformation *in vitro* and tumor formation *in vivo* (Zeng et al., 2002). Elevated levels of PDK-1 phosphorylation was also reported in metastasized breast tumors (Lin et al., 2005). Studies also showed that PDK-1 phosphorylation is ovarian carcinomas (Ahmed et al., 2008). PDK-1/AKT pathway is also reported to be activated in Rhabdomyosarcoma (RMS) (Cen et al., 2007). Targeting PDK-1 with anti-sense oligonucleotides has showed a marked reduction of cell proliferation and survival and also an increased rate of apoptosis than that observed in PI3K or PKB inhibition (Flynn et al., 2000). PDK-1 was also reported as a potential target for sensitizing breast cancer cells to chemotherapeutic agents (Liang et al.,2006). Knock down of PDK-1 was recently reported to enhance the anti-tumor effect of EGFR inhibitor (Zhang et al., 2006). These studies also show that PDK-1 inhibitors will help to improve the clinical response to EGFR inhibitors. PDK-1 is also reported to play a role in the motility of cancer cells (Pinner and Sahai., 2008). Lack of PDK-1 is reported to cause inhibition of cell proliferation in mouse embryonic fibroblasts (MEFs) (Nakamura et al., 2008). A recent study shows that PDK-1 mediates cell survival through another distinct IκB kinase-β (IKKB)/NFkB pathway in addition to AKT pathway (Tanaka et al., 2005).PDK-1-hypomorphic mice which express only 10% of normal levels of PDK-1 were reported to be viable and fertile(Lawlor et al., 2002). This finding shows that inhibition of

PDK-1 can be achieved without severe toxicity. A more recent study using PDK-1-hypomorphic mice has shown that the reduced PDK-1 expression in PTEN+/- mice markedly protected the animal from a wide range of tumors (Bayascas et al., 2005).Thus PDK-1 has become a well validated anticancer target. ATP competitive PDK-1 inhibitors will compete with ATP molecule to bind in the PDK-1 active site and prevents the transfer of phosphate group from ATP to downstream substrate proteins .This will result in the blockade of PDK-1 mediated signal transduction. Development of PDK-1 inhibitor could lead to development of better treatment options for cancer.

**2.3 Flavonoids**

Plants are rich sources of chemically diverse class of compounds, many with valuable benefits to human health. Subsequently, nearly 50% of the anticancer therapeutic agents are derived from plants (Ross and Kasum., 2002). Polyphenolic compounds are one of the most abundant class of compounds derived from plant source (Beecher., 2003). They are basically the secondary metabolites (Nijeveldt et al., 2001) performed many important functions in plants. Some of these functions are protection from UV light , tolerance of biotic and abiotic stresses, allelopathic interactions , pigmentation, and normal cell growth and development (Gould and Lister., 2005; 2011; Hassan and Mathesius., 2012; Samanta et al.,). Polyphonels are further divided into 10 general classes containing more than 8,000 compounds identified to date (Harborne and Williams 2000; Ververidis et al. 2007; Chahar et al. 2011). Among these polyphenols flavonoids are most abundant class of polyphenols, accounting for nearly 60% of the total polyphenols

(Harborne 1994). Flavonoids are further divided into six sub-classes based on their chemical structure, including flavanols, flavones, flavanones, flavonols, isoflavones, and anthocyanidins. The basic chemical structure of flavonoids is characterized by a presence of diphenylpropane carbon skeleton of C6-C3-C6, with two benzene rings are fused by a 3-carbon chain forming a heterocyclic pyran or pyrone ring with an oxygen (Bazzano et al.2002; Atmani et al. 2009; Xiao et al. 2011;Fang et al. 2010; Clifford 2004;). Differences in the number and arrangement of the hydroxyl groups lead to variations within each group.

Among flavonoids, especially flavanols, flavonols and anthocyanins are relatively most abundant in human diet and probably more involved in prevention of cancers, cardiovascular diseases and neurodegeneration (Xiao et al., 2011; Bazzano et al., 2002; Atmani et al., 2009; Clifford ., 2004; Fang et al., 2010)

**2.4 Minimization**

The Minimizer uses an algorithm to identify the geometrics of the molecule corresponding to the minimum points on the potential surface energy (Fox and Stanton., 1968). The minimizer reduces the unwanted forces which are present in the molecule and lowers the energy level of the molecule. There are many algorithms available in the minimization process. Some of the minimization methods used in the Smart Minimizer is Steepest Descent method, Conjugate Gradient method, Newton Raphson method and Quashi-Newton method (Curry., 1944; Hestenes and Stiefel., 1952; Payne *et al*., 1992; Ypma., 1995).

## 2.5 Molecular Docking:

Molecular docking is the technique used to study molecular binding that aims in placing the ligand into the active site of protein to ensure proper ligand-protein interactions (Lengauer and Rarey., 1996). The term **"Docking"** is mostly related to protein-ligand interactions . Knowledge of preferred orientation in turn may be used to predict strength association and binding energy between two molecules by using scoring functions. Molecular docking may be defined as an optimization problem, which would describe the "best-fit" orientation of a ligand that binds to a particular protein of interest (Fletcher., 1969). Docking is useful for predicting both strength and type of signal produced.

The focus of molecular docking is to computationally stimulate the molecular recognition process. The aim of molecular docking to achieve an optimized conformation for both the protein and ligand and relative orientation between protein and ligand such that the free energy of overall system is minimized. Docking is frequently used to predict the binding orientation of small molecule drug candidates to their protein target in order to in turn predict the affinity and activity of small molecule. Hence molecular docking plays an important role in the rational design of small molecule drugs. On the basis of given the biological and pharmaceutical significance of molecular docking, significant efforts have been directed towards improving the methods used to predict docking.

There are two popular docking approaches exist. The conformational search approach uses a matching technique that describes the protein and ligand as complementary surfaces. The second approach using Scoring methods simulates the actual docking process in which the ligand- protein

pair wise interaction energies are calculated. The energies are three types: Force field based, Empirical based and knowledge based methods. Both approaches have their own significant advantages as well as some limitations too.

Scoring is the process of evaluating a particular pose (binding mode) by counting the number of favorable intermolecular interactions such as hydrogen bonds and hydrophobic contacts.

There are several docking methods which are used to dock ligands in different docking algorithms. Each method has its own advantages as well as disadvantages. Here I used two docking methods for docking studies implemented in Discovery Studio By Accelrys are LigandFit (Venkatachalam, 2003) and CDOCKER (Wu et al., 1997).

### *2.5.1 C-DOCKER:*

C docker (Wu et.al; 2003) is a gird based molecular docking method that employs CHARMm forcefield (Brooks et al., 1983). During the refinement the receptor is held rigid whereas the ligands are flexible. Random ligand conformations are generated from the initial ligand structure through high temperature molecular dynamics followed by random rotations. Scoring is given by evaluation of individual poses. The random conformations are refined by gird based simulated annealing. Many force field based methods are based on the following simple relationship:

$$E_{binding} = E_{complex} - (E_{receptor} + E_{ligand})$$

The binding energy is calculated by removing the internal energy of individual components.

In CDOCKER the receptor is held rigid while the ligands are allowed to be flexible during refinement process. Random conformations of ligands are generated from initial structure of ligand through high temperature molecular dynamics followed by random rotations. To adequately explore the conformation space, many different optimization methods and search strategies have been developed, including distance-geometry, Monte Carlo (MC) simulated - annealing, genetic algorithm (GAs), and molecular dynamics. Further the generated random conformations are refined by grid based simulated annealing and a final grid based or full force field minimization. Soft core potentials are found to be effective in exploring the conformational space of small organic and macromolecules and are being used in various applications, including docking and the prediction of protein loop conformations. During the docking process, the non bonded interactions Vander Wall (vdW) and electrostatics) are softened at different levels, but this softening is removed for the final minimization.

#### 2.5.1.1 *CDOCKER Protocol*

The standard protocol of DCOKER docking method, 50 replica of each ligand is generated and randomly distributed around the centre of active sites. The internal coordinates of each replicas are kept same as those originally generated from CORINA (used to 2D structure of molecules). MD simulated annealing process is performed using rigid protein and flexible ligand. The ligand protein interactions are calculated either GRIDI or GRID II or complete force field. Final minimization step is performed on the each of ligand docking pose. Minimization consists of 50 steps of steepest descent followed by 200 steps of conjugate gradient

using an energy tolerance of 0.0001 kcal/mol. The minimized docking poses are then clustered based on a heavy atom RMSD approach using 1.5 A tolerance. The final ranking of the ligand's docking pose is based on the total docking energy (Intra molecular energy of the ligand's for ligands and the ligand -protein interactions).

**Steps:**

- Minimization of the receptor was done using CHARMm forcefield.
- An active site sphere was defined for the receptor.
- The ligand was given as an input to run the protocol.
- C docker energy for all the conformations were obtained in the outfile.

### *2.5.2 LIGAND FIT:*

Ligand Fit is a shape-based method for accurately docking the ligands into protein active sites (Morris *et al.*, 1998).The method employs the well known cavity detection algorithm for finding and invaginations in the protein for candidate active regions. A shape comparison filter is combined with a Monte Carlo conformational search for generating ligand poses consistent with the active site shape (Li and Scheraga., 1987). Candidate poses are further minimized in the context of the active site by using a gird-based method for evaluating protein-ligand interaction energies. Further errors arising from grid interpolation are dramatically reduced using a new non-linear interpolation scheme.

#### *2.5.2.1 Site search*

Sites are defined based on the shape of the protein. An “eraser” algorithm (Savage., 1997) is used to clean all the grid points outside the protein. The boundary between inside and outside is determined by defining the

opening size parameter. Within the boundary of flood filling algorithm is employed to search unoccupied grid points which form the cavities (sites). All detected sites can be browsed according to their size, and size less than user defined size cut-off eliminates.

#### *2.5.2.2 Conformational search*

The Monte Carlo simulation is employed in the conformational search of the ligand (Li and Scheraga., 1987) . During the search, bond lengths and bond angles are untouched only torsional angles (except those in a ring) are randomized. Therefore, the ligand molecules should be energy minimized to ensure correct bond lengths and bond angles before using ligand fit.

#### *2.5.2.3 Ligand fitting*

After a new conformer is generated, the ligand fitting is carried out in two steps. First the non mass-weighted principle moment of inertia (PMI) of the binding site is compared with non mass-weighted principle moment of inertia (PMI) of the ligand. If the value ($Fit_{value}$) is above the threshold or not better fitting results previously saved, no further docking process will be performed. If the value ($Fit_{value}$) is better than previously saved results the ligand is positioned into the binding site according to the PMI. Because PMI is a scalar property, there four possible positions for the ligand to orient in the binding site. For each position, the corresponding docking score is computed.

The docking score is negative value of the non-bonded inter molecular energy between ligand and protein. After the docking score is calculated, for each orientation it is compared with the results saved previously. If the new one is better, it is saved, and then the process of conformational

search and ligand fitting is iterated until number of trials is reached. Finally rigid body minimization is applied to the saved conformations of the ligand to optimize their positions and docking scores.

## 2.6 Pharmacophore

First time the concept of pharmacophore was introduced in 1909 by Ehrlich, who defined the pharmacophore as ‘a molecular framework that carries (phoros) the essential features responsible for a drug’s (pharmacon) biological activity’. This concept remain unchanged after a century's development, but international meaning and application changed continuously. According to recent definition of IUPAC (Wermuth et al., 1998), 'a pharmacophore is the ensemble of steric and electronic features that is necessary to ensure the optimal supramolecular interactions with a specific biological target and to trigger (or block) its biological response (Gund., 2000; Wermuth., 2006;Dror et al., 2006). A suitable pharmacophore model enables understanding the interaction between a receptor and ligand. A pharmacophore model or hypothesis consists of a three dimensional configuration of chemical functions surrounded by tolerance sphere (Barnum, et al., 1996). A tolerance sphere defines that area in space that should be occupied by specific type of chemical functionality (Barnum, et al., 1996). Pharmacophore models are routinely used in lead identification and optimization in the areas of library focusing, evaluation and prioritization of virtual high throughput screening (VHTS) results, denovo design and scaffold hopping (Schneider et al., 1999).

### *2.6.1 List of features available for pharmacophore generation*

#### *2.6.1.1 HB ACCEPTOR (vector)*

It matches sp or $sp^2$ nitrogen's that have a lone pair and charge less than or equal to zero with surface accessibility. $sp^3$ oxygen's or sulfur's that have a lone pair and charge less than or equal to zero and non-basic amines that have a lone pair. It does not match to basic, primary, secondary and tertiary amines that are protonated at physiological pH. There is no exclusion of electron-deficient pyridines and imidazoles.

#### *2.6.1.2 HB ACCEPTOR lipid (vector)*

It matches nitrogen, oxygen or sulfur atoms that have a lone pair and charge less than or equal to zero. It includes the basic nitrogen. There is no exclusion of electron-deficient pyridines and imidazoles.

#### *2.6.1.3 HB DONOR (vector)*

It matches Thiols, non acidic hydroxyls, Acetylenic hydrogen's and NHs. It does not match electron-rich pyridines and imidazoles that would be protonated or nitrogen's that would be protonated due to their high basicity.

#### *2.6.1.4 HYDROPHOBIC (scalar)*

It matches a contiguous set of atoms that is not adjacent to any concentrations of charge (charged atoms or electronegative atoms) in a conformer such that the atoms have surface accessibility to groups like cycloalkyl, isopropyl and methyl.

#### *2.6.1.5 HYDROPHOBIC ALIPHATIC (scalar)*

It matches a contiguous set of atoms that is not adjacent to any concentration of charge (charged atoms or electronegative atoms) in a conformer such that the atoms have surface accessibility to groups like cycloalkyl, isopropyl and methyl.

#### *2.6.1.6 HYDROPHOBIC AROMATIC (scalar)*

It matches a contiguous set of atoms that is not adjacent to any concentrations of charge (charged atoms or electronegative atoms) in a conformer such that the atoms have surface accessibility to groups like phenyl and indole.

#### *2.6.1.7 NEG CHARGE (atom)*

Matches negative charges not adjacent to a positive charge.

#### *2.6.1.8 NEG IONIZABLE (point)*

Matches atoms or groups of atoms that are likely to be deprotonated at physiological pH, such as Triflouromethyl sulfonamide hydrogens, sulfonic acids (centroid of three oxygens), phosphoric acids (centroid of the three oxygens), sulfinic, carboxylic or phospinic acids (centroid of the two oxygens) and tetrazoles.

#### *2.6.1.9 POS CHARGE (atom)*

Matches positive charges not adjacent to a negative charge.

#### *2.6.1.10 POS IONIZABLE (point)*

Matches atoms or groups of atoms that are likely to be protonated at physiological pH. Such as Basic amines, Basic secondary amines (iminyl nitrogen), Basic primary amidines, except guanidines (centroid of the two

nitrogens), Basic guanidines (centroid of the three nitrogens). Positive charges adjacent to a negative charge do not match weakly basic aromatic nitrogen's such as pyridine and imidazole.

#### *2.6.1.11 RING AROMATIC (vector and plane)*

Matches 5- and 6-membered aromatic rings. The feature defines 2 points, the ring centroid and projected point normal to the ring plane. The projected point can map both above and below the ring.

#### *2.6.1.12 Principal*

It specifies the reference molecules and reference configuration models are considered to be potential centers for hypotheses.

- 0-Indicates do not consider this molecule
- 1-Consider configurations of this molecule
- 2-Use this molecule as a reference molecule

#### *2.6.1.13 MaxOmitFeat*

- 0-All features must map to generate hypotheses
- 1-All but one feature must map to generate hypotheses
- 2-No feature need to map to generated hypothesis

**Steps:**

- Using common feature hypothesis, training set was given as an input.
- Out of 11 catalyst features only 10 were selected to run the protocol.
- The best common feature pharmacophore model with maximum fit value and rank file was obtained from exporting hypotheses.

### *2.6.2 Hypogen*

Hypogen attempts to derive SAR models for a set of molecules for which activity value ($K_i$) on a given biological target are available (Li et al., 2000). Hypogen optimizes hypothesis that are present in the highly active compounds in the training set. But missing among the least active (or inactive) ones. It attempts to construct the simplest hypothesis that best correlates that activity (estimates Vs. measured) the predicted models are created the predicted models are created in three stages namely:

- Constructive
- Substractive
- Optimization

The constructive phase identifies hypothesis that are common to the most active set of compounds. The most active set is determined by the following equation of the compounds. The most active set is determined by the following equation:

$$\mathbf{MA \times Unc_A = (A/Unc_A) > 0.0}$$

Where Unc is the uncertainty in the measured activity and A is the activity of the compounds.

The objective of the substractive phase is to identify those pharmacophore configurations. The optimization phase involves improvement of the hypothesis score. Small perturbations are applied to those pharmacophore configurations that survived the subtractive phase and that are scored based on the errors in activity estimates from regression and complexity of the hypothesis.

#### *2.6.2.1 Hypogen training set conditions*

- At least 16 compounds are necessary to assure statistical power.
- Activities should span 4 orders of magnitude.
- Each order of magnitude should be represented by at least 3 compounds.
- No reduction information.
- No excluded volume problems.

#### *2.6.2.2 Hypothesis considerations*

- Configuration values should be around 17.
- RMS should be as low as possible, preference nearer to zero.
- Correlation should be around 1.0.
- Cost factor difference between fixed cost and null cost should be between 40-80 bits.

The factors that determine the quality of Pharmacophore include Weight cost, error cost, configuration cost, null hypothesis and correlation. An uncertainty value of 2 or 3 is given for all the molecules in the training set.

### 2.7 Molecular Dynamic Simulation

In 1950s Molecular dynamics methods were used in the theoretical physics community. Later in 1957, Alder and Wainwright performed the first MD simulation by using the hard-sphere model. During the 1970s, as computers became more widespread, MD simulations were developed for more complex systems; culminating in 1976 with the first simulation of a protein (McCammon., 1976; McCammon et al., 1977) using an empirical

energy function constructed using physics-based first-principles assumptions.

Molecular dynamics (MD) simulation is a computational method that calculates the time dependent behaviour of a biological system (Cheatham and Kollman., 2000; . Karplus and McCammon., 2002) . MD simulations generate complete information about variation and conformational changes of proteins and nucleic acids in biological system, and now routinely used to investigate the structure, dynamics and thermodynamics of biological molecules and their complexes (Norberg and Nilsson., 2002; Moraitakis et al., 2003). The basic idea of molecular dynamics (MD) is to study atomic fluctuations in solvated system. The basic principle behind this is application of classical Newton's equation. Simulations reported in this thesis were performed with GROMACS molecular simulation package (Berendsen *et al.,* 1995). Before MD, the ionization states of amino acids were set to mimic a neutral PH environment i.e. all Lys and Arg carried net positive charge, and all glutamic acid and Aspartic acid carried a net negative charges. The histidine residues were in doubly protonated condition.

A random generation of 100 models from the starting structure was calculated and subsequently the best model with the low RMS value of superposition using Swiss-pdb viewer (Guex *et al*., 1999) was subjected for further analysis. The best modeled protein were solvated with water molecules in a truncated octahydron box. The size of the box was set to 1.1 nm distance from the surface of the protein. The Single Point Charge (SPC) water model (Berendsen *et al.,* 1987) and ions (Na+ and Cl −) was built. The box model, first with explicit water and then with ions was

added to protein containing truncated octahydron box, this was submitted to 500 steps of energy minimization using the steepest descent algorithm till an energy gradient was reached and it was found to be the most appropriate energy gradient to relax the models and afford well Ramachandran plots. In order to constrain all bond length in protein, the LINCS (Hess *et al.,* 1997) algorithm was used. For water molecule bond length constrain, the SETTLE algorithm was implemented (Miyamoto and Kollman, 1992). The electrostatic and Van Der Waal forces are implemented using particle mesh Ewald potential method (Essmann *et al.,* 1995) and Lennard-Jones potential method respectively. All full MD simulations were performed at 10ns with no restriction using two fs of integration time, constant temperature and pressure. The temperatures of the proteins and solvent molecules were each coupled separately, using (V-resacle) Berendsen thermostat algorithm (Berendsen *et al.,* 1984). The pressure was coupled using (Parrinello-Rahman) Berendsen algorithm at 1 bar with coupling constant ôP = 1 ps. Co-ordinates and energy terms (total, kinetic and potential for the whole system and electrostatic, distance dependent, distance-independent reaction force field) were saved for each ps. With the aim of evaluating the system stabilization throughout the molecular dynamics time, the total, kinetic and potential energy was plotted versus time. The stabilization was assessed by graphics visualization using VMD (Humphrey*et al*., 1996) and Xmgrace.

### *2.7.1 Performance analysis*

After successful completion of molecular dynamic simulation set up and force field parameterization, MD simulations are to be implemented. In present study, Leap-frog version of Verlet algorithm was exploited for

generating time averaged structural conformations with respect to forces that act on individual atoms (van Gunsteren and Berendsen, 1988). Leap-frog version of Verlet algorithm (Verlet., 1967) uses Newton's second law of motion.

$$\mathbf{F}_i = \frac{\partial(U(\mathbf{r}_1 \ldots \mathbf{r}_N)}{\partial \mathbf{r}_i}$$

Newton's law of motion is thus used in MD simulations to calculate the forces up on successive configurations of the system. The velocities, constant temperatures and pressure were constantly maintained. The MD (Molecular Dynamic) simulation algorithm used in present work is implemented in GROMACS package. The system temperature and pressure were kept constant throughout the MD simulation period.

### *2.7.2 Analysis of result*

MD simulations produce bunch of structural conformations at different time scale. Therefore, well planned analysis of bunch of structural conformations can provide vital clues of molecular function exactly. From all simulations generated from starting experimental model the Root Mean Square Fluctuation (RMSF), Root mean square deviation (RMSD), potential, kinetic and total energies were analyzed. The stabilities of intramolecular hydrophobic interactions were evaluated in terms of Lennard-Jones potential. Lennard-Jones potential is a good approximation of Van Der Waal (VDW) stabilization energies.

## 2.8 ADME Analysis

Absorption, distribution, metabolism, elimination and toxicity (ADME/Tox) are important in determining approval of the drug molecule. Half of the molecule fails because of its unacceptable efficacy, which may due to poor pharmacokinetic properties, poor absorption and poor safety profiles (Li, 2001). A reliable procedure to predict ADME properties is Lipinski Rule of Five formulated by Lipinski (Lipinski et al., 2001).

### *2.8.1Lipinski rule of five*

For qualifying compound, the compound should have molecular weight less than500, number of hydrogen bond donors is less than 5, number of hydrogen bond acceptor should be below 10 and octanol-water partition coefficient should be below 5 (Choy and Prausnitz, 2011).Among all the designed leads the molecules of high ranking which follow Lipinski's rule were selected and further analysed for binding with the protein model using docking tools.

### *2.8.2 LogP (octanol/water partition coefficient)*

LogP is calculated by the methodology developed by Molinspiration as a sum of fragment-based contributions and correction factors. Method is very robust and is able to process practically all organic, and most organometallic molecules.

### *2.8.3 Octanol-water partition coefficient logP*

LogP is used in QSAR studies and rational drug design as a measure of molecular hydrophobicity. Hydrophobicity affects drug absorption, bioavailability, hydrophobic drug-receptor interactions, metabolism of molecules, as well as their toxicity. LogP has become also a key parameter in studies of the environmental fate of chemicals.

### *2.8.4 Molecular Polar Surface Area TPSA*

It is calculated by based on the methodology published by Ertl *et al.* as a sum of fragment contributions. O and N atoms cantered polar fragments are considered. PSA has been shown to be a very good descriptor characterizing drug absorption, including intestinal absorption and bioavailability. It is a very useful parameter for prediction of drug transport properties. Polar surface area is defined as a sum of surfaces of polar atoms (usually oxygen, nitrogen and attached hydrogen atoms) in a molecule. This parameter has been shown to correlate very well with the human intestinal absorption, Caco-2 monolayer permeability, and blood-brain barrier penetration. The calculation of PSA in a classical way, however, is rather time consuming, because of the necessity to generate a reasonable 3D molecular geometry and determine the surface itself. Additionally, calculations require specialized software to generate the 3D molecular structures and to determine the surface. In today's era of drug development shaped by high-throughput screening and combinatorial chemistry, fast bioavailability screening of virtual libraries consisting of

hundreds of thousands, even millions of molecules is required. That is the reason why in our molecular property prediction toolkit so called topological polar surface area - TPSA is implemented. Briefly, the procedure is based on the summation of tabulated surface contributions of polar fragments (atoms regarding also their environment).

### *2.8.5 ADMET-Absorption*

This model predicts human intestinal absorption (HIA) after oral administration. Intestinal absorption is defined as a percentage absorbed rather than as a ratio of concentrations (cf. blood-brain penetration). A well-absorbed compound is one that is absorbed at least 90% into the bloodstream in humans (Waterbeemd et al., 2001).The intestinal absorption model includes 95% and 99% confidence ellipses in the ADMET_PSA_2D, ADMET_AlogP98 plane. The ellipses define regions where well-absorbed compounds are expected to be found: 95% of well-absorbed compounds are expected to fall within the 95% ellipse, while 99% of well-absorbed compounds should fall within the 99% ellipse. Note that the location of any particular compound does not necessarily imply whether it will be well, moderately, or poorly absorbed. In general, however, absorption tends to drop off quite rapidly outside the 95% ellipse. There are four prediction levels:

- 0 (Good)
- 1 (Moderate)
- 2 (Poor)
- 3 (Very Poor)

### *2.8.6 Protein Binding module*

It predicts plasma protein bound fraction and the equilibrium binding constant to blood serum albumin of a compound in blood. The protein binding properties are predicted from automatically calculated physicochemical properties such as lipophilicity, ionization constants, and hydrogen bonding capacity (Waterbeemd and Gifford., 2003)

### *2.8.7 Volume of Distribution module*

Contains a predictive model which generates a quantitative estimate of the apparent volume of distribution of a compound. Physicochemical parameters, charge state, lipophilicity and hydrogen bonding capacity are automatically calculated and used as inputs to the predictive model of the volume of distribution (Waterbeemd and Gifford., 2003).