Chapter 2

LITERATURE RIVIEW

2.1 Molecular targets of dietary polyphenols for prevention and therapy of cancer

There are numerous proteins which may be designated as target for cancer prevention and therapeutics as evidenced by recent research on cancer chemoprevention:

2.1.1 B-cell lymphoma 2 (Bcl-2)

Apoptosis (programmed cell death) is a molecular process by which cells commit suicide after completing physiological function, or after a severe genetic damage. A common apoptotic mechanism appears to be preserved throughout evolution and regulated mainly by the Bcl-2 family of proteins (Reed, 2002; Igney and Krammer, 2002). Some of them, such as Bcl-2, Bcl-xL, Bcl-w or Mcl-1 block apoptosis while others, such as Bad, Bak, Bax, Bid, Bim or Hrk, induce it. It has been established that overexpression of Bcl-2 and Bcl-xL proteins is related to the initiation and development of different types of cancer, as well as to resistance to chemotherapeutic treatments (Pintoa et al., 2004). Bcl-2 family proteins are characterized by containing up to four conserved sequences of amino acids which are known as Bcl-2 homology (BH) domains (Adams and Cory, 2001; Opferman and Korsmeyer, 2004). They are usually grouped into three distinct subclasses: (1) Bax and Bak (contain the BH1 to BH3 domains) that mediate apoptosis by triggering destabilization of the outer mitochondrial membrane and releasing the cytochrome c from mitochondria to the

cytosol (Green and Kroemer, 2004; Antignani and Youle, 2006), (2) another subclass is composed of the BH3-only proteins (including Bim, Bad, Puma and Noxa) that communicate pro-death signals and ultimately activate downstream Bax and Bak (Kim et al., 2006; Willis et al., 2007), (3) the remaining subclass composed of Bcl-2, Bcl-xL, Bcl-w, Mcl-1, A1 and Bcl-B, all of which contain all the four BH domains (BH1 to BH4 domains), suppress the activation of Bax/Bak (Sattler et al., 1997; Petros et al., 2000; Liu et al., 2003; Ku et al., 2011). Bcl-2 prevents the Bax activation/oligomerization by forming heterodiner (Bcl-2-Bax). Several mutagenesis studies showed that the BH1–3 regions of Bcl-2 and the BH3 region of Bax are critically required for their interaction (Yin et al., 1994; Zha et al., 1996; Wang et al., 1998; Lin et al., 2004; Dlugosz et al., 2006). In all the complex structures determined so far, the BH3 peptide forms an amphipathic α-helix whose hydrophobic surface interacts with a hydrophobic groove on the Bcl-2-like protein formed by the BH1–3 regions (Suzuki et al., 2000; Ding et al., 2010). The structures of antiapoptotic protein (Bcl-2, Bcl-xL, and Mcl-1) consist primarily of two central hydrophobic helices surrounded by amphipathic helices (Petros et al., 2004; Day et al., 2005). The binding groove is formed mainly by the α2, α3, α4, and α5 helices (Acoca et al., 2011). This protein-protein interaction might be an important target for cancer chemoprevention.

2.1.2 Heat shock protein-90 (Hsp90)

Molecular chaperones, which perform correct folding of a number of proteins inside the cell, found to be play a significant role in cancer proliferation. Hsp90 (heat shock protein 90) was first characterized with regard to role of molecular chaperones in cancer (Whitesell and Lindquist, 2005). Hsp90 stabilizes a wide range of mutated and over expressed oncogenic proteins (Pearl, 2005; Calderwood et al., 2006). Hsp90 expression is constitutive in tumor cells at 2–10-fold higher levels compared to normal cells (Ferrarini et al., 1992; Welch and Feramisco, 1982). More than 200 proteins molecules folded by Hsp90 in their stabilized and active conformation state (Kamal et al., 2004). Protein kinases are known to be involved in cell proliferation and may cause cancer when deregulated. Protein kinases depend on Hsp90 for proper folding to be functional. Hsp90 requires a series of co-chaperones to form functional super-chaperone complex. These co-chaperones regulate the chaperoning process by association and dissociation with Hsp90. Cdc37 (cell division cycle protein 37) is a co-chaperone of Hsp90 has been characterized as a protein kinase recruiting subunit of the Hsp90 machinery (Hunter and Poon, 1997). Cdc37 facilitates the maturation of these kinases by acting as an adaptor and load kinases “onto” the Hsp90 complex (Pearl, 2005; Vaughan et al., 2006; Silverstein et al., 1998; Grover et al., 2011). The Hsp90 structure consists of three domains: (1) an amino terminal region (N terminal domain, NTD) that for ATP binding, (2) a middle region (M domain) that provides interaction sites for client proteins and co-chaperones to form an active ATPase and

(3) a carboxy-terminal region (C domain) that contains a dimerization motif (Donnelly and Blagg, 2008; Ali et al., 2006; Prodromou and Pearl, 2003).

The Cdc37 (44.5-kDa) protein can be dissected into three domains (MacLean and Picard, 2003; Shao et al., 2003): an N-terminal domain (residues 1–127, 15.5 kDa), a middle domain (residues 147–276, 16 kDa), and a C-terminal domain (residues 283–378, 10.5 kDa). The middle domain Cdc37M is highly resistant to proteolytic digestion and was found to be the most stable domain of Cdc37 (Zhang et al., 2004). Cdc37 associates with the N-terminal portion of protein kinases (Terasawa and Minami, 2005; Zhao et al., 2004). The 20 residue client binding site of Cdc37 (residues 181–200) and the glycine rich loop in the N-terminal portion of the protein kinases are both necessary for physical interaction between the two proteins (Terasawa and Minami, 2005; Terasawa et al., 2006). The middle segment of Cdc37 has also been shown to interact with the N-terminal ATP-binding site domain of HSP90 (Roe et al., 2004).

Chaperone function necessarily requires dimerization of two Hsp90 protomers through their C domains (Wayne and Bolon, 2007; Trepel et al., 2010). There are two main ways to modulate chaperoning process by small molecule inhibitors: (1) inhibition of ATPase activity of Hsp90 which is essential for dimerization and successful performance of chaperone cycle, (2) prevention of proper association of co-chaperones with Hsp90 which is mandatory process for recruitment of client proteins.

2.1.3 Mouse double minute 2 (MDM2)

The p53 (tumor suppressor protein) is one of the key players which regulate the cell cycle, apoptosis, and DNA repair to protect cells from malignant transformation (Levine, 1997; Vogelstein et al., 2000; Vassilev et al., 2004). However, activity of p53 is regulated by MDM2, a protein that inhibits the ability of p53 to bind to DNA and activate transcription. Tumors have over expressed MDM2 which led to down regulated tumor suppressor activity of p53 (Dastidar, 2009). The p53 interacts with MDM2 by inserting its hydrophobic residues (Phe19, Trp23 and Leu26) into a deep groove in MDM2 (Allen et al., 2009). Human MDM2 is a 491amino acid long phosphoprotein that interacts through its N terminal domain with an ahelix present in the N terminal transactivation domain of p53 (Kussie et al., 1996). This entails several negative effects on p53. MDM2 binding to the N terminal transactivation domain of p53 blocks its transcriptional activity directly (Chen et al., 1993; Oliner et al., 1993). More importantly, MDM2 functions as the E3 ligase that ubiquitinates p53 for proteasome degradation (Haupt et al., 1997; Kubbutat and Vousden, 1997). The crystal structure of the p53-MDM2 complex has been solved (Kussie et al., 1996). The biochemical basis of MDM2-mediated inhibition of p53 function was further elucidated by crystallographic data that showed that the amino terminal domain of MDM2 forms a deep hydrophobic cleft into which the transactivation domain of p53 binds, thereby concealing itself from interaction with the transcriptional machinery (Kussie et al., 1996). This has been confirmed by biochemical analysis. The direct interaction between the two proteins has been localized to a relatively small (aa 25–

109) hydrophobic pocket domain at the N terminus of MDM2 and a 15-aa amphipathic peptide at the N terminus of p53 (Chen et al., 1993; Kussie et al., 1996). The minimal MDM2-binding site on the p53 protein was subsequently mapped within residues 18–26 (Chen et al., 1993; Bottger et al., 1996; 1997a, b). Site-directed mutagenesis has shown the importance of p53 residues Leu14, Phe19, Leu22, Trp23, and Leu26, of which Phe19, Trp23, and Leu26 are the most critical (Chen et al., 1993, Bottger et al., 1997). Accordingly, the MDM2-binding site p53 mutants are resistant to degradation by MDM2 (Haupt et al., 1997; Kubbutat et al., 1997; Kubbutat and Vousden, 1997). Similarly, mutations of MDM2 at residues Gly58, Glu68, Val75, or Cys77 result in lack of p53 binding (Freedman et al., 1997). The interacting domains show a tight keylock configuration of the p53-MDM2 interface. The hydrophobic side of the amphipathic p53 a-helix, which is formed by residue 19–26 (Phe19, Trp23, and Leu26), fits deeply into the hydrophobic cleft of MDM2. Thr18 is very important for the stability of the p53 α-helix (Kussie et al., 1996). The MDM2 cleft is formed by the residues 26–108 and consists of two structurally similar portions that fold up into a deep groove lined by 14 hydrophobic and aromatic residues (Kussie et al., 1996).

Many peptide inhibitors that mimic the MDM2–p53 interaction have been reported but these inhibitors display only modest potency because they have poor membrane permeability (Bautista et al., 2010; Fasan et al., 2004; Kritzer et al., 2004; Stoll et al., 2001; Zhao et al., 2002). Several different small-molecule inhibitors have been designed by structure-based methods to interrupt the binding of p53 to MDM2

which mainly include Nutlins (based on cis-imidazolidine) (Chene, 2003; Vassilev et al., 2004), benzodiazepinedione derivatives (Koblish et al., 2006; Popowicz et al., 2010) and spirooxindole (Shangary and Wang, 2008; 2009). These studies make MDM2 apo form and MDM2-p53 attractive target for transcription regulation and induction of apoptosis in cancer cells.

2.1.4 Vascular epithelial growth factor receptor 2 (VEGFR-2) kinase

Vascular endothelial growth factor (VEGF) is a potent growth promotor that is highly specific for vascular endothelial cells (Dvorak et al., 1995). VEGF is a strong angiogenic agent that increases vessel permeability and enhances endothelial cell growth, migration, proliferation and differentiation (Ferrara et al., 2003). Angiogenesis plays crucial role in the pathogenesis of cancer alongwith proliferative retinopathies, and rheumatoid arthritis (Folkman, 1990; Klagsbrun et al., 1991; Folkman et al., 1992). VEGFs and their tyrosine kinase receptors (VEGFRs) have been recognized as attractive targets for the inhibition of angiogenesis (Shawver et al., 1997; Traxler, 2003). The growth promoting and angiogenic effects of VEGF are mainly mediated by two receptor tyrosine kinases (RTKs): VEGFR-1 kinase and VEGFR-2 kinase (kinase insert domain receptor (KDR) kinase) (De Vries et al., 1992; Shalaby et al., 1995). Expression of VEGFRs varies in specific endothelial cell layers. The VEGFR-2 is located on almost all endothelial cells; however, the VEGFR-1 and -3 are alternatively located on endothelial cells in distinct vascular layers (Hicklin et al., 2005). The inhibition of VEGF signaling not only blocks

angiogenesis in tumors but can also change or destroy tumor vessels (Yang et al., 2003). Therefore, VEGFR-2 is an attractive target for biological cancer therapies (Lee et al., 2010). It is possible to interfere with VEGF signailing from the extracellular as well as from the intracellular site. In the extracellular region antibodies, soluble receptors and VEGF antagonists can avoid binding of the VEGF to the ligand binding site of the receptor (Los et al., 2007), or the inhibition of the VEGFR-2 in the intracellular region by blocking the ATP-binding site of the tyrosine kinase which is required to perform kinase activity (Underiner et al., 2004; Schmidt et al., 2008). The difficulties associated with the competitive inhibition of protein–protein interactions by small molecular weight compounds made targeting the catalytic site of kinases with ATP-competitive inhibitors is a more promising approach for drug intervention (Cochran et al., 2000, Yang et al., 2010).

2.2 Quercetin

Quercetin is a well known molecule present in most plants, fruits and vegetables (Hertog et al., 1993a). Quercetin has several beneficial properties in several diseased state including inflammation, atherosclerosis, thrombosis, hypertension, and arrhythmia as well as modulation of cancer-related multidrug resistance (Hertog et al., 1993b; Formica and Regelsont, 1995; Bischoff, 2008; Chen et al., 2010; Kressler et al., 2011; Mendoza and Burd, 2011; Larson et al., 2010; Russo et al., 2012).

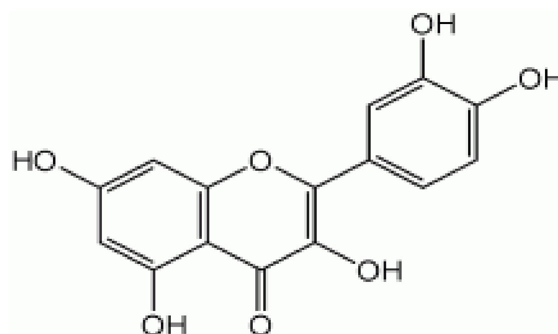

Fig 2.1. Structure of quercetin

2.4.1 Ethnopharmacological relevance of quercetin

Quercetin is present in 19 of 112 plants that are major constituent in Chinese medicine (Caia et al., 2004). Quercetin is also found in plant extracts that are well known for neuroprotective and anticancer effects. Several plants utilized in Ayurvedic medicine, Such as *Ginkgo biloba*, Kava-kava, St John's wort, Valerian, *Bacopa monniera* and *Convolvulus pluricaulis*, widely used for their effectiveness in brain disorders identified for containing quercetin (Kumar, 2006, Kumar et al., 2007). *G. biloba* is widely used to treat dementia (Weinmann et al., 2010) and *B. monniera* has been regarded as a ‘‘brain tonic’’ and has also been demonstrated to elicit neuroprotective effects (Jyoti and Sharma, 2006). *Centella asiatica*, a plant that is utilized for its activities on the nervous system, which have been experimentally demonstrated to affect the amyloid cascade and to alter the amyloid beta pathology in the brains of the transgenic PSAPP mice model of Alzheimer's disease (Dhanasekaran et al., 2009).

The ethnopharmacological relevance of several quercetin containing plants that are commonly used for cancer treatment: *Hydrocotyle sibthorpioides*,

Glyptostrobus pensilis, *Solanum lyratum* (from China), *Fabiana imbricata* (from Chile), *Castilleja tenuiflora* (from Mе́xico), *Uncaria tomentosa* (from Perú), *Oldenlandia diffusa* (from Singapore) and *Azadirachta indica* (from Nigeria) (Graham et al., 2000).

2.2.1 Quercetin and anticancer effects

Numerous in vitro studies have shown anticancer effects of quercetin in a variety of cancer cell lines including HeLa (cervical cancer, Vidya Priyadarsini et al., 2010); CWR22Rv1(prostate cancer, Hsieh and Wu, 2009); U2.US/ MTX300 (osteosarcoma, Xie etal., 2010); HT-29 (colorectal xenografts, Priego et al.,2008); MDA-MB-453 (breast cancer, Choi et al., 2008); U138MG (glioma, Braganhol et al., 2006); myeloid leukemia (Duraj et al.,2005); and oral cavity cancer (Kang et al., 2010). Oral administration can prevent induced carcinogenesis in the colon (Murakami et al., 2008). Quercetin can inhibit melanoma growth, invasion and metastatic potential (Caltagirone et al., 2000). When administered in the diet, quercetin was able to inhibit the initiation, growth and/or dissemination of induced tumors in experimental animal models (Yang et al., 2001).

2.3 Taxifolin

Taxifolinis a flavonoid commonly found in onions (Slimestad et al., 2007), milkthistle (Wallace et al., 2005), Frenchmaritime bark (Rohdewald, 2002) and Douglas fir bark (Kiehlmann and Edmond, 1995).

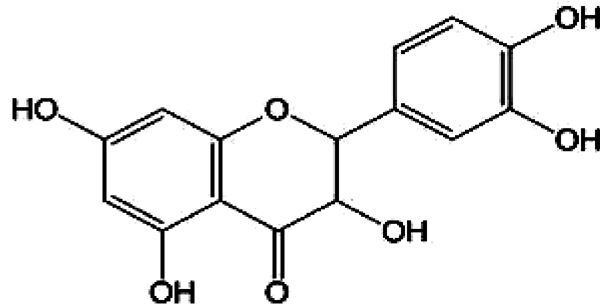

Fig 2.2. Structure of taxifolin

To date taxifolin is rarely used as a single compound but it is found in complex preparations such as silymarin (Legalon™), Pycnogenol® and Venoruton® (Weidmann, 2012). Silymarin, for example, is an extract from the seeds of the Milk Thistle plant (*Silybum marianum*) licenced for the treatment of chronic liver disease in Germany (Blumenthal and Busse, 1998) and more recently for the prevention of recurrent hepatitis C in liver transplant recipients by the European Medicines Agency (EMEA, 2010) (Weidmann, 2012). Initially silymarin was considered to be a pure compound (7-chromanol-3-methyl-taxifolin) (Hahn et al., 1968) but quantitatively determined 7 flavonolignans from the silymarin mixture (Ding et al., 2001) and one flavonoid (taxifolin) in addition to fatty acids and other polyphenolic compounds (Kim et al., 2003). Kvasniča et al., (2003) were the first to suggest that silymarin may contain "impurities" which have a more potent antioxidant capacity than any of the flavonoid isomers themselves. As the attention for silymarin in the treatment of liver diseases has shifted to its therapeutic potential in cancer treatment (Flaig et al., 2007, 2010), the investigation of taxifolin as a therapeutic agent is attracting more and more interest. Although the majority of research on taxifolin to date has focused on the

identification of molecular targets in silico and in vitro it is conceivable that taxifolin, although a minor component in extracts such as silymarin, Pycnogenol® and Venoruton® may contribute to their therapeutic efficacy (Weidmann, 2012).

2.4 Molecular Docking

In silico methods became widely used in the fields of structural molecular biology and structure-based drug design with the rapid increase in computational power. Molecular docking (Lengauer and Rarey, 1996; Gschwend et al., 1996; Kitchen et al., 2004) is one of these in silico techniques.

Docking is a method which predicts preferred orientation (on the basis of binding energy) of one molecule to the second to form a stable complex. In the field of drug design, first molecule is usually protein/enzyme/DNA and the second one is small organic molecule/small peptides as potential drug candidate.

Knowledge of preferred orientation of ligand and protein used to predict binding affinity and to discriminate high-affinity drug candidates from the low-affinity compounds.

2.4.1Lock and key analogy

Molecular docking is sometimes described as a problem of lock and key, where one is interested in finding the correct orientation of a key (ligand) that will open the lock (protein). Although this analogy is simple to understand but it does not account for inherent flexibility of proteins and ligands which is why more appropriate term hand-in-glove is used sometimes (Rangaraju and Rao, 2013).

2.4.2 Rigid-body docking vs. flexible docking

The docking problem involves many degrees of freedom (Leach, 2001). There are three translational and three rotational degrees of freedom for each molecule as well as the conformational degrees of freedom for both molecules. The simplest approach to docking is to take into account only translational and rotational degrees of freedom and treat both receptor and ligand as rigid objects. This approach is known as rigid-body docking (Leach, 2001). It depends from case to case, whether this approximation is accurate enough or not. If there are significant conformational changes within the molecules during the complex formation, this approach is inadequate. However, generation and scoring of all possible conformations is prohibitively expensive in computer time. Flexible docking algorithms (Baxter et al., 1998) must therefore take into consideration only a selected subset of possible conformational changes. Today with continual increase in computational resources, ligands are often considered flexible and depending on required accuracy, flexibility of amino acid side chains in the vicinity of active site may be considered as well.

2.4.3 Search algorithms

The search space which the docking software should take into account theoretically consists of every possible conformation and orientation of the receptor and ligand. While it is impossible to exhaustively explore this search space, efficient search algorithm is able to explore its large portion and identify global extrema (i.e. minima in the energy corresponding to the preferred conformations) (Halperin et al., 2002).

The docking problem can be handled manually with help of interactive computer graphics.

2.4.4 Genetic algorithms

Genetic algorithms (Oshiro et al., 1995) are search methods that mimic the process of evolution by incorporation of techniques inspired by natural evolution, such as inheritance, mutation or crossover. In genetic algorithm, an initial population of one-dimensional strings (called chromosomes), which encode candidate solutions (individuals) evolves toward better solutions (Khan and Alam, 2012). In case of a molecular docking, each individual may represent one possible system configuration and each string may contain information about its conformation. At the beginning, initial population is randomly generated. In the next step, a subset of the initial population is chosen on the basis of the results of the fitness function, which evaluates the quality of a particular individual. This subset is subsequently used to produce next generation. New generations are produced until a certain number of steps are performed or until a required level of fitness is reached. One example of genetic algorithm which was used in this work is Lamarckian Genetic Algorithm (LGA) (Ryška, 2011).

2.4.5 Lamarckian Genetic Algorithm

LGA is hybrid genetic algorithm named after Jean-Baptiste Lamarck (1744-1829), a French soldier and academic, who proposed an idea that organisms can pass on characteristics that they learned or acquired during their life to their offspring

(Bowler, 2003). While this idea contradicts Mendelian genetics and was later disproved, its implementation into genetic algorithm may lead to more accurate docking results. Genome is in LGA represented by floating point genes (unlike classical genetic algorithms, which use binary representation), each of which encodes one state variable describing molecular position, orientation and conformation.

2.4.6 Scoring function

Search algorithms are able to quickly generate large number of possible conformations. The "quality" of these possible solutions needs to be compared, so that best binding modes can be selected. This is the purpose of a scoring function used in docking software (Ajay and Murcko, 1995). Many of the scoring functions in common use attempt to approximate the binding free energy for the ligand binding to the receptor. A low (negative) energy indicates stable system and thus a likely receptor-ligand binding interaction. While many ways to predict free energies of binding exist, most of them are too computationally expensive to be of use in the field of molecular docking. These simplified scoring functions usually assume that the binding free energy can be written as a sum of several additive components representing various contributions to the binding free energy. An equation of this kind would have the following contributions:

$$\Delta G_{bind} = \Delta G_{solvent} + \Delta G_{conf} + \Delta G_{int} + \Delta G_{rot} + \Delta G_{r/t} + \Delta G_{vib} \quad (2.1)$$

$\Delta G_{solvent}$ represents contributions of solvent effects, which arise from the interaction of the solvent and ligand, receptor and the intermolecular complex. ΔG_{conf} arises from the conformational changes in both protein and especially more flexible ligand. ΔG_{int}

stands for the free energy of specific protein-ligand interactions. ΔG_{rot} is the free energy loss caused by freezing of the internal rotations. $\Delta G_{r/t}$ is a change in rotational and translational free energy due to association of receptor and ligand, forming a single body and ΔG_{vib} corresponds to free energy changes in vibrational modes (Bostrøm et al., 1998; Leach, 2001, Ryška, 2011).

2.5 Molecular dynamics simulation

Molecular dynamics simulations of proteins were initially developed in the early 1980s (McCammon et al., 1980) to use the emerging power of computers to study the motions of proteins and other biopolymers. Molecular dynamics simulations have been particularly successful in studying the protein folding problem (Zhang et al., 2009; Cho et al., 2009) and the impact of ligand binding on protein motion (Radkiewicz et al., 2000; Salsbury et al., 2001). Classical all-atom molecular dynamics simulations start simple as simulations where Newtonian equation of motions is solved for each atom in the system (Schlick, 2008). These simulations require only three items (Salsbury, 2010):

(a) Initial coordinates of molecules,

(b) A potential,

(c) Algorithms for propagation.

The initial coordinates can be obtained from experimental structures by nuclear magnetic resonance (NMR), X ray crystallography etc, from models generated by sequence of amino acids or some combination these (Salsbury, 2010). The potential is obtained from a force field along with the coordinates. A force field is simple

parameterization of the energy surface of the protein. The most common and currently used force fields include the CHARMM (MacKerell et al., 1998), AMBER (Ponder and Case, 2003) and GROMACS (Oostenbrink et al., 2004). These force fields are associated with particular modeling suites to simulate a wide-variety of macromolecules. It was reported often that simulations performed on the same structure with different force fields generate consistent results (Price and Brooks, 2006; Salsbury et al., 2001; Salsbury et al., 2009).

The CHARMM (Brooks et al., 2009), AMBER (Case et al., 2005), GROMACS (Hess et al., 2008), and NAMD (Phillips et al., 2009) are four most commonly used simulation suites. CHARMM at one extreme is a very complete modeling program that perform the widest variety of simulation analyses but this flexibility comes at a cost a steep learning curve, although there is an ongoing attempt to alleviate this with a GUI (Graphics user interface) (Jo et al., 2009) and poorer parallel performance. NAMD has a much simpler scripting language and reduced functionally but the reduction in functionally is belongs to analysis and simulation methods but does contain the simulation methods needed for classical all-atom simulations (Salsbury, 2010). NAMD is the most capable of performing large classical all-atom simulations with particularly large proteins and protein complexes (Vasilyeva et al., 2009; Gumbart et al., 2009; Yao et al., 2008; Ghosh et al., 2007). GROMACS and AMBER are closer to NAMD in scope and complexity. GROMACS is not using a scripting language and possessing a large number of external tools for analysis. Gromacs has the advantage of being the only one of these four that is open-

source. The 'classical' force field consists of two major parts: the first describes interactions between atoms connected via covalent bonds and the second describes the non-covalent interactions using the Lennard–Jones and Coulomb terms (Ponder and Case, 2003). The Lennard–Jones term models the van der Waals interactions and the Coulomb term describes electrostatic interactions between charges (Werner et al., 2012).

2.5.1 Energy minimization

Energy minimisation in refers to strategies used to traverse the energy surface of a molecule in order to find a conformation close to the global minimum. The energy minimisation procedure is always trying to walk downhill to find a structure of lower energy. The process is usually continued until the difference in calculated energy between successive structures less than a specified threshold is found. Different starting conformations of the protein should converge to the same energy minimized conformation providing confidence that the minimum is global (Werner et al., 2012). Steepest descent, conjugate gradients and Newton–Raphson are three common algorithms used for energy minimization and often combined in various ways to provide a convergence to the global minimum. The energy minimisation procedure is mainly used to refine protein structures. Energy minimisation can also be used to study the effect of mutations in a proteinby re-orienting residues, mostly in the region of the mutation, to minimise unfavourable contacts and steric constraints and maximize favorable contacts (Werner et al., 2012).

2.5.2 Molecular dynamics

Before the molecular dynamics simulation can be performed, the simulation system has to be defined by setting up the atoms (protein, lipids, water, and ion) and their positions, the temperature, the pressure and also the system cell inside which the simulation takes place. An equilibration run is performed to avoid unnecessary distortion of the protein when the molecular dynamics simulation is started, during which all heavy protein atoms are restrained to their starting positions (Werner et al., 2012). One of the most important questions is how long one should simulate the system in order to see effects and whether these effects are stable or only temporary. The simulation times vary between a few nanoseconds and 1 ms (Voelz et al., 2010; Shaw et al., 2009) depending, in part, on the size of the system. In most molecular dynamics simulations, a system is considered stable when it is seen to equilibrate over a longer timescale. However, the timescale to simulate depends on the effects the user wants to observe and the available computer power (Werner et al., 2012).

It has become increasingly evident that molecular dynamics simulations have become valuable tools for analyses of protein structure, dynamics, protein-ligand interactions and protein-protein interactions.