Chapter I

Introduction
Molecular Recognition

The essential elements of life like information processing, replication, and metabolism occur largely by specific interaction between biological molecules protein, DNA etc. Molecular recognition implies specific interaction of two molecules through different types of bonding and is the fundamental basis in biological processes like enzyme catalysis, cellular signaling, transcription, translation, DNA replication and transport (Baron et al., 2013). These processes mainly involve protein-protein and protein-DNA interactions. Understanding how two molecules recognize each other is thus one of the fundamental issues in biochemistry. Molecular recognition is also a central topic in applied biochemistry because it determines whether a compound possesses useful clinical properties. The first understanding of this concept emanated from the field of enzymology. Dutch chemist Emil Fisher proposed in 1894 that the enzyme and substrate fit together "like lock and key". He first clearly formulated the idea that molecular recognition lies in the complementarities of interacting surfaces. A modern view on molecular recognition is that the interacting molecules are flexible and can change their shape during the recognition process. It has been observed experimentally for many protein-ligand interactions. At the molecular level, the complementarities between two molecules depend on several factors like: shape of two molecules, hydrogen bonding, electrostatic interaction, hydrophobic interaction, ion-ion interaction and van der Waals force etc (Chothia et al., 1975; Morgan et al., 1979; Connolly et al., 1986; Johnson et al., 1996; Jones et al., 1996). The molecular recognition may be static or dynamic in nature. Static recognition occurs between a single guest and a single host binding site. In dynamic recognition binding of first guest at the first
binding site induces a conformation change that affects the association of the second guest at the second binding site.

Figure 1.1: Association of a macromolecule with its ligand using static and dynamic molecular recognition
(http://en.wikipedia.org/wiki/Molecular_recognition#/media/File:Molecular_Recognition_Static_vs_Dynamic_cartoon.png)

The dynamic nature of molecular recognition is particularly important since it provides a mechanism to regulate binding in biological systems (Rebek, 2009). The strength of binding of a ligand to a macromolecule is governed by the free energy change in the binding process. This can be determined experimentally. The association of proteins with DNA can be followed by using a variety of in vitro and in vivo assays. The interactions of proteins with DNA influence the structure and function of the corresponding nucleic acid. Elucidating the roles that protein-nucleic acid complexes play in regulation of transcription, translation, DNA replication, RNA processing and translocation continues to revolutionize our understanding of cell biology and the mechanisms of disease. The biological importance of protein–
DNA interactions has been recognized since the early 1960s, starting with the discovery by Jacob and Monod of the lac operon and its regulation in *Escherichia coli* (Jacob et al., 1961). Next 50 years, studies of protein–DNA interactions have made significant contributions to most areas of molecular, cellular, and developmental biology. The first binding site sequences were determined in the early 1970s, which led to hypotheses about recognition mechanisms and the information required for regulatory systems to function (Stormo, 2013). Here in this work we try to understand the factors affecting the molecular recognition and their effects, using synthetic ligands. For this purpose we have taken the help of two systems 1) protein-DNA interaction 2) protein-protein interaction.

**Systems: On which studies have been done**

1. Gal Operon
2. S100A4/ Nonmuscle myosin II A
3. Beta-Catenin mediated Wnt signaling
Gal Operon

Structural genes:

Gal operon is one of the best-studied catabolite-sensitive operons of *Escherichia coli*. It contains four structural genes, *galK*, *galT*, *galE* and *galM*, which specify the enzymes galactokinase, galactose transferase, galactose epimerase, and mutarotase, respectively. Mutarotase converts the β-D-galactose into α-D-galactose. The former is formed when β-galactosidase hydrolyzes lactose. Then galactokinase, galactose transferase, galactose epimerase act in a sequence of steps to yield the overall reaction Galactose + ATP → glucose-1-phosphate + ADP (Figure 2) (Adhya *et al.*, 1979; Aiba *et al.*, 1981; Michaelis *et al.*, 1967; Buttin *et al.*, 1963).

![Figure 1.2: Operation of Gal Structural genes](image-url)
The promoters

The *gal* operon has a promoter region containing two overlapping promoters P1 and P2, separated by 5 bp and as a result, the transcription initiation site for P1 is 5 bp upstream from that for P2 (Adhya *et al.*, 1979; Mulligan *et al.*, 1984). Both *in vivo* and *in vitro* it was seen that the gal operon is controlled by two promoters: P1 and P2 (Musso *et al.*, 1977).

![Figure 1.3: Schematic diagram of Gal operon (Adopted from Adhya, S. (1996)).](image)

Left arrow indicates the transcription initiation point of P1 and right arrow indicates that for P2

Activation of promoters P1 and P2

The P1 promoter requires cyclic AMP (cAMP) and its receptor protein (CRP) for its activation. P1 is mostly responsible for expression of gal operon genes in wild-type cells. The activation of P2 does not require the presence of cyclic AMP-CRP (Adhya, *et al.*, 1979). In fact, presence of cyclic AMP-CRP inhibits the transcription initiation at P2 (Musso *et al.*, 1977). The target sequence (AS); where cAMP•CRP binds with DNA, lies 41.5 bp upstream from the P1 transcription start site. Binding of cAMP-CRP has opposite effects on activation of two promoters. Binding of
cAMP-CRP to AS represses transcription from P2 but activates transcription from P1. Moreover, P2 can support gal mRNA transcription in the presence of glucose when intracellular cAMP-CRP levels are low but P1 cannot do so. The presence of two promoters in the gal operon helps galactose, which is a carbon source and a precursor for lipopolysaccharide synthesis, to execute two roles in cellular metabolism. When galactose is not available in the growth medium, cells require the epimerase specified by the gal operon to convert glucose to galactose, which is used to synthesis lipopolysaccharide. Synthesis from P2 promoter permits the low level of epimerase formation required to convert glucose to galactose-1-phosphate so that lipopolysaccharide can be formed. If P1 was the only promoter, then epimerase could not be formed in presence of glucose as P1 requires cAMP-CRP activation. If P2 was the only promoter then in presence of galactose, the operon could not be fully induced by galactose because cAMP-CRP inhibits P2. Thus P2 which is a cAMP-CRP-independent promoter is required for background constitutive synthesis and cAMP-CRP-dependent promoter P1 is required to regulate high-level synthesis.

The Operators

The gal operon contains two operators. One is the external operator, \( \text{galO}_E (O_e) \) and the other one is the internal, \( \text{galO}_I (O_i) \). The two operators contain a 16 bp dyad symmetry and are separated by 113 bp. \( \text{galO}_E \) is present upstream (-60.5) from the start site of P1 transcription and \( \text{galO}_I \) is present in \( \text{galE} \) (+53.5). Both operators participate in repression in a constitutive manner, such that mutation of either operator destroys the switching off mechanism (Irani et al., 1983).
The Gal repressor (GalR)

The structural gene responsible for the repressor protein, GalR, is located far from the structural genes for the galactose enzymes. The gal repressor (GalR) is a homodimer having 37KDa subunits (von Wilcken-Bergmann et al., 1982). It contains a single tryptophan at position 162 and four cysteine molecules at positions 121, 159, 186 and 244. One dimer binds to galO_E and another to galO_I. The gal repressor has a C-terminal domain that binds the inducer D-galactose and an N-terminal domain with a helix-turn-helix motif that binds to a galO_E or galO_I half-site. The histone-like protein, HU, binds to the gal promoter region (+ 1). This binding causes DNA super coiling, resulting in transcription repression (Roy et al., 2005; Adhya et al., 1989, 1998; Choy et al., 1995; Aki et al., 1996).

Expression of gal gene

The expression of the gal operon of E. coli is under both positive and negative control. The former consists of a requirement for a cyclic adenosine 3’:5’-monophosphate and cyclic AMP receptor protein, which act together with RNA polymerase to form a pre-initiation complex at the gal promoter (De Crombrugghe et al., 1969; Nissley, et al., 1971). Negative control is exerted by the gal repressor protein that binds specifically to gal DNA (Adhya et al., 1966). Repression of the gal operon is overcome by the addition of inducers such as D-fucose or D-galactose (Parks, et al., 1971). Nakanishi and his co-workers showed that the repressor acts specifically to repress the synthesis of gal mRNA (Nakanishi, et al., 1973a).
Mechanism of repression

Repression of \textit{gal} mRNA synthesis by \textit{gal} repressor might occur through a variety of molecular mechanism. It might prevent formation of the pre-initiation complex between RNA polymerase and \textit{gal} DNA or it might act subsequent to the formation of the pre-initiation complex by preventing either initiation or elongation of transcription. According to Nakanishi, pre-initiation complex will not form when the repressor is pre-bound to the \textit{gal} DNA. Conversely, \textit{gal} repressor does not inhibit transcription once a pre-initiation complex is formed at the \textit{gal} promoter (Nakanishi, \textit{et al.}, 1973a).
S100A4/ Nonmuscle myosin IIA

S100 protein family

The S100 proteins are a group of small Ca\(^{2+}\) binding modulator proteins. This group contains 24 family members, S100A1, S100A2, S100A3, S100A4, S100A5, S100A6, S100A7, S100A8, S100A9, S100A10, S100A11, S100A12, S100A13, S100A14, S100A15, S100A16, S100B, S100G, S100P, S100Z, Calbindin D9k, Profilaggrin, Trychohyalin, and Repetin. S100A14, S100A15, S100A16, S100G and S100Z are newly discovered members in this family (Gribenko et al., 2001; Pietas et al., 2002; Donato et al., 2013). They are called S100 because they are soluble in 100%-saturated solution of ammonium sulfate at neutral pH (Moore et al., 1965).

The S100 proteins are initially characterized as a group of abundant low molecular weight (10-12 kDa) acidic proteins having two EF-hand Ca\(^{2+}\) binding loops: one of them N-terminal pseudo EF-hand contains 14 residues (EF1) and another C-terminal canonical EF-hand contains 12 residues (EF2) (Zimmer et al., 1995; Marenholz et al., 2004). The canonical EF-hand helices have oxygen-containing ligands in Ca\(^{2+}\)-binding domain and they use their backbone or side-chain oxygen contain ligand for binding (Strynadka et al., 1989). S100 proteins are specific EF-hand proteins because they have one canonical and other pseudo-canonical EF-hand. S100 protein uses backbone carbonyl oxygen for Ca\(^{2+}\) binding rather than that of side-chain oxygen (Kligman et al., 1988). The majority of members of this family exist as symmetrical, antiparallel homodimers, or antiparallel heterodimers in which the N- and C-terminal helices (helices 1 and 4) from each subunit form a tight four helix bundle that forms the dimer interface. The Ca\(^{2+}\) ion, which binds to the C-terminal EF-hand which is flanked by 3 and 4 helices, helps the S100-terget proteins to bind it. Calcium binding to EF2 induces a conformational rearrangement...
that changes the angle between the 3 and 4 helices and exposes a hydrophobic cleft that provides a binding surface for target proteins (Rety et al., 1999; Rustandi et al., 2000; Bhattacharya et al., 2003; Yap et al., 1999; Zimmer et al., 2003). Thus the biological activity of S100-family members depends upon the binding of Ca\(^{2+}\) with it (Donato, et al., 2001; Zimmer et al., 2010; Charpentier et al., 2010). Only one member, S100A10, does not undergo Ca\(^{2+}\) dependent conformational changes as it is in a permanent “Ca\(^{2+}\) on” state (Gerke et al., 1985; Donato et al., 1991; Rety et al., 1999; Hiezmann, 2002). The hinge region and the C-terminal extension play a critical role in the interaction of S100A1, S100B, S100A10, and S100A11 with several target proteins (Donato, 2001; McClintock et al., 2002). All S100 protein share common amino acid sequence domain. Hydrophobic regions are found in both the N- and C-termini. The N-terminal EF hand is located in a region rich in basic amino acids and the C-terminal EF-hand is located in a region rich in acidic amino acids. Some of the proteins have a blocked N terminus (Kligman et al., 1988).

![Schematic diagram of Ca\(^{2+}\) dependent activation of S100 protein](image)

**Figure 1.4:** Schematic diagram of Ca\(^{2+}\) dependent activation of S100 protein
Functions of S100 Proteins

The members of this multigenic S100 protein family play regulatory roles within cells and exert regulatory effects on target cells once released into the extracellular space (Donato et al., 1999, 2001; Schafer et al., 1996; Zimmer et al., 1995). S100 proteins are distributed in a cell- and tissue-specific manner (Donato et al., 2003). They are mainly expressed in vertebrates (Donato et al., 2001). Due to this type of distribution they interact with several types of effectors proteins within cells. Through these interactions these proteins are responsible for multiple cellular processes such as Cell growth and differentiation, cell cycle progression, cell motility, transcription, contraction, structural organization of membranes, dynamics of cytoskeleton constituents, and protection from oxidative cell damage, protein phosphorylation and secretion (Donato, 2003; Santamaria-Kisiel, et al., 2006).

Along with the regulation of normal physiology, increased expression of specific S100 family members is associated with many human diseases such as cancer, cardiac disease, and neurodegenerative and inflammatory disorders (Odink et al., 1987; Van Eldik, et al., 1994; Heizmann et al., 2007).
S100A4

S100A4 is a specific member of this large S100 family. It is located in a 2.05 Mbp segment of the genomic DNA of chromosome 1q21 region (Marenholz et al., 1996; Mischke et al., 1996; Schafer et al., 1996), where the S100 family of gene cluster, with the exception of S100B, occurs. S100A4 is known under several synonyms, which include 18A2/mts1, CAPL, PEL-98, 42A and p9Ka, and metastasin (Sherbet et al., 2009). S100A4 forms homodimers and interacts with many target proteins in a calcium-dependent manner. It possesses four helices and each subunit has two calcium-binding domains known as the pseudo and canonical (typical) EF hands, characterized by helix-loop-helix calcium-binding motifs. These EF-hands are Ca\(^{2+}\)-binding domains and differ greatly in binding affinity. The conformation of the pseudo-EF hand, but not the canonical EF hand, is similar to that of other S100 proteins. It is responsible for cell motility, invasion, cell growth, cell-cell adhesion in normal conditions and over-expression of it is associated with fibrosis and inflammation, rheumatoid arthritis, cardiac hypertrophy and cancer (Schneider et al., 2008; Grigorian et al., 2008). In case of cancers, S100A4 has been shown as a promoter of metastasis (Helfman et al., 2005; Garrett et al., 2006). Human S100A4 is one of the best-characterized members of the S100 protein family in terms of its role in contribution to tumor-induced angiogenesis (Semov et al., 2005; Barraclough et al., 1994). Over-expression of S100A4 has been observed in several metastatic cancers, including breast, pancreatic, prostate, gallbladder and lung cancer (Pedersen et al., 2002; Rosty et al., 2002; Gupta et al., 2003; Nakamura et al., 2002; Kimura et al., 2000). High levels of S100A4 protein have been shown to be present in human carcinoma cells, primary colorectal carcinomas, gastric adenocarcinoma and invasive pancreatic carcinomas (Rosty et al., 2002; Simpson et
In addition, S100A4 has been immunocytochemically detected in the carcinoma cells of human breast cancers (Platt-Higgins et al., 2000; Rudland et al., 2000). The metastatic phenotype in *in vitro* and *in vivo* models of mammary carcinogenesis has been strongly associated with up regulated levels of S100A4 (Barraclough et al., 1998). S100A4 has been shown to be secreted by cancer cells and is detectable in the serum of cancer patients indicating that S100A4 may serve as a prognostic marker in human cancers (Ambartsumian et al., 2001, Garrett et al., 2006). The main features of these pathologies are cell motility and invasion (Boye et al., 2010). Thus inhibition of S100A4-dependent cell motility is a potentially valuable target for cancer therapy. S100A4 has no enzymatic activity but it exerts its function through binding with other proteins in cytoplasm, in nucleus and also in the intercellular spaces (Boye et al., 2010). There are many protein targets, such as non-muscle myosin IIA (NMIIA) (Kriajevska et al., 1994), actin, tumor suppressor p53 (Grigorian et al., 2001), S100A1, tropomyosin (Takenaga, et al., 1994) and liprin β1 (Kriajevska et al., 2002). Among them, interaction between S100A4 and NMIIA is one of the most important one, as this interaction is an established regulator of cell motility (Li et al., 2006, 2010; Ismail et al., 2008). S100A4 is also up regulated in beta-catenin mediated Wnt signaling pathway with resulting stimulation of tumor cell migration (Stein et al., 2006). Inhibition of beta-catenin down regulates S100A4 expression followed by cell migration and invasion (Stein et al., 2011). Human S100A4 is around 10 kDa molecular weight protein having 101 amino acids. The sequence of human S100A4 is

MACPLEKALDVMVSTFHKSFGKEGDKFKLKSNLDSNREVFQFQEFYCVFLSICAMMCNEFEGFDPDKQPRK.
Myosin II proteins

Myosin IIs are actin-based motor proteins in eukaryotic cells. They constitute a super family of motor proteins that play important parts in several cellular processes that require force and translocation (Holmes, 2007; Mooseker et al., 2007; El-Mezgueldi et al., 2007). They form bipolar filaments and are presumed to contract the actin cytoskeleton. Lower eukaryotes, such as Dictyostelium, express a single myosin II protein. In higher eukaryotes, expressions of variety of myosin II proteins occur. They are classified as muscle myosin IIs (MIIs) and non-muscle myosin IIs (NMIIIs) (Sellers, et al., 2000). Activities of NMIIIs play important roles in a variety of cell functions ranging from mitotic spindle assembly (Rosenblatt, et al., 2004) to cytokinesis (Straight, et al., 2003), cell migration (Totsukawa, et al., 2004), and growth cone outgrowth (Tullio, et al., 2001) and cell spreading (Wakatsuki, et al., 2003; van Leeuwen, et al., 1999; Arthur, et al., 2000).

Nonmuscle myosin II (NMII)

In higher eukaryotes, three different non-muscle myosin II isoforms have been identified. They are NMIIA, NMIIB, and NMIIC. Non-muscle myosin IIs are widely distributed in human and mouse organs but exhibit differential tissue expression patterns (Golomb, et al., 2004). NMIIIs have important roles in several aspects of cell motility including traction force generation, retrograde F-actin flow, and cell spreading (Brown et al., 2003; Lin et al., 1995, 1996; Diefenbach et al., 2002). NMII activity is also critical for fibroblasts to probe the extracellular substrate during spreading and migration (Giannone, et al., 2004). NMIIC is absent during the earliest stages of development (Golomb et al., 2004). In vertebrates most cells express comparable levels of NMIIA and NMIIB but neuronal cells
predominantly express NMIIB (Sellers et al., 2000; Rochlin et al., 1995). NMIIA and NMIIB have distinct but overlapping distributions in both neuronal (Rochlin et al., 1995) and non-neuronal cells (Maupin et al., 1994; Lo et al., 2004; Kolega et al., 1998, 2003; Saitoh et al., 2001). Depending on cell types, the same NMII isoform may be distributed differently. Additionally, both NMIIA and NMIIB interact with different proteins (Obungu et al., 2003; Huang et al., 2003; Krzewski et al., 2006; Clark et al., 2006) indicating their distinct functions. Finally, NMIIA and NMIIB undergo dynamic reorganization in motile cells (Kolega et al., 2003; Saitoh et al., 2001; Brown et al., 2003), implying that their biological functions are related to their dynamic reorganization. Deletion of NMIIB results in a decrease in cellular traction force (Lo et al., 2004; Bridgman et al., 2001; Meshel et al., 2005), the rate of neurite outgrowth (Tullio et al., 2001; Wylie et al., 1998), and the size of growth cones (Tullio et al., 2001). NMIIB null fibroblasts have defects in directional migration as a consequence of the multiple, unstable and disorganized protrusions of the cell edge, but cell movement rates are in the normal range (Lo et al., 2004). NMIIA has important roles in neurite retraction in neuronal cells, rearrangement of the actin cytoskeleton and also it decreases cell-matrix adhesion in neuroblastoma cells (Wylie et al., 2001, 2003). A similar phenotype is also observed in Hela cells when a truncated fragment of the myosin IIA heavy chain is over expressed (Wei et al., 2000). Knockout of NMIIA leads to impaired embryonic cell-cell adhesion, and similar effect occurs when E-cadherin and beta-catenin disappear from cell-cell adhesion sites (Conti et al., 2004). NMIIA has a primary role in developing a coherent, contractile network from one side of the cell to the other (Cai et al., 2006).
S100A4 and metastasis

Metastasis is defined as the spread of malignant cells from the primary tumor through the circulation to establish secondary growth in a different organ far from origin. There are many evidences to prove the direct involvement of S100A4 in the formation of metastasis from several different tumor types. From that evidences it is clear that S100A4 does not affect the initiation and growth of the primary tumors, only responsible for the diffusion of malignant cell (Maelandsmo et al., 1996; Levett et al., 2002; Ambartsumian et al., 1996; Davies et al., 1996). Metastasis is the main cause of death in patients with cancer and expression of S100A4 in tumor cell regulates the patient-survival very significantly. Patients with negative expression of S100A4 remain alive longer than the patients having positive-S100A4 expression (Rudland et al., 2000).
Beta-catenin mediated Wnt signaling

Wnt signaling

The Wnt Signaling pathways are a group of signal transduction pathway made by special proteins (known as Wnt protein) that transmit external signals to nucleus through cytoplasm. The first detail of the Wnt network was reported in 1982 with the identification of the proto-oncogene int-1 in mice (Nusse et al., 1982). Later its homolog in Drosophila, Wingless, was shown to be required for proper wing formation (Baker, 1987). Nusse and his co-workers discovered this signaling pathway by research on oncogenic retroviruses (Nusse et al., 2012; Logan et al., 2004). Wnts are secreted as lipid-modified signaling proteins (Willert et al., 2003) that influence multiple processes in animal development (Nelson et al., 2004).

Although several growth factors are known to affect both gene expression and cell migration (Thiery et al., 2002), recent focus has been on the Wnt signaling pathway. The Wnt signaling pathway is highly conserved from C.elegans to humans and it plays crucial roles in cell-fate decisions throughout development (Wodarz et al., 1998).

Three Wnt signaling pathways have been characterized: the canonical Wnt pathway, the non-canonical planar cell polarity pathway, and the non-canonical Wnt/calcium pathway. All three Wnt signaling pathways are activated by the binding of a Wnt-protein ligand to a Frizzled family receptor, which passes the biological signal to the protein Dishevelled inside the cell. The canonical Wnt pathway leads to regulation of gene transcription, the non-canonical planar cell polarity pathway regulates the cytoskeleton that is responsible for the shape of the cell, and the non-canonical Wnt/calcium pathway regulates calcium inside the cell. Thus, there are two main categories: one is canonical and another is non-
canonical. Canonical pathway involves the beta-catenin protein and non-canonical pathway operates without the beta-catenin protein.

**Canonical Wnt pathway**

The canonical Wnt pathway plays critical roles in embryonic development, stem cell growth, and tumorigenesis. Deregulation of Wnt pathway leads to the formation of several types of cancers. In absence of Wnt, GSK-3, which is a kinase, phosphorylates the beta-catenin. Beta-catenin then associates with axin, complexed with GSK-3 and APC. The creation of the said complex increases the phosphorylation of beta-catenin by facilitating the action of GSK-3. When beta-catenin is phosphorylated it is degraded and cannot build up by sufficient amount in the cytosol. The above complex is known as the destruction complex. In the canonical Wnt pathway, when Wnt protein is activated, it binds to the cell surface receptor of the Frizzled family member, causing the activation of Dishevelled (DSH) family of proteins. Activation of DSH facilitates the formation of cell surface receptor complex. As axin is an important member of the receptor complex, with activation of DSH it binds to the axin. Then axin cannot form destruction complex with GSK-3. So beta-catenin is inhibited and it facilitates the accumulation of beta-catenin in the cytosol. The accumulation of sufficient amount of beta-catenin in cytosol shuttles it to the nucleus. In nucleus, beta-catenin binds to BCL9 and transcription factor (Tcf), resulting in the transactivation of Wnt targeted genes (Behrens et al., 1996; Korinek et al., 1997).
Figure 1.6: Activation of the signaling pathway and fate of beta-Catenin in absence and presence of Wnt (Adopted from MacDonald et al., 2009)

Non-canonical Wnt pathway

The non-canonical Wnt pathway is mainly characterized in two subdivisions; one is the non-canonical planar cell polarity (PCP) pathway, and other is the non-canonical Wnt/calcium pathway. Both of these pathways do not involve the beta-catenin protein. But like the canonical pathway both of the non-canonical pathways are activated by binding of the Wnt protein to the cell surface receptor Fz protein and its co-receptor. Then the receptor binds with Dsh, which uses its PDZ and DEP domains to form a complex with downstream proteins to continue the signaling pathway.
Figure 1.7: The non-canonical PCP pathway

Figure 1.8: The non-canonical Wnt/Calcium pathway
Beta-catenin

Beta-catenin is the most important member in the canonical Wnt signaling pathway. Beta-catenin was originally identified in cell adherens junctions to connect cadherin to the alpha-catenin and the actin cytoskeleton (Hulsken et al., 1994; McCrea et al., 1991). Not only the cell adhesion beta-catenin plays critical roles in embryonic development and stem cell growth through Wnt signaling (Moon et al., 1998; Peifer et al., 2000; Reya et al., 2005). Deregulation of beta-catenin is associated with several types of human diseases like cancers, Alzheimer’s disease etc (Moon et al., 2002, 2004; Peifer et al., 2000). Deregulation of beta-catenin occurs due to mutation in beta-catenin itself or in its binding partners (Bienz et al., 2000; Moon et al., 2004; Polakis, 1997, 2000; Kinzler et al., 1996). Oncogenic mutations of beta-catenin and the prevention of their degradation results in intracellular accumulation. These mutants can induce tumor formation in transgenic animals (Gat et al., 1998; Harada et al., 1999). The importance of beta-catenin in abnormal cell proliferation attained prominence after the discovery of oncogenic beta-catenin mutations in colon cancers bearing the wild-type APC allele (Morin et al., 1997; Sparks et al., 1998). The mutant beta-catenin protein is not degraded by APC, thus leading to its accumulation in the cytoplasm, resulting in uncontrolled cellular proliferation (Sparks et al., 1998).

The frequency of oncogenic mutations in beta-catenin is low but has been reported in a variety of human cancers (Polakis et al., 2000). Inhibition of beta-catenin can be done using small molecule inhibitors or siRNA (Verma et al., 2003). As beta-catenin is a necessary oncogene, pharmacological inhibition of oncogenic beta-catenin is likely to be an effective strategy for reversing the malignant properties of advanced human tumors (Kim et al., 2002). Many beta-catenin signaling
pathway inhibitors are under investigation with the potential aim of disrupting beta-catenin activity and its interaction with the transcription factors. Lepourcelet et al., made initial attempts to screen and identify compounds capable of disrupting Tcf/beta-catenin complexes (Lepourcelet et al., 2004). The following table (Table 1) contains some important inhibitors of beta-catenin and its binding partners in the Wnt signaling pathway.

**Table 1.1: Inhibitors of Beta-catenin**

<table>
<thead>
<tr>
<th>Inhibitor</th>
<th>Target</th>
</tr>
</thead>
<tbody>
<tr>
<td>PKF118-310, CGP049090, PKF115-584, PKF222-815, PKF118-744, iCRT-3,5,14, NC043</td>
<td>Beta-catenin/TCF interaction</td>
</tr>
<tr>
<td>Carnosic acid</td>
<td>Beta-catenin/BCL9</td>
</tr>
<tr>
<td>ICG001</td>
<td>Beta-catenin/CBP interaction</td>
</tr>
<tr>
<td>SB239063, SB203580</td>
<td>p38 MAPK</td>
</tr>
<tr>
<td>Bosutinib</td>
<td>Src-kinase</td>
</tr>
</tbody>
</table>

Beta-catenin contains total 781 amino acids among which residues from 138-664 form the central structural core of 12 armadillo repeats (Figure 1.9) (Huber et al., 1997) and residues 665-781 form the C-terminal domain which consist an alpha-helix (residues 667-683, referred to as helix C) and a loop with a $3_{10}$-helical turn (residues 684-691). The armadillo repeat region constitutes the positively charged groove which helps to form the binding surface for the most of the binding
partners (many of which are critical for the cell adhesion and Wnt signaling) of beta-catenin (Daniels et al., 2002; Eklof Spink et al., 2001; Graham et al., 2000, 2002; Ha et al., 2004; Huber et al., 1997, 2001; Poy et al., 2001; Xing et al., 2003, 2004, 2008; Zhurinsky et al., 2000). Sequences of beta-catenin terminal domains are less conserved than the armadillo repeat domain. Both the N- and C-terminal domains are sensitive to mild protease digestion and they are unable to form stable folded structure by themselves (Huber et al., 1997). But beta-catenin terminal domains together with armadillo repeats enable the functions of beta-catenin during protein-protein interaction. N-terminal domain connects the beta-catenin/E-cadherin complex to alpha-catenin, which is a key regulator of the actin cytoskeleton (Drees et al., 2005; Nagafuchi, 2001; Yamada et al., 2005). The majority of beta-catenin partners are involved in the regulation of Wnt-responsive gene transcription in the nucleus. Among them Tcf family members interact with beta-catenin armadillo repeats 3-10. In addition to the Tcf binding region, the first armadillo repeat (R1) and armadillo repeat 11 to the C-terminal (R11-C) were identified as essential regions for transactivation of Wnt target genes (Stadeli et al., 2006; Willert et al., 2006). Beta-catenin R1 interacts with BCL9 that in turn recruits pygopus which is a critical transcriptional co activator in Wnt signaling (Kramps et al., 2002). The beta-catenin R11 region has been shown to interact with many transcriptional co activators, such as parafibromin, Brg1, CBP/p300, MED12, which function in different stages of transcription, as well as transcriptional inhibitors such as ICAT and Chibby (Barker et al., 2001; Hecht et al., 2000; Kim et al., 2006; Mosimann et al., 2006; Sierra et al., 2006; Tago et al., 2000; Takemaru et al., 2000, 2003). The sequence of full length beta-catenin protein is as follows.
Figure 1.9: Diagram of full length Beta-catenin protein with its armadillo repeats
BCL9

Human BCL9 was originally identified as an oncoprotein associated with precursor B cell acute lymphoblastic leukaemia (Willis et al., 1998). BCL9 has been proposed both to interact with beta-catenin in the nucleus to activate transcription and to shuttle beta-catenin in and out of the nucleus (Townsley et al., 2004; Krieghoff et al., 2006). BCL9 has three conserved regions termed as homology domains (HD) 1-3. Genetic and biochemical studies have shown that HD2 of BCL9 (residue 349-377) interacts directly with the first four armadillo repeats of beta-catenin (Kramps et al., 2002; Sampietro et al., 2006). Residues 352-374 of human BCL9- HD2 form a continuous α helix that packs against the groove formed between helices 2 and 3 of armadillo repeat 1 of beta-catenin and forms a helix bundle with the three helices of the first armadillo repeat of beta-catenin. BCL9 acts as the core components of the transcription-activation complex of the Wnt signal transduction pathway. Thus beta-catenin/BCL9 interaction may be an ideal drug target for the inhibition of Wnt pathway target genes in cancer patients.
ICAT

ICAT, a negatively charged 81 residue protein is an inhibitor of beta-catenin/Tcf interaction in Wnt signaling pathway (Danette et al., 2002; Tago et al., 2000; Tutter et al., 2001). ICAT is required for normal embryonic development in *Xenopus* (Tago et al., 2000). ICAT and its homologous gene, LZIC, are localized in a human chromosome region that is frequently rearranged or deleted in various cancers (Katoh, 2001). Reduced level of ICAT was also found in many cancer cells (Reifenberger et al., 2002). The crystal structure of beta-catenin/ICAT complex shows that ICAT contains an N-terminal 3-helix bundle that binds armadillo repeats 10-12 and a C-terminal tail that, similar to Tcf and E-cadherin, binds in the groove formed by armadillo repeats 5-9 of beta-catenin. From the structural analysis it is clear that full-length ICAT selectively inhibits beta-catenin/Tcf binding without disrupting beta-catenin/cadherin binding. The 3-helix bundle provides the critical domain for anchoring ICAT to beta-catenin, whereas the tail is necessary for excluding Tcf from beta-catenin (Graham et al., 2000, 2001, 2002; Poy et al., 2001). Whereas the full length ICAT is required for the disruption of the interaction of beta-catenin with its partners, only the helical domain alone selectively blocks binding to p300 (Daniels et al., 2002). As ICAT has been found to present in both cytosol and nucleus, and since cadherin and Tcf bind beta-catenin in very similar ways, ICAT might also be expected to regulate cell adhesion. Thus, it is possible to design cancer therapeutics that inhibits beta-catenin mediated transcriptional activation without interfering cell adhesion from ICAT.
Tcf

Tcf family proteins are the downstream DNA-binding transcription factors and they are main partners of beta-catenin in the gene regulation in the canonical Wnt signaling pathway (Arce et al., 2006; Hoppler et al., 2007). In Drosophila and worm, a single Tcf gene is present, but mammals contain four Tcf genes, Tcf1, LEF1, Tcf3 and Tcf4. Tcf/LEF-1 proteins by themselves have no inherent transcriptional activity and repress transcription of Wnt target genes by recruiting co-repressors to the promoter (Roose et al., 1999; Barker et al., 2000). Wnt-induced beta-catenin stabilization and nuclear accumulation leads Tcf to complex with beta-catenin, and recruits other co-activators such as p300/CBP and TATA binding protein to the promoter region for gene activation (Hecht et al., 1999, 2000; Kato et al., 1999; Takemaru et al., 2000). The very N-terminal domain of Tcf/Lef-1 family proteins containing 60 amino acids is responsible for the beta-catenin binding (Beherens et al., 1996; Molenaar et al., 1996; Korinek et al., 1997; van de Wetering et al., 1997). Among those 60 residues of Tcf, Asp-16 and
Leu-48 are identified as critical residues for the interaction with beta-catenin (Omer et al., 1999). All Tcf/LEF-1 family members also have a highly conserved HMG DNA binding domain, located within the C-terminal half of the protein. Tcf family members interact with beta-catenin armadillo repeats 3-10 and anchor beta-catenin to specific promoters of transcription.

Figure 1.13: Schematic diagram of the different binding domain of Tcf
Techniques used to study the molecular recognition

Protein-protein, protein-DNA or interactions of small molecules with proteins or DNA are common phenomena of molecular recognition occurring in biology. Thus, many tools and techniques are considered to explain different types of molecular recognition. NMR-study, X-ray crystallography, Mass spectroscopy, Fluorescence spectroscopy, Circular dichroism (CD), Analytical ultra centrifugation, surface Plasmon resonance (SPR), Isothermal titration calorimetry (ITC) and Absorption spectroscopy are the common methods used for the above purpose. In our cases, fluorescence spectroscopy, ITC and CD were widely used to study the protein-protein and protein-DNA interactions.

Fluorescence Spectroscopy

Fluorescence is a phenomenon which described the emission of light by a molecule when that molecule relaxes back to the ground state from electronically excited singlet states. Absorption of a photon leads to excitation of a molecule from its ground state to a higher energy state. Fluorescence is one of the many processes by which the molecule returns to its ground state. In 1935, Jablonski described that this process occurs between the absorption and emission of light and is usually illustrated by the Jablonski diagram. Fluorescence spectral data are generally presented as emission spectra and emission spectra are typically independent of the excitation wavelength. In 1852, sir G.G. Stokes revealed that fluorescence occurs at longer wavelengths i.e. the energies of the emission is less than that of the absorption. This is known as Stokes shifts.
Fluorophore

Fluorescence classically occurs from aromatic molecules, known as fluorophores. With respect to biological macromolecules, fluorophores are generally divided into two main classes: intrinsic and extrinsic. Intrinsic fluorophores are those that occur naturally and extrinsic fluorophores are those added to the sample that does not show any desired fluorescence properties. Proteins containing the tryptophan, tyrosine and phenylalanine show fluorescence emission, which can be utilized for study of proteins. Most widely used fluorophores in experimental studies include dansyl chloride (DNS-Cl), carboxy-fluorescein and fluorescein isothiocyanate (FITC) etc.

Fluorescence of Tryptophan

In proteins, the dominant intrinsic fluorescence originates due to the indole group of tryptophan. Indole absorbs near 280 nm, and emits near 340 nm. The emission of tryptophan is highly sensitive to its local environment and thus it is used as the reporter for protein conformational changes during the interaction with its binding partners or solvent and protein unfolding. The emission of indole may be shifted if the group buried within a native protein and its emission may shift to longer wavelengths when the protein is unfolded.

Fluorescence measurements

Fluorescence measurements are classified into two types: one is steady-state and another is time-resolved. Steady-state measurements are performed with constant illumination and observation. The sample is illuminated with a continuous beam of light, and the emission spectrum is recorded. The time-resolved measurement is
used for measuring intensity decays or anisotropy decays. For these measurements the sample is exposed to a pulse of light, where the pulse width is typically shorter than the decay time of the sample. This intensity decay is recorded with a high-speed detection system that permits the intensity or anisotropy to be measured on the nanosecond scale.

**Fluorescence Anisotropy**

Fluorescence anisotropy measurement is one of the most helpful applications of the fluorescence spectroscopy. Anisotropy measurement provides information on the size and shape of a molecule or the rigidity of various molecular environments. Thus, this technique is very helpful to study the association of proteins with its ligands. The anisotropy measurements are based on the principle of photo selective excitation of fluorophores by polarized light. The origin of anisotropy is the existence of transition moments for absorption and emission that lie along specific directions within the fluorophore structure. Fluorophores preferentially absorb photons whose electric vectors are aligned parallel to the transition moment of the fluorophore. Hence, the excited state population is partially oriented. Upon excitation with polarized light the emission from the samples is also polarized. The extent of polarization of the emission is described in terms of anisotropy (r). Anisotropy measurements reveal the average angular displacement of the fluorophore that occurs between absorption and subsequent emission of a photon. The angular displacement is dependent upon the rate and extent of rotational diffusion during the lifetime of the excited state. Rotational diffusion changes the direction of the transition moments and is one common cause of depolarization. Increased rate of rotational diffusion is associated with
the decrease in the anisotropy value. The dependence of fluorescence anisotropy upon fluorophore motions has resulted in numerous applications of this technique in biochemical research.

**Measurements of fluorescence anisotropy**

For most of the experiments the sample is excited with vertically polarized light and the intensities of emission are measured parallel ($I_{\parallel}$) and perpendicular ($I_{\perp}$) direction with respect to the excitation direction. These emission intensity values are used to determine the anisotropy. Thus anisotropy is described as follows:

\[
\frac{I_{\parallel} - I_{\perp}}{I_{\parallel} + 2I_{\perp}}
\]

For some instrumental complexity, the measurements of the actual intensities ($I_{\parallel}$ and $I_{\perp}$) are difficult. To indicate the orientation of the excitation and emission polarizer two subscripts are used i.e. $I_{HV}$ corresponds to horizontally polarized excitation and vertically polarized emission. Similarly $I_{VV}$ indicates the vertically polarized excitation and vertically polarized emission. These notations are easier to recall. Now to measure the sensitivity of the instrument $S_H$ and $S_V$ are introduced, where $S_H$ and $S_V$ are the sensitivities of the emission channel for the horizontally and vertically polarized components, respectively. Now,

$I_{VV}=k S_V I_{\parallel}$

$I_{HV}=k S_H I_{\parallel}$

$\frac{I_{VV}}{I_{HV}}= \frac{S_V}{S_H} \frac{I_{\parallel}}{I_{\perp}} = G \frac{I_{\parallel}}{I_{\perp}}$

$\frac{I_{\parallel}}{I_{\perp}} = \frac{I_{VV}}{I_{HV}} \frac{1}{G}$

\[ \text{…………………………………………………………… ... (1)} \]
where $G = (S_v/S_H)$, the ratio of the sensitivities of the detection system for vertical and horizontal polarized light.

Thus, the measured intensities ratio is different from the true value by a factor $G$. To calculate the actual intensities ratio ($I_\perp/I_\parallel$), $G$ factor has to be determined. The $G$ factor is easily determined using the horizontally polarized excitation. With horizontally polarized excitation the excited-state distribution is rotated to lie along the observation axis, and then both horizontally and vertically polarized components are equal and proportional to $I_\parallel$. Thus

$$
(I_{HH}/I_{VV}) = (S_v I_\perp/S_H I_\parallel) = (S_v/S_H) = G
$$

(2)

Now putting the value of $(I_\perp/I_\parallel)$ from equation (1) in the equation of anisotropy ($r$), we get

$$
r = (I_{VV} - G I_{VH}) / (I_{VV} + 2G I_{VH})
$$

For a mixture of fluorophores, the average anisotropy is given by, $r = \Sigma f_i r_i$, where $r_i$ indicates the actual anisotropy observed from $i$th component and $f_i$ is the fractional contribution of the $i$th component to the total emission intensity.

As fluorescence anisotropy is very sensitive to the size and internal flexibility of a sample, it is very useful to study the ligand binding to the protein or DNA by measuring the anisotropy values at different concentrations of the samples. Those obtained anisotropy values are fitted in the proper equation to get the dissociation constant for a particular reaction.
Circular Dichroism (CD)

Dichroism is the phenomenon in which light absorption differs for different directions of polarization. Linear dichroism involves linearly polarized light where the electric vector is confined to a plane and it tells us about the structure of a molecule. Circular dichroism involves the circularly polarized light and here the magnitude is constant and the direction is modulated. The electric vector of circularly polarized light describes a helix and it may be right handed or left handed.

Circular dichroism (CD) spectroscopy is an optical technique that measures the difference in the absorption of left and right circularly polarized light (Woody, 1995). This phenomenon originated from the absorption of optically active chiral molecules. The compound may be active either by virtue of its intrinsic structure, or by being covalently linked to a chiral centre, or by being placed in an asymmetric environment. Circular dichroism is defined as the difference in extinction coefficients ($\varepsilon$) of right-handed and left-handed polarized light at a particular wavelength ($\lambda$).

$$\Delta A = A_L - A_R = (\varepsilon_L - \varepsilon_R)lc = \Delta \varepsilon lc$$ ...................................................(3)

Or the CD is also presented in terms of ellipticity ($\theta$), because ellipticity relates the measurement to optical rotator dispersion. Ellipticity in degrees is related to the difference in absorption by

$$\Delta A = \theta / 32.98$$ ...........................................................................................................(4)

We will view CD as a special kind of absorption spectroscopy and it will be expressed in terms of $\Delta \varepsilon$ and followed by molar ellipticity ([\theta]). Molar ellipticity is expressed as

$$[\theta] = (M^* \theta) / (10^* d^* c)$$ ...........................................................................................................(5)
Where, θ is observed ellipticity in degrees, M is the molecular weight, d is the path length in cm and c is the concentration in gm/ml.

Now combining the equations (3), (4) and (5) with Beer-Lambert law, we get the relation between molar ellipticity and Δε as

\[ [\theta] = 3298 \Delta \varepsilon \]

CD spectroscopy is very sensitive to the secondary structure of polypeptides and proteins and thus has been employed in many different fields, such as investigation of the proteins’ secondary structure and spectral studies of DNA in free state and also in case of ligand association (Lewis et al., 1974; Holm et al., 2010; Matsuo et al., 2004; Sreerama et al., 2000). Nowadays, the structures of small organic molecules such as oligo-peptides and oligo-nucleotides have been studied with this technique (Porumb et al., 2002; Zhou et al., 2007; Andrushchenko et al., 2010; Daura et al., 2003).

**Isothermal Titration Calorimetry (ITC)**

ITC is a physical technique to measure the thermodynamic parameters of molecular interaction in solution phase. This method has been an invaluable tool for understanding the forces that stabilize the folded conformations of proteins. Thus, this technique is mainly used to study the interaction of protein-protein, protein-DNA or small molecule with protein or DNA. ITC is a quantitative technique that is used to determine the binding affinity \( (K_a) \), enthalpy changes \( (\Delta H) \), and binding stoichiometry \( (n) \) of the interaction between two or more molecules. From these initial
measurements, Gibbs free energy changes ($\Delta G$) and entropy changes ($\Delta S$) can be determined using the relationship:

$$\Delta G = -RT \ln K_a$$

$$= \Delta H - T \Delta S$$

(Where $R$ is the gas constant and $T$ is the absolute temperature).
References


Chapter I

Chapter I

Chapter I


• Morgan, R.S., Miller, S.L. and McAdon, J. (1979): *J. Mol. Biol.*, 127, 31-39


Chapter I


