REVIEW OF LITERATURE
CHAPTER-2
REVIEW OF LITERATURE

THE IMMUNE SYSTEM

The immune response in higher vertebrates is a remarkably adaptive defense mechanism that has evolved to evade myriads of invading pathogens and malignant cellular growths. It is orchestrated through a complex interplay of wide variety of cell types and soluble molecules, collectively referred to as 'the Immune System'. The immune system functions through two interrelated components: the innate (non-adaptive) and the adaptive immune system. The important differences between these two components being specificity and memory associated with the response mounted against the pathogens. Several physical and chemical factors, apart from some specialized cells such as the blood monocytes, neutrophils, tissue macrophages, dendritic cells, NK cells, constitute the innate immune system. The adaptive immune system, however, functions centrally through specialized cells termed as the lymphocytes, which are derived from pleuripotent hematopoetic stem cells in the bone marrow. Two major populations of the lymphocytes, the B and the T lymphocytes, are distinguished primarily on the basis of function, maturation and the expression of cell surface markers.

ANTIGEN RECONGNITION BY T CELLS

A critical event in the activation of the adaptive immune response is the recognition of the antigen by T cell receptor (TCR) expressed on the surface of T lymphocytes. T lymphocytes recognize the antigen as a peptide bound to the major histocompatibility complex (MHC) molecule on the surface of antigen-presenting cells (APC). The interaction of TCR with the peptide-MHC complex on the surface of APCs, besides the interaction between the other cell surface molecules of the two
interacting cells, is referred to as 'the immunological synapse' and is a critical determinant of the activation of T-cells (Huppa and Davis, 2003). Sensitivity, specificity and context discrimination are three key properties of T cell antigen recognition. Since the MHC molecules are highly promiscuous in peptide binding, epitopes for T cell recognition are generally displayed at low copy number and consequently, the T cell recognition must be exquisitely sensitive (Davis et al., 1998). In addition, recognition must be highly specific, since the T cells must discriminate few antigenic peptide-MHC molecules among a vast excess of irrelevant, yet highly homologous, complexes of the same MHC molecules carrying a variety of self-peptides. Finally, T cells must be able to interpret the context in which a given antigen is presented to mount an effector response to infectious antigens or ignore innocuous or self-antigens (Lanzavecchia et al., 1999).

**T CELL ACTIVATION**

To explain the phenomenon of lymphocyte activation, a two-signal concept was proposed. According to this model, optimal activation of antigen-specific lymphocytes require specific antigen recognition by lymphocytes (signal 1) and additional signals (called the signal 2 or the co-stimulatory signals). In the absence of signal 2, lymphocytes fail to respond effectively to the antigenic stimulation and are rendered 'anergic'. For T cell activation, signal 1 is provided by the interaction of the peptide-antigen-MHC complex with the TCR while the co-stimulatory molecules expressed on the surface of APCs deliver the 'signal 2' (reviewed in Bretscher, 1999).

Though, the two signal model provided a useful model for explaining the phenomenon of T cell activation and 'Tolerance', it over simplified the contributions of each signal. The strength of TCR signal has a quantitative influence on T cell activation and differentiation. T cell activation can occur in the absence of signal 2 if
the TCR signal is very strong (Sharpe and Freeman, 2002). The number of TCRs triggered is an important parameter that determines the fate of antigenic stimulation. Irrespective of the nature and affinity of the ligand, T cells are activated to proliferate and produce cytokines when a threshold number of triggered TCRs are reached. Furthermore, this threshold is not absolute and it depends on whether or not co-stimulation is present. It is therefore a characteristic for a particular T cell-APC interaction. The support for this statement comes from the observation that the human T cell clones expressing ~ 30,000 TCRs, have a threshold of ~ 8000 TCRs in the absence and ~1000 in presence of CD28 mediated co-stimulation (Viola and Lanzavecchia, 1996) for their effective activation.

In addition to a quantitative effect of TCR ligation, both positive and negative second signal can be delivered to the T cells. The balance of these stimulatory and inhibitory signals is crucial to maximize protective immune responses while maintaining immunological tolerance and preventing autoimmunity (Sansom, 2000).

**ROLE OF B LYMPHOCYTES IN T CELL PRIMING**

The ability of B cells to present antigen to CD4+ T cells is well established. However, there are differences in opinion on when and where they perform the function of modulating T cell responses *in vivo*. The earliest evidences suggesting the significance of B-cell as important APCs came from the studies employing μSM (μ suppressed) mice that lacked peripheral B-cells due to administration of anti-μ antibodies from birth. Such studies employing μSM mice showed a severe impairment of CD4+ T cell priming when the antigen was injected locally in 'complete freund’s adjuvant' (Ron *et al.*, 1981; Ron *et al.*, 1983; Hayglass *et al.*, 1986; Janeway *et al.*, 1987; Ron and Sprent, 1987). These effects on T cell priming could be reversed with the adoptive transfer of B-cells prior to antigenic challenge suggesting a
direct role of B-cells as APCs for T cell priming in vivo (Ron and Sprent 1987; Janeway et al., 1987; Abbas et al., 1985). But, a few subsequent studies utilizing gene disruption approach to generate B cell deficient mice (such as disruption of transmembrane domain (μMT) (Kitamura et al., 1991) or JH segment of Ig heavy chain (JHD) (Chen et al., 1993) generating conflicting results, ranging from near normal (Epstein et al., 1995; Topham et al., 1996; Phillips et al., 1996) to severely impaired T cell responses (Liu et al., 1995; Linton et al., 2000; 2003). The gene disrupted, μMT, mice however, were shown to suffer severe developmental abnormalities associated with the spleen (Crowley et al., 1999). It was demonstrated that B cells play an important role for normal development of spleen and hence the latter approaches, in which B cells were absent even during the embryonic development, did not result in normal spleen development (Asano et al., 1996; Mandik-Nayak et al., 2001). Several subsequent reports described methods to overcome drawbacks of μMT mice and demonstrated that the B cell were indeed capable of efficiently priming T cell in vivo (Constant et al., 1999; Evans et al., 2000; Rivera et al., 2001; Rodriguez-Pinto and Morreno, 2005; Crawford et al., 2006).

Further, the role of B cells in productive T cell activation was suggested to be restricted to the priming of previously activated T cells only, on the basis of two major arguments. First, B cells only with the antigen-specific BCR efficiently take up the antigen for presentation to T cells and the frequency of such antigen-specific B cells would be too low for them to participate significantly in T cell priming. The second argument suggested that only the activated B cells could function in T cell priming and the activation of B cells was dependent entirely on the help from T cells themselves. However, there was no consensus in literature about the ability of naïve T cells to express molecules such as CD40 ligand which could help B cells enhance
their own activation status (Roy et al., 1995; Jaiswal et al., 1997; Croft et al., 1997). Thus, in the proposed scheme of events, productive T cell activation essentially required T cell priming by another APC type such as dendritic cells, bearing adequate antigenic complexes and costimulatory molecules, before a subsequent productive interaction with B cells can take place. However, the evidences in support of B cells being able to stimulate resting naïve T cells in vivo are gradually emerging. Studies have shown that naïve T cells induce small B cells to express the costimulatory signal that T cells need to proliferate in vitro and in vivo in absence of any other APC type or antigen receptor signaling, though some what less effective than dendritic cells (Evans et al., 2000). Further, studies employing B cells from CD40 deficient mice demonstrated that CD40 on B cells is necessary for T cell proliferation and sustained induction of various costimulatory molecules. The requirement of CD40 mediated costimulation during the process of T cell activation was more pronounced with B cells as APCs (Ozaki et al., 1999).

THE CO-STIMULATORY MOLECULES

The family of T cell co-stimulatory molecules is composed of several proteins expressed on the T cell surface, referred to as the CD28 family, that interact with the proteins expressed on the surface of APCs, classified into the B7 family. Ever since the identification of CD28, the first member of this family, several new members have been added to both B7 and CD28 family of molecules. CD28, ICOS, PD-1, and BTLA are expressed on T cells and constitute the CD28 family of receptors. B7-1 (CD80), B7-2 (CD86), ICOS ligand, PD-L1 (B7-H1), PD-L2 (B7-DC), B7-H3, and B7-H4 (B7x/B7-S1) constitute the B7 family of molecules expressed on APCs (Greenwald et al., 2005). The identification of several critical pathways and a growing number of members in B7:CD28 family suggests that there might be a
functional hierarchy of these co-stimulatory molecules for regulating the response of naïve, effector and memory T cells. Further, given the significance of these pathways in regulating T cell responses, they provide promising therapeutic targets (Carreno and Collins, 2002).

**B7-1/B7-2/CD28/CTLA-4 Pathway**

The B7-1/B7-2/CD28/CTLA-4 pathway is the best-characterized T cell co-stimulatory pathway and is crucial to T cell activation and tolerance. It includes two B7 family members, B7-1 (CD80) and B7-2 (CD86) that have dual specificity for two CD28 family members, the stimulatory receptor CD28 and the inhibitory receptor CTLA-4 (CD152) (Sharpe and Freedman, 2002). CD28 is constitutively expressed on the surface of T-cells while the expression of CTLA-4 is rapidly upregulated following T cell activation (Freeman *et al.*, 1993). On most of the APCs populations, B7-2 is constitutively expressed at low levels and rapidly upregulated while B7-1 expression is inducible and expressed much later after activation. The difference in the pattern of expression of B7-1 and B7-2 could be suggestive of the differences in their function during an immune response (Hathcock *et al.*, 1994).

CD28 transmits a signal that synergizes with a signal from TCR to promote T cell activation. Engagement of CD28 alone has no physiological effect in the absence of TCR stimulation. CD28 signaling regulates the threshold for T cell activation and significantly decreases the number of TCR engagement needed for effective T cell activation (Lanzavecchia *et al.*, 1999). CD28 optimizes the responses of previously activated T cells promoting production of interleukin 2 (IL-2) and T cell survival. CTLA-4 engagement delivers a negative signal, inhibiting TCR and CD28 mediated signal transduction. It inhibits IL-2 production and progression through cell cycle and terminates T cell responses (Walunas, 1996). CTLA-4 blockade *in vivo* augments
anti-tumour immunity (Leach et al., 1996) and exacerbates autoimmune responses (Perrin et al., 1996). The functional outcome of an immune response involves a balance between CD28-mediated T-cell activation and CTLA-4 mediated inhibition. Though, it is not entirely clear how signals through CD28 and CTLA-4 are coordinated. It is likely that CTLA-4, the higher affinity receptor for B7-1 and B7-2, might out-compete CD28 for binding to B7-1 and B7-2, induce immunosuppressive cytokines or directly antagonize CD28-mediated signaling and/or TCR-mediated signaling (Sharpe and Freeman, 2002).

**CD80 And Its Role In Immune Response**

CD80 has been suggested to be one of the most potent of all costimulatory molecules known. Also, it is one the best-characterized molecule in this family. Several lines of evidence from both *in vitro* and *in vivo* studies have shown that the interaction of CD80 molecules with their counter-receptors, CD28 and CTLA-4, plays an irreplaceable role in T cell clonal expansion and induction of effector functions (Powers et al., 1994 Dubey et al., 1995). Inhibition of the interaction between CD28 with its ligand blocks T cell activation and results in clonal anergy (Sperling et al., 1993). Many studies have suggested the significance of this signaling pathway in tissue transplantation, autoimmune diseases and protective immunity against tumors (Blazar et al., 1994; Verwilghen et al., 1994; Matulonis et al., 1995).

The expression of CD80 is regulated by a variety of stimuli including lipopolysaccharide (Hathcock et al., 1994), CD40 crosslinking (Caux et al., 1994) IL-4 (Stack et al., 1994) and oxidative stress (Donepudi et al., 2001). It has been demonstrated that non-APCs can be induced to express CD80 when transformed which may provide a mechanism for immnosurveillance against tumor cells (Antonia et al., 1995). The administration of low-doses of melphalan to mice bearing
a large MOPC-315 plasmacytoma led to a rapid upregulation of CD80, but not CD86 expression on the surface of MOPC-315 tumor cells. *In vitro* studies also demonstrated a preferential upregulation of CD80 surface expression following *in vitro* exposure of tumor cells to melphalan, γ-irradiation and mitomycin C. It required *de novo* protein and RNA synthesis and was associated with accumulation of mRNA for CD80 within 4-8 hrs (Sojka *et al.*, 2000).

**The Genomic Organization Of CD80**

The genomic organization of both human and mouse CD80 loci were first reported by Selvakumar *et al.*, 1992; 1993. The human gene was found to contain six exons spanning approximately 32 kb. The first exon was non-protein coding and the second contained the translational start site. Exons 3 and 4 corresponded to the extracellular Ig-like domains, whereas the fifth and sixth, encoded the transmembrane and cytoplasmic domains, respectively. The murine CD80 gene was found to be composed of five exons spanning more than 39 kb. The first exon corresponded to the 5' untranslated region and signal peptide, exons 2 and 3 encoded the IgV-like and IgC-like domains, respectively, exon 4 encoded the transmembrane domain and exon 5 encoded the cytoplasmic domain and 3' untranslated regions. Borriello *et al.*, 1994 reported the identification of a previously unrecognized sixth exon located approximately 7 kb downstream of the cytoplasmic domain-encoding exon 5. RT-PCR analysis demonstrated that exon 6 could be spliced to exon 4, leading to generation of a transcript encoding the CD80 protein with a novel cytoplasmic domain. This transcript could be readily detected in the splenic RNA. The presence of the alternative cytoplasmic domain suggested that distinct signals might be transduced upon engagement of the external domains of murine CD80. Further, the murine CD80 gene was found to exist on band 5 of chromosome 16.
Murine CD80 Promoter And Transcriptional Regulation

Selvakumar et al., 1993 reported the identification of the transcription start site of the murine CD80 gene. They employed primer extension analysis and found the transcription start site 67 nucleotides upstream of the 5' end of the cDNA clone. An inspection of ~ 200 bp sequence around the transcription start site revealed that the proximal promoter was devoid of any canonical core promoter elements such as the TATA box, initiator sequence, CCAAT box or any conserved SP1 binding elements. Further analysis by Borriello et al., 1994 confirmed the previously reported translational start site but the transcription start site was found to exist 1742 nucleotides upstream of the start codon as analysed by 5' RACE. This corresponds to an extremely long 5' untranslated region of atleast 1700 bp, contiguous with the known exon 1 open reading frame.

Zhang et al., 1996 also reported the isolation, mapping and partial characterization of the murine CD80 promoter. They employed transient transfection assays in mouse L(tk') cells and suggested the proximal promoter to contain three positively regulated regions from -2597 to -1555, -130 to -110 and +25 to +269. The deletion of region located between -2597 and -1555, was found to account for reduction in luciferase activity by ~36%. Further, the results from a set of 5' deletion constructs of the promoter region showed that a 20 nt sequence between -130 to -110 accounted for almost 70% of the luciferase activity in their experiments. This region was found to contain a tandem repeat sequence 5' GTGTTCTAGTGTT 3'. Though, this sequence was suggested to be an authentic transcription factor-binding site, as shown by EMSA, the identity or any further information about the putative transcription factor binding this sequence was not provided in this study. Though, this study suggested that the CD80 gene could be potentially regulated by multiple
promoters, as suggested by existence of CD80 transcripts originating at different positions along the promoter, it primarily generated support for the initial observation made by Selvakumar et al., 1993. The sequences around the transcription start site reported by Selvakumar et al., 1993 were found to be important determinants of promoter activity and the deletion of the region in which the putative promoter for the transcript suggested by Borriello et al., 1994 existed had no effect on promoter strength. However, in the same study, an alternatively spliced isoform of CD80 transcript, lacking the sequences between -188 to +111, with respect to the transcription start site reported by Selvakumar et al., 1993 was detected.

An independent study by Antonia et al., 1996 also addressed the cloning and partial characterization of the mouse CD80 promoter. They cloned a 758 bp fragment spanning region between -692 and +67, relative to the transcription start site, described by Selvakumar et al., 1993 into p758LUC to generate the reporter constructs and showed that this region harbored the promoter activity. Further analysis employing 5' deletion mutants of the promoter showed that a region spanning nucleotides between -692 to -480 accounted for ~ 54% of the luciferase activity suggesting the existence of some key transcriptional control elements in this region. The sequence analysis of the promoter sequence revealed the presence of several sequences for transcription factor binding, though these were not characterized for any functional relevance.

Further, Jennings et al., 1999 suggested the presence of an enhancer element in the region between -597 and -588 of the murine CD80 promoter. EMSA analysis showed that this region could associate with a unique transcription factor from the purified nuclear extracts, which could not be competed out using commercially available AP2, Sp1 and NF-1 consensus oligonucleotides.
Review of Literature

Zhang et al., 2000 raised transgenic mice with three different CD80 promoter driven luciferase transgenes spanning the regions -3034/+269, -1073/+269 and -215/+269 bp, relative to the start site. Their results showed that the -3084/+269 transgene was sufficient to confer the tissue-specific expression of the CD80 gene in transgenic mice, as assessed by luciferase assay. When the promoter was deleted to -1073, the normal tissue specific expression was lost. They suggested a brain-specific element, the pentanucleotide AGGGA, to exist in the murine CD80 promoter between -1073 nt and -215 nt. However, this study did not provide any conclusive evidence that the pentanucleotide, AGGGA, was the brain specific activator for CD80 expression. None-the-less, it suggested that the region between -3084 to -1073 nt could potentially contain the tissue specific elements controlling transcription of this gene.

In a recently reported study from our lab (George et al., 2006), we have worked out the nucleosomal organization of the promoter region encompassing the transcription start site reported by Selvakumar et al., 1993 and shown that the upregulation of transcript from this start site, in response to a brief oxidative stress was critically dependent on the remodeling of a solitary nucleosome centered ~ 545 nucleotides upstream of the start site. The upregulation of transcript was controlled by the cooperativity between a NF-κB half site and a non-canonical SP1 binding element near the transcription start site. The activation process was facilitated by an induced DNA bending between these transcription factor-binding sites leading to the appropriate positioning of the basal transcriptional machinery.

SIGNAL AND TISSUE SPECIFIC REGULATION OF GENE EXPRESSION

The genome of an organism represents the 'blue print for life'. This master plan is realized through the process of gene expression resulting in the production of
several different kinds of RNA and a huge diversity of proteins. Metazoans consist of hundreds of different types of cells, each one of these containing the same genetic information but designed to perform a special role that contributes to overall functioning of organism. This remarkable diversity in cell specialization and the ability to respond to different physiological and environmental cues is achieved through the tightly controlled expression and regulation of a precise subset of genes. For this temporally and spatially correct expression of genes, metazoans have evolved a diversity of regulatory mechanisms scrutinizing the expression program at several levels. Amongst these diverse regulatory mechanisms, the regulation of transcription of a gene is the first stage at which control is exercised to regulate its expression. The fact that 5% of the human genome is predicted to encode for transcription factors alone, a class of transcription regulatory proteins, underscore the significance of regulation exercised at this stage in eukaryotic gene expression (Tupler et al., 2001).

The Template, In Vivo, For Transcription

The compartmentalization of DNA within the nucleus of the cell requires an extreme compaction of this highly charged polymer, over two meters in length. Faced with this problem of organization and compaction, the eukaryotic cell evolved a repetitive structure consisting of 146 base pairs of DNA wrapped around an octamer which is composed of two sets of four extremely basic histone proteins namely H2A, H2B, H3 and H4. This repetitive structure referred to as ‘the nucleosome’ is further folded and organized into higher order structure achieving the required compaction (Luger et al., 1997). This nucleosomal organization, which in principle is capable of impeding any cellular pathways involving nuclear DNA such as transcription, replication or repairs, plays a pivotal regulatory role in the nucleus. Several proteins must coordinate, spatially and temporally, in response to specific stimuli to make this
The Transcription Cycle

The synthesis of mature and functional messenger RNA for the protein coding genes by eukaryotic RNA pol II is a complex multistage process requiring the cooperative action of several different classes of proteins. This process - collectively referred to as the transcription cycle - proceeds via following stages: pre-initiation, initiation, elongation (further divided into the stages of promoter clearance, promoter proximal pausing and productive elongation) and termination. Interestingly, multiple steps in mRNA maturation, including pre-mRNA capping, splicing, 3’-end processing, surveillance and export are coordinated in a co-transcriptional manner.

Though, historically, the initial stages of transcription namely pre-initiation and initiation have received more attention, recent studies have generated compelling evidence that the regulation of transcription can be achieved by controlling the subsequent stages during the transcription cycle too.

INITIATION OF TRANSCRIPTION

RNA pol II catalyzes DNA-dependent synthesis of mRNA but in order to transcribe a gene, it needs to be recruited to the promoter and identify the transcription start site. Since, RNA polymerase II does not have the property of identifying the transcription initiation site on its own, it requires the help of several other proteins (collectively referred to as the eukaryotic transcription apparatus) to accomplish this task. This protein complement to RNA pol II consist of a common set of initiation factors (the basal transcription factors) and several other classes of individual proteins and proteins complexes. A series of interactions between these
protein factors and the target DNA sequences ensure the initiation of transcription by RNA pol II.

A breakthrough in the understanding of the mechanism of transcription initiation by RNA pol II followed the discovery that purified mammalian pol II would selectively initiate transcription from the template DNA when supplemented with the crude cell extract (Weil et al., 1979). This led to the fractionation and subsequent identification of the general transcription factors (GTFs), defined by their requirement for accurate initiation by Pol II in vitro (Orphanides et al., 1996).

The Basal Transcription Machinery

The basal transcription machinery can be defined as the factors, including RNA pol II, that are minimally essential for transcription in vitro from an isolated promoter DNA sequence. A short contiguous stretch of DNA extending ~ 35 bp upstream or downstream of the transcription start site, referred to as 'the core promoter', is the minimal region required to specify non-regulated or basal transcription and is identified by this basal transcription machinery. It directs an appropriate positioning of the RNA pol II over the transcription start site (Smale and Kadonaga, 2003). The complex consisting of the template DNA, the general transcription factors (GTFs) and RNA pol II is termed as 'the pre-initiation complex' or PIC. A minimum of five general transcription factors are found in PIC: TFIID, TFIIB, TFIIE, TFIIF and TFIIH aggregating to a molecular mass of nearly 2 megadaltons (Orphanides et al., 1996, Dvir et al., 2001).

The basal transcriptional apparatus is remarkably conserved amongst the eukaryotic organisms. The full complement of general transcription factors has been identified among disparate organisms, with one-to-one correspondence between each of the subunits. Extensive work carried out in past two and a half decades has lead to
a significant advance in our understanding of the initiation of transcription by RNA pol II and the general transcriptional machinery involved in this process.

RNA Polymerase II - The Centre Of Attraction

RNA pol II lies at the centre of transcription machinery, interacting with the general transcription factors in the PIC, breaking these interactions upon initiation, promoter clearance and associating with another set of factors during elongation and termination. Polymerase II subunits can be classified into three overlapping categories: subunits of the core domain having homologous counterparts in bacterial polymerase (Rbp1, 2, 3 and 11), subunits shared between all three nuclear polymerases (Rpb5, 6, 8, 10, and 12) and the subunits specific to pol II but not essential for transcription elongation (Rbp4, 7 and 9) (Hahn, 2004). A breakthrough in understanding the mechanism of transcription was achieved with the high-resolution structures of the RNA pol II enzyme (Cramer et al., 2001). The pol II structure and models for protein interactions have been determined for the 10- and 12-subunit enzymes without DNA (Armache et al., 2003; Bushnell et al., 2003), the 10-subunit enzyme in two different transcribing complexes (Gnatt et al., 2001; Westover et al., 2004), in complex with the general transcription factor TFIIB (Chen and Hahn, 2003; Bushnell et al., 2004) and the transcription elongation factor TFIIS (Kittenberger et al., 2003). Further, low-resolution EM structures have also been obtained for pol II binding to the mediator complex (Davis et al., 2002) as well as general transcription factor TFIIF (Chung et al., 2003). The structures of these multifactor complexes are beginning to reveal the assembly mechanism for the general transcription machinery and identifying the conformational changes in protein and DNA that must occur during transcription initiation.
The General Transcription Factors

TFIID

TFIID is a multi-protein complex composed of TBP (TATA binding protein) and about ten TBP-associated factors or TAFs (Tora, 2002). TFIID recognizes the core promoter through its components and provides a scaffold upon which rest of the transcriptional machinery can assemble. The molecular structure of TFIID complex, as studied by electron microscopy at 35 Å resolution (Andel et al., 1999, Brand et al., 1999a), resembles a lobular, horse-shoe shaped structure organized around a solvent-accessible groove that could accommodate a double-stranded DNA molecule.

TBP is a single polypeptide that identifies the ‘TATA’ sequence on the promoter. TBP sits astride the TATA sequence as a molecular ‘saddle,’ inducing a sharp bend in the DNA (Nikolov and Burley, 1997). Though, required for identification of the core promoter, TBP alone is incapable of supporting activator-dependent transcription (Hoey et al., 1990; Pugh and Tjian, 1990; Dynlacht et al., 1991). Apart from RNA pol II, TBP has also been shown to be required for initiation of transcription by polymerase I and III and hence was considered a universal transcription factor. However, recently several other homologues of TBP such as the TLF/TRF2 have been reported that bind conserved TATA sequences (Dantonel et al., 1999) but their physiological role is not well defined. Furthermore, the TBP free TFIID (TFTC) complexes have also been shown to exist in vivo (Brand et al., 1999a and b).

The TAF subunits of TFIID can also function to identify several core promoter elements (Albright et al., 2000) but, importantly several TAF subunits of TFIID interact with a number of transcription activator proteins in vivo like VP16, p53, SP1 (Sauer and Tjian, 1997; Albright and Tjian, 2000). Though, TAF dependence of
activation-driven transcription is not universal, some genes are indeed affected by the absence of TAFs in the assay system. Furthermore, the requirement for TAFs in these systems was shown to be largely specified by core promoter rather than distal regulatory elements (Green, 2000). Some TAFs like TAF$_{II}^{250}$ have also been shown to possess enzymatic activities (Dikstein et al., 1996; O'Brien and Tjian, 1998). Several TAFs exhibit sequence and structural similarity with histones (Burley and Roeder, 1996), which could imply similar rules of DNA identification between the two classes of proteins. Though, some studies supported this hypothesis (reviewed in Hoffmann et al., 1997) others argued against such possibilities (reviewed in Albright et al., 2000). None-the-less, there is enough consensus to suggest that such conserved structural motifs could facilitate compaction and tight protein-protein interactions in vivo during PIC formation (Birck et al., 1998).

**TFIIB**

TFIIB is a single polypeptide basal transcription factor that includes an N-terminal zinc binding domain (nTFIIB), a core domain that encompasses the C-terminal two thirds of the molecule (cTFIIB), and a phylogenetically conserved sequence that links the two domains (Woychik and Hampsey, 2002). It enters PIC subsequent to TFIID-DNA (core promoter) interaction and is a prerequisite for polymerase binding. TFIIB has also been shown to interact in a sequence specific manner to the promoter DNA (Lagrange et al., 1998). The asymmetric binding of TFIIB, as shown by X-ray crystal structures (Tsai and Sigler, 2000), accounts for the unidirectional binding of PIC and the direction of transcription. Further, the electron crystallography of a TFIIB-RNA Pol II complex, in combination with the structure of the DNA-TBP-TFIIB ternary complex, suggests a mechanism for start site selection: TFIIB appears to bridge the interaction between TBP and RNA Pol II such that the
DNA template follows a straight path from the TATA box, to position the start site in the active center of RNA Pol II (Leuther et al., 1996).

**TFIIF**

TFIIF was initially identified based on its physical association with RNA pol II and its requirement for accurate initiation. It performs a dual role in the assembly of PIC by binding to and strongly stabilizing DNA-TBP-TFIIB ternary complex and by recruiting TFIIE and TFIIH to the PIC (Dvir et al., 2001). TFIIF is a heterotetramer composed of two large (TFIIFα/RAP74) and two small (TFIIFβ/RAP30) subunits. Both the kinds of subunits contain multiple domain polypeptides that play distinct roles in initiation, elongation, and regulation of the Fcp1 CTD phosphatase activity. Site-specific protein DNA crosslinking experiments have defined contact points of TFIIF with template DNA. RAP30 interacts with template DNA on either side of the TATA box, whereas RAP74 interacts only downstream of TATA (Kim et al., 1997).

**TFIIE**

TFIIE is a heterotetramer containing two large (TFIIEα) and two small (TFIIEβ) subunits. The central core domain of the small subunit of TFIIE contains a helix-turn-helix motif that is predicted to bind DNA (Okuda et al., 2000). TFIIE affects late events in PIC assembly, including the recruitment of TFIIH and subsequent regulation of TFIIH catalytic activities. TFIIE and TFIIH are required for ATP-dependent formation of the open promoter complex prior to formation of the first phosphodiester bond. TFIIE, TFIIF and TFIIH also cooperate to suppress the promoter-proximal stalling, thereby facilitating early events in the transition of RNA Pol II to productive elongation (Dvir et al., 2001). TFIIE forms promoter contacts in and immediately downstream of the transcription bubble region, without altering protein-DNA interactions by RNA Pol II or the other GTFs (Kim et al., 2000).
TFIIH

TFIIH is the largest and most complex of all GTFs, consisting of nine subunits with a molecular mass comparable to that RNA pol II. TFIIH is the only GTF with defined enzymatic activities, including two ATP-dependent DNA helicases of opposite polarity (XBP and XPD) and a cyclin-dependent protein kinase (cdk7-cyclinH). TFIIH can be resolved into two subcomplexes: core-TFIIH and the cyclin-kinase complex. In addition to its role in transcription, core-TFIIH is also an essential component of the nucleotide excision repair (NER) machinery. Holo-TFIIH is essential for transcription, affecting steps before, during and immediately after initiation. Two different models for XPB-mediated promoter melting were put forth (Douziech et al., 2000; Kim et al., 2000) but whether XPB can function as a classical helicase to unwind the promoter is not clear.

TFIIA

TFIIA functions in part by binding to TBP and possibly stabilizing the TBP-DNA interactions (Buratowski et al., 1989). It interacts with numerous activators and may also function by antagonizing transcriptional repressors. It is also known to function by physically displacing or blocking several negative transcriptional regulators from the TFIID complex (Lee and Young, 2000).

DNA ELEMENTS DIRECTING THE RECRUITMENT OF THE BASAL MACHINERY: THE CORE PROMOTER ELEMENTS

Specific DNA elements within the core promoter bind the factors that nucleate the assembly of a functional preinitiation complex and integrate stimulatory and repressive signals from factors bound at distal sites. The widely recognized sequence motifs constituting the core promoter include the TATA box, the Initiator element (Inr), the downstream promoter element (DPE), the TFIIB recognition element (BRE)
and presumably others that remain to be identified. Although core promoter structure was thought to be invariant, a remarkable degree of diversity has become apparent (reviewed in Smale and Kadonaga, 2003) in recent years.

**TATA Box**

The TATA box was the first eukaryotic core promoter motif identified in the eukaryotic protein-coding genes (Breathnach and Chambon 1981). A binding site selection analysis identified the sequence 5'-TATATAAG-3' as the optimal TBP recognition sequence (Wong and Bateman, 1994). However, based on the functional TATA sequences from other reported studies, the consensus for TATA box can be generalized as TATAWAAR (degenerate nucleotides are designated according to the IUPAC code). Furthermore, it has been observed that a wide range of other non-conserved sequences can also function as TATA sequence and can interact with TBP (Hahn *et al.*, 1989; Singer *et al.*, 1990; Wobbe and Struhl, 1990; Zenzie-Gregory *et al.*, 1993; Patikoglou *et al.*, 1999). Though, majority of studies have suggested that TBP is the predominant TATA box-binding protein, identification of TBP related factors (TRFs) suggests that some other proteins can also identify it.

**Initiator Element**

The initiator element was first defined as a discrete core promoter element that can function independently of a TATA box, during an analysis of the lymphocyte-specific terminal nucleotydyl transferase gene promoter (Smale and Baltimore, 1989). It encompasses the transcription start site in a variety of promoters, both TATA-containing as well as TATA-less. Analysis of randomly generated and specifically targeted initiator mutants by *in vitro* transcription and transient transfection using human cell lines led to the identification of the functional consensus sequence (YYA⁺₁NWYY) in eukaryotes (Lo and Smale, 1996). A variety of protein factors
including TAFs of TFIID subunits, TFII-I and YY1 have been shown to interact with the eukaryotic initiator sequence.

**Downstream Promoter Element (DPE)**

DPE was identified as a downstream core promoter element that is required for the binding of purified TFIID to a subset of TATA-less promoters (reviewed in Butler and Kadonaga, 2002). DPE is conserved from *Drosophila* to humans and is typically, but not exclusively, found in TATA-less promoters. It acts in conjunction with the initiator element, and the core sequence of DPE is located at precisely +28 to +32 nucleotides relative to the A₁ in the initiator motif. The DPE consensus motif, RGWYV(T) (Kutach and Kadonaga, 2000), is identified by certain TAFs such as TAF₁40 and TAF₁60 (Tora, 2002).

**TFIIB Recognition Element**

BRE (TFIIB binding element) is located immediately upstream of some TATA boxes in some promoter (Lagrange, 1998). The BRE consensus, G/C-G/C-G/A-C-G-C-C, with 3’ C of BRE followed by 5’ T of TATA box was found conserved in ~12% of 315 TATA-containing promoters (Lagrange, 1998). The interaction of TFIIB with BRE was further illuminated by x-ray crystallography of a TFIIB-TBP-DNA complex (Tsai and Sigler, 2000). Interestingly, this motif is missing in yeast and plants, which suggests that the BRE may not contribute to gene regulation in these organisms.

**CpG Islands**

The start site proximal sequences of some promoters have been observed to be relatively GC rich and have over-represented CpG dinucleotides sequences. These generally lack TATA or DPE sequences but contain multiple sequences bound by SP1 and related transcription factors (Macleod *et al.*, 1994). SP1, through its ability to
interact with the components of basal machinery, mediates the assembly of PC on such promoters. In contrast to TATA or initiator containing promoters, transcripts from CpG containing promoter originate weakly at multiple positions.

### Motif Ten Element (MTE)

Motif ten was recently identified as a downstream, over-represented sequence in a computational analysis of *Drosophila* core promoters. It encompassed +18 to +29 relative to A+1 of the initiator element. Scanning mutational analysis of the motif ten sequence revealed that the sequence, CSARCSSAACGS, was capable of promoting transcription when located precisely at positions +18 to +27, relative to A+1 of the initiator element. The MTE promotes transcription by RNA pol II in conjunction with the initiator, but functions independently of the TATA-box or DPE (Lim *et al.*, 2004).

### Other Promoter Elements

In addition to the core promoter motifs described above, other DNA sequences in the transcription start site proximal regions have also been shown to contribute to transcriptional activity in a variety of genes. The human β-globin promoter was shown to possess a downstream promoter sequence from +10 to +45 termed as the downstream core element (DCE) (Lewis *et al.*, 2000). The human glial fibrillary acidic protein (gfa) gene was shown to contain a TATA box as well as a downstream sequence from +11 to +50 that is required for TFIID binding and transcriptional activity (Nakatani *et al*. 1990 a,b). Further, the analysis of TATA-less promoters with unclustered, multiple start sites led to the identification of a downstream motif termed multiple start site element downstream-MED-1 (Ince and Scotto, 1995).

### The Events During Initiation Of Transcription

In view of the structural data reported for several different forms of RNA pol II and general transcription factors (as described above), the chronology of events
occurring during the initiation of transcription can be summarized as follows: TBP bends TATA box DNA around the C-terminal domain of TFIIB. The N-terminal domain of TFIIB brings the complex to a point on the polymerase surface from which the DNA need only follow a straight path and by virtue of the conserved spacing from TATA box to transcription start site in pol II promoters, the start site would be juxtaposed with the active center. Entry of TFIIB and promoter DNA in the complex leads to binding of TFIIE, which in turn recruits TFIIF. An ATPase/helicase subunit of TFIIF torques the DNA, introducing negative superhelical tension in the region over the active center cleft. Consequent thermal unwinding is followed by capture of the non-template strand by TFIIF and the template strand descends to the active site, where it interacts with the B finger (a special structural loop formed by interaction of RNA pol II active site and general transcription factor TFIIB). Initiation and the synthesis of RNA greater than 10 residues in length leads to the displacement TFIIB (Bushnell et al., 2004) and the onset of next phase in transcription cycle, namely promoter clearance (Boeger et al., 2005).

THE GENE SPECIFIC REGULATORS OF TRANSCRIPTION

Apart from the basal machinery and RNA pol II subunits, several other classes of proteins and protein complexes are involved in the regulation of genes and tissue specific transcription. These proteins are required differentially for specific loci or the cell type in response to various extracellular and intracellular cues. These proteins serve as a critical link between the cellular signaling and transcriptional machinery.

DNA Binding (Sequence Specific) Activator Proteins

In eukaryotes, each protein coding gene has its own transcriptional program in terms of the signal it responds to and the transcriptional machinery required for the execution of that program. The end point of many, if not all, signal transduction
pathway is the covalent modifications of the sequence specific DNA binding proteins, the transcription factors, which bind the proximal and distal region of the genes (referred to as the 'enhancers' and 'silencer' element). A typical sequence-specific factor has a DNA binding module linked to one or more activation or repression modules as well as a multimerization and regulatory module. These proteins are targets of extensive post-translational modifications, many of which occur in response to a stimulus, and are important for the biological activity of these proteins at most of the occasions. Current evidence indicates that the sequence-specific factors function mainly by recruitment of transcriptional coactivators and corepressors to the DNA template via protein-protein-interactions, affecting directly the recruitment of basal machinery or effecting a regulatory conformation in the target DNA stretch. Further, the cofactors can then act directly and indirectly to regulate the activity of the RNA pol II transcriptional machinery at the core promoter (Ptashne and Gann, 1997).

Coactivators And Corepressors

This class of transcriptional regulators includes a diverse array of cellular factors and complexes that primarily function to transduce the signal between the DNA binding activator proteins and the general transcriptional apparatus utilizing diverse mechanisms. Coactivator functions can be broadly divided into two classes: adaptor function, directing activator-dependent recruitment of the general transcriptional apparatus, and enzymatic function that alters chromatin structure to facilitate transcription. The enzymatic activities can further be classified into covalent histone modifications such as acetylation/deacetylation, methylation/demethylation, ubiquitylation (Berger, 2001) and the ATP-dependent remodeling activities (Narlikar et al., 2002). Remarkably, these coactivator complexes share several common subunits (Naar et al., 2001).
Histone Acetyltransferases

A wealth of studies during past three decades has suggested that post-translational modifications of the histones modulate chromatin structure and hence transcription. In particular, an increase in acetylation of the N-terminal tails of the core histones occurs concomitant with the activation of transcription on a locus, first proposed 40 years ago (Allfrey et al., 1964). Subsequently, several proteins termed the ‘histone acetyltransferases’ (HATs) were identified having the ability to transfer acetyl groups from acetyl coenzyme A onto histone ε-amino groups of conserved lysine residues within the core histones. Though, several HATs may have overlapping function, based on their origin and function, HATs were divided into two groups: A-type HATs refer to the nuclear HATs which catalyze transcription-related acetylation events while B-type HATs refer to the cytoplasmic proteins which catalyse acetylation events linked to the transport of newly synthesized histones from the cytoplasm to the nucleus for deposition onto newly replicated DNA. A-type HATs include mammalian GCN5 (Yang et al., 1996), and its orthologs, P/CAF (Yang et al., 1996), CREB-binding protein, CBP (Ogryzko et al., 1996; Bannister and Kouzarides, 1996), p300 (Ogryzko et al., 1996), and TAF\textsubscript{II}250 (Mizzen et al., 1996; Dunphy et al., 2000) whereas the B-type HATs include HAT1 from yeast (Sobel et al., 1995). Further, HATs can also be classified based on the presence of acetylation-related structural motifs (reviewed in Roth et al., 2001). Though, some HATs have shown site and histone specificity, precise specificities of others need to be determined \textit{in vitro and in vivo} (Roth et al., 2001). Importantly, many other non-histone substrate for several HATs have also been shown recently, including transcription factors such as p65 (Kiernan et al., 2003), p53 (Gu et al., 1997), E2F1 (Martinez-Balbas et al., 2000), GATA-1 (Boyes et al., 1998) and HMGI(Y) (Munshi et al., 1998).
Histone Deacetylases

Histone deacetylase (HDAC) activity was also identified the same year when first histone acetyltransferase activity was defined (Taunton et al., 1996). Mammalian HDACs can be categorized into three classes: class I- HDAC 1- 3 and 8 (related to Rpd3), class II- HDAC 4-7, 9 and 10 (related to Hda1) and, the Sir2 family the activity of which is dependent on NAD (Yoshida et al., 2003). Similar to several HATs, most of the deacetylases are found in multiprotein complexes and the best characterized are the ones containing HDAC1 and HDAC2. These complexes include the Sin3, NuRD, and CoREST, which are recruited by various DNA-binding transcriptional regulators to DNA (Ayer, 1999; Knoepfler and Eisenman, 1999).

Histone Methyltransferases

The presence of different ratios of methylated species of lysine and arginine, depending on cell type or tissue source, was suggested almost three decades back (reviewed in Bannister et al., 2002). The histone lysine side chains may be mono-, di-, or trimethylated whereas the arginine side chain get mono- or dimethylated (symmetrically or assymmetrically) (Kouzarides, 2002). At present, there are 24 known sites of methylation on histones (17 of which are lysine and 7 arginine residues). Ever since the discovery of first histone methyltransferase (Rea et al., 2000), a considerable amount of information about the role of this covalent modification in regulating transcription has accumulated. Once thought of as a ‘permanent’ mark, several recent studies, reporting the reversibility of this modification, has shaken this dogma.

Methylation of the histones affect chromatin structure, primarily, by creating binding sites for the regulatory proteins such as HP1, CHD1, that contain specialized binding domains. This is in contrast to acetylation, which can also neutralize the charge on lysine residues of the histone tails. Most methylated lysine or arginine
residues have been identified on basis of their association with either heterochromatin or euchromatin. It is however, becoming clear that not all heterochromatin or euchromatin are same with respect to methylation status of histones that it contains (reviewed in Bannister and Kouzarides, 2005).

**Histone Demethylases**

Histone demethylases constitute one of the most recently acknowledged classes of enzymes involved in the process of transcription. Up until recently, demethylation of histones was a remote possibility, if not completely unfeasible. Recently, an enzyme lysine-specific demethylase (LSD1) was identified, which could demethylate a specific lysine (K4) residue of H3 histone using amine oxidation reaction (Shi et al., 2004). But surprisingly, this enzyme lacked any obvious homologues that could carry out demethylation of other lysine residues. However, alternative pathways for the reversal of arginine methylation have been proposed. Peptidyl arginine deiminase 4 (PADI4), converts unmodified and mono-methylated (but not di-methylated arginine) to citrulline at specific sites on the tails of H3 and H4 histones (Cuthbert et al., 2004; Wang et al., 2004). This deiminating enzyme, too, lacks obvious orthologues in lower eukaryotes and appears to be restricted to higher mammals only.

**Mediators**

Mediators are evolutionary conserved and ubiquitously expressed multiprotein complexes that play an important role in activation and repression of RNA pol II transcription. This family of transcriptional regulators lack an intrinsic DNA binding ability and primarily function as ‘adaptors’ that bridge RNA pol II (or the basal machinery) and myriads of DNA binding regulatory proteins (activators) and transduce both positive and negative signal through distinct mechanisms. Mediator
complex was first identified and purified to near homogeneity in yeast, by virtue of its ability to promote activator-dependent transcription when purified RNA pol II and GTFs were used (Kim et al., 1994). Further, this complex was shown to consist of around 20 polypeptides. In mammals, several yeast mediator-like complexes were subsequently identified and characterized, namely TRAP/SMAC (Fondell et al., 1999, Gu et al., 1999) ARC-L (Naar et al., 1999), DRIP (Rachez et al., 1999), PC2 (Malik et al., 2000), CRSP (Taatjes et al., 2002), mouse med (Jiang et al., 1998) and rat med (Brower et al., 2002). Though functionally distinct, these complexes include several orthologs of *Saccharomyces cerevisiae* mediator subunits, suggesting a significant evolutionary conservation. To account for such a functional divergence and evolutionary conservation, it has been proposed that a minimal module in these complexes could serve as a 'core mediator' enabling the cell to custom-design distinct mediator complexes in response to a specific signal or the associating transcription factor (Rachez and Freedman, 2001). Interestingly, mammalian mediators have actually been suggested as the possible end-points for signal transduction pathways (Jiang et al., 1998).

**ATP-Dependent Remodeling Complexes**

The term 'ATP-dependent remodeling' subsumes a large number of ATP-dependent changes of canonical nucleosome structure brought about by the dedicated nuclear enzymes, which are usually part of large, multifactorial complexes termed the 'remodeling complexes' (Becker and Horz, 2002). These complexes endow nucleosome with the dynamic properties that implement the state of 'plasticity' and 'fluidity' to facilitate processes such as transcription, repair and recombination. ATP-dependent changes in a chromatin substrate, include disruption of histone-DNA contacts with nucleosomes, movement of histone octamers in *cis* and in *trans*, loss of
negative supercoils from circular mini-chromosomes, and increased accessibility of nucleosomal DNA to transcription factors and restriction endonucleases (Peterson and Workman, 2000). All remodeling complexes belong to SNF2 family of ATPases (Lusser and Kadonaga, 2003) and can be divided into three main classes based on the identity of their catalytic ATPase subunit namely, SWI2/SNF2 family, ISWI family and the Mi-2 family (Narlikar et al., 2002). The ATPase subunits, alone, within a class show homology and associate with additional subunits to form distinct remodeling complexes. Remodeling complexes formed with all three family of ATPase are capable of increasing accessibility of nucleosome bound DNA by translational repositioning of the nucleosomes, apart from other characteristic mechanisms (Narliker et al., 2002). The ability of these complexes to alter the structure of chromatin has been studied using a variety of different assays indicating that these complexes have different substrates requirement and might act via different mechanisms.

The distinguishing feature of SWI2/SNF2 family ATPases is the presence of a bromodomain (Horn and Peterson, 2001; Martens and Winston, 2003). In humans, remodeling complexes are formed around two ATPase subunits, hBRM and hBRG, of this family. The members of ISWI family of remodellers are characterized by C-terminal SANT-like domain (Grune et al., 2003). SWI2/SNF2 and ISWI are two best-characterized families of remodeling complexes with their substrate requirement and the remodeled species generated as a consequence of their enzymatic action well established (reviewed in Narliker et al., 2002). Another family of remodeling enzymes, the Mi-2 family (also referred to as the CHD family, Eberharter and Becker, 2004), is characterized by presence of a chromodomain (Delmas et al., 1993; Kelley et al., 1999).
TRANSCRIPTION ELONGATION

Following recruitment of the PIC and formation of an open complex between RNA pol II and the DNA template, the formation of first phosphodiester bond between the two initiating ribonucleoside triphosphates marks the onset of elongation phase of transcription cycle (Goodrich and Tjian, 1994). Transcription elongation can be divided into three phases: Promoter escape (Stage I), promoter-proximal pausing (Stage II) and productive elongation (Stage III). Each of these stages is defined by marked differences in the stability and behaviour of the transcription complex, as well as a distinct repertoire of factors that associate with it (Saunders et al., 2006). The factors which influence transcript elongation by RNA pol II can be broadly classified into two groups: Active elongation factors, those which can affect the catalytic activity of RNA pol II and Passive Factors, those which associate with the moving polymerase but do not affect the rate of catalysis (Sims et al., 2004).

Elongation Stage 1 (Promoter Clearance (Escape))

Promoter escape describes the earliest steps in this maturation process, during which the polymerase breaks its contacts with the promoter bound sequence elements and at least some promoter-bound factors and simultaneously tightens its grip on the nascent RNA. During this stage PIC partially diassembles but a subset of GTFs remain at the promoter providing a scaffold for the subsequent initiation complexes (Zawel et al., 1995, Yudlovsky et al., 2000). Each of these steps depend on the intrinsic interactions of the polymerase with the nascent RNA and the template DNA sequences, and are vulnerable to regulation by extrinsic factors (Dvir, 2001). Basal transcription factor TFIIH plays an integral role in facilitating promoter clearance, in part by preventing premature arrest (Goodrich and Tjian, 1994; Dvir et al., 1997). Furthermore, recent cocrystal of RNA pol II-TFIIIB suggested that TFIIB might
impede the exiting path of the newly formed RNA (Bushnell et al., 2004). The authors suggested that continued presence of TFIIB might result in abortive initiation and removal of TFIIB might be essential for promoter clearance. Promoter escape, rate-limiting for the transcription elongation in vitro, begins with transcription initiation (the transcription complex at this stage is referred to as the initially transcribing complex (ITC)) and is considered complete when the nascent RNA associates stably with the transcription complex. The transcription complex then is referred to as the early elongation complex (EEC).

During early transcript synthesis, the ITC undergoes abortive initiation. It has been suggested that RNA length is a crucial determinant of promoter escape through its interaction with pol II. In eukaryotes, this abortive initiation is markedly reduced on the addition of the fourth nucleotide, though promoter escape is still not complete by then (Holstege et al., 1997; Kugel and Goodrich., 2000). This transition to metastable stage is referred to as ‘escape commitment’. The rate-limiting step of promoter escape that coincides with the transition to the EEC occurs after the addition of the eighth nucleotide and when the pol II active site is translocated to ninth position (Hieb et al., 2006). This stage is marked by the end of abortive transcript release (at +10/+11 nucleotide), end of the requirement of TFIIH and ATP (at +8/+10 nucleotide) and a sudden collapse of an extended transcription bubble (reviewed in Saunders et al., 2006). Some studies have also suggested that rate of promoter escape is also sequence-dependent and a complex that contains weak DNA-RNA hybrid is slow to complete promoter escape (Weaver et al., 2005) but further studies are required to better understand the role of initially transcribed sequences and the core promoter elements in the various steps of promoter escape.
The early elongation complexes retain a measurable tendency to undergo lateral slippage till about +23 nucleotides (Pal and Luse, 2003) and are prone to backtracking and arrest until +30 nucleotides. EECs arrested in this stage can be reverted to competency by the transcription factor TFIIS, which stimulates the intrinsic RNA-cleavage activity of RNA pol II so as to generate a 3’OH properly aligned with the RNA pol II active site (Cramer, 2004). The relationship between distance of EEC from the transcription start site and formation of mature elongation complex is not clear but it has been hypothesized that the emerging RNA interacts with the polymerase and/or itself in some way that can effect elongation potential (Ujvari et al., 2002). Consistent with this, RBP7 subunit of RNA pol II has been shown to interact with the nascent emerging RNA (Ujvari and Luse, 2006) through its oligonucleotide binding domain (Orlicky et al., 2001). RBP7 probably directs the nascent transcript towards the CTD of pol II and the associated CTD-dependent-RNA-processing enzymes (Ujvari and Luse, 2006). The nascent RNA can also interact with CTD to suppress transcription coupled 3’end processing (Kaneko and Manley, 2005).

**Elongation Stage 2 (Promoter-Proximal Pausing)**

Promoter-proximal pausing (or transcriptional stalling) is a phenomenon whereby RNA pol II pauses in the 5’ region of the transcription unit and only progresses efficiently into productive elongation on stimulation by appropriate signals. Escape from pausing is a rate-limiting step and constitutes an important step in the regulation of pol II elongation *in vivo*. It functions as a check point before committing into productive elongation phase and represents a dynamic level of regulation capable of altering the transcriptional output *in vivo*. For several years this stage of regulation was under-appreciated. However, the promoter-proximal pausing
has now been shown to be a widespread phenomenon as suggested by analysis of several genes in isolation (Lis, 1998) and a recent genome-wide study (Kim et al., 2005). It has been suggested that promoter proximal pausing might facilitate correct capping and a correctly capped nascent RNA might be a prerequisite for escape from pausing (Pei et al., 2003).

Transcriptional pausing is dependent on the presence or absence of some ancillary factors. NELF and DSIF are two such factors which cooperate to repress transcription elongation and result in promoter proximal pausing. DSIF (DRB-sensitivity inducing factor), identified based on its ability to confer DRB sensitivity in a partially purified transcription reaction, (Wada et al., 1998) consists of the elongation factors Spt4 and Spt5 which are conserved from yeast to humans (hartzog et al., 1998). NELF comprise four subunits, NELF-A, B, C/D and E and is conserved in mammals. DSIF, in absence of NELF, has been suggested to have a positive role in transcript elongation (Yamaguchi et al., 1998; Rondon et al., 2004). DSIF and NELF have been reported to be present in unstimulated Drosophila melanogaster heat shock genes in vivo (Andrulis et al., 2000) and implicated for promoter proximal pausing on this gene.

Stage 3 (Productive Elongation)

Release from pausing

Several factors control entry of RNA pol II into the productive elongation phase. P-TEFb (Positive transcription-elongation factor b) which consists of a protein kinase CDK9 (cyclin dependent protein kinase 9) and a cyclin partner either cyclin T1, T2a, T2b or K (Peng et al., 1998 a,b) has been shown to relieve DSIF and NELF mediated repression of elongation on several genes. It is dependent on its kinase activity for its effect and hence sensitive to a kinase inhibitor DRB. P-TEFb
phosphorylates DSIF, NELF and CTD of RNA pol II and is required for the production of long transcripts \textit{in vitro} and \textit{in vivo} for several genes. It is localized with active genes \textit{in vivo} (Lis \textit{et al.}, 2000) and its depletion causes defects in gene expression (Ni \textit{et al.}, 2004; Shim \textit{et al.}, 2002). Several gene specific regulators have been shown to interact with P-TEFb \textit{in vivo} (Garriga and Grana, 2004). Recently, a bromodomain containing protein Brd4 which associates with active P-TEFb complex has been shown to be responsible for the recruitment of P-TEFb through its interaction with acetylated histones in the nucleus.

\textbf{RNA polymerase II catalysis rate and processivity}

A class of factors stimulate the rate of elongation and read-through of intrinsic arrest and pause sites. TFIIF, ELL (Eleven-nineteen lysine rich in leukemia) and Elongin have such effects. While these factors function at different steps, they stimulate elongation, at least in part, by suppressing transient pausing.

\textbf{TFIIF}

The general transcription factor TFIIF consist of a heterodimer of RNA pol II associating protein 30 (RAP30) and RAP74. It plays an integral role in recruiting RNA pol II to PIC. Though, more-well characterized in context of PIC formation, several lines of evidences indicate that it plays an important role in elongation too. TFIIF activity has been shown to be important for suppressing abortive initiation and acts in concert with TFIIFH to facilitate promoter escape (Yan \textit{et al.}, 1999). It has also been shown to be important for release of pol II from stalled site (Zawel \textit{et al.}, 1995). It has been shown to associate with RNA pol II in the promoter region only and not with the body of the gene, implying that it does not travel with RNA pol II (Krogan \textit{et al.}, 2002; Pokholok \textit{et al.}, 2002). TFIIF activity is regulated by phosphorylation. TAF\textsubscript{II}250 (TAF I) and TFIIF phosphorylate TFIIF RAP74 subunit (Ohkuma and
Roeder, 1994; Dikstein et al., 1996) though, RAP74 itself also has been suggested to have an autophosphorylation activity (Rossignol et al., 1999).

**Elongins**

The elongin complex consists of three subunits: the transcriptionally active elongin A, the regulatory elongin B and C. Unlike TFIIF, elongins affect transcription elongation through the body of the genes (Moreland et al., 1998). In vitro elongins exert their effect on transcripts longer than 8-9 nucleotides and maximal effects require dissociation of TFIIF from PIC. Based on protein homology, a role for elongins in ubiquitylation of RNA pol II and other cofactors has also been speculated (Shilatifard et al., 2003). There are several elongin-related factors reported recently, which are also implicated in stimulation of elongation rates (Aso et al., 2000; Yamazaki et al., 2002).

**ELL (Eleven-Nineteen Lysine Rich In Leukemia)**

ELL is functionally analogous to TFIIF and elongins due to its ability of suppressing transcriptional pause and stimulation of the rate of transcript elongation. The ELL gene was identified as a chromosomal translocation partner to MLL gene in acute myeloid leukemia (Thirman et al., 1994). However, the major determinant of acute myeloid leukemia seems to be dysfunctional MLL rather than loss of ELL, because of high number of MLL translocations to other genes (Ayton and Cleary, 2001). While in humans, ELL constitutes a family of related proteins (Shilatifard et al., 1997a), Drosophila ELL was identified as the sole representative member. ELL, similar to elongins, functions to stimulate elongation throughout the body of the gene but the mutations in ELL seem to have pronounced effects, preferentially, on the transcription of longer genes (Eissenberg et al., 2002; Shilatifard, 2004). ELL can also
have negative effects on transcription through uncharacterized mechanism (Shilatifard et al., 1997b).

**Role of CTD in regulating transcript elongation by RNA polymerase II**

The carboxyl terminal domain (CTD) of Rpb1, the largest subunit of RNA pol II, itself has also been shown to represent a key player in regulating entry into the productive elongation phase. CTD consists of multiple repeats of the heptapeptide sequence YSPTSPS, and hyperphosphorylation of the serine residues in this domain of Rpb1 is a hallmark of the productive elongation phase of the transcription cycle (Dahmus, 1996). Ser 2 and Ser 5 of CTD have been identified as the major modification sites. Although, several kinases have been reported to phosphorylate CTD (Palancade and Bensaude, 2003; Prelich, 2002), CDK7, CDK8 and CDK9 are of particular importance. CDK7, a kinase subunit of TFIIH, specifically phosphorylates Ser 5 residues of CTD while CDK9, a kinase subunit of P-TEFb, phosphorylates Ser 2 of CTD. In contrast, CDK8, which associates with Srb/mediator complex, is an inhibitory kinase that phosphorylates CDK7 and also CTD on Ser 5 (Cho et al., 1998). Recent studies have revealed that these two modifications can be detected differentially in different regions of a gene and are associated with distinct stage of the transcription cycle (Cheng and Sharp, 2003; O'Brien et al., 1994; Komarnitsky et al., 2000).

**Factors regulating elongation through histone tail modifications**

**Histone acetylation**

Histone acetylation is required for productive transcription but evidences for role of histone acetylation during elongation stage of transcription are limited (Walia et al., 1998). One of the links between elongation and specific histone acetylation is suggested by interaction of a component of NuA3 histone H3 acetyltransferase
complex with Spt16 subunit of FACT (John et al., 2000). Another acetyltransferase complex, elongator, has also been implicated in regulation of elongation, though there are conflicting reports about its role (reviewed in Sims et al., 2004).

**Histone methylation**

Methylation of several histone tail residues has been linked to the process of elongation. Members of Set 1 family of proteins methylate histone H3 at K4. Set 1 family includes *Drosophila melanogaster* trithorax, and human MLL1, MLL2 and Set1 methyltransferases. H3K4 trimethylation is concentrated at the 5' end of the genes and correlates strongly with active transcription (Pokholok et al., 2005; Schneider et al., 2004). Set 1 associates preferentially, with ser 5 phosphorylated form of CTD, which is closely associated with early transcriptional events (Ng et al., 2003). Another family of methyltransferases, Set 2 (which includes mammalian NSD1) methylate histone H3-K36, detected all across the body of the gene (Dhillon et al., 2005), and is strongly correlated with elongation (Krogan et al., 2003b; Schaft et al., 2003). Set 2 interacts with dual, serine 2 and serine 5, phosphorylated form of RNA pol II and this interaction is important for it's functioning (Kizer et al., 2005). The recruitment of Set 2 methyltransferase on the coding region is dependent on PAF elongation complex (Krogan et al., 2003a). The PAF elongation complex is conserved from yeast to mammals and apart from recruiting Set2, is also required for other modification during elongation phase (Sims et al., 2004).

**Other histone modifications during transcript elongation**

Several other covalent modifications of histones have been implicated in generation of full length transcript in eukaryotes. Monoubiquitylation of histone H2B at K120 stimulates the production of full length transcript *in vitro* (Pavri et al., 2006). Further, monoubiquitylated H2B (K-120 in mammals) and the corresponding
ubiquitylation enzyme complex are present at promoters and coding regions of active
genes (Zhu et al., 2005). Phosphorylation of H3 serine 10 has been localized to active
heat shock genes and a critical role for protein phosphatase 2A (PP2A) in
dephosphorylating and regulating global transcription in *Saccharomyces Cerevisiae*
(Nowak et al., 2003).

**Histone Chaperones**

DNA compaction and protection are two essential properties of chromatin but
a disruption of this structure occurs concomitant with catalysis by DNA and RNA
polymerases. It has been reported that elongation by RNA pol II on chromatin
templates is accompanied by loss of H2A-H2B dimer from the nucleosomal particles
(Kireeva et al., 2002). Furthermore, since the histones themselves contain pertinent
information relevant to gene expression in the form of posttranslational modifications,
the eukaryotic cell has evolved a system to maintain the required integrity of
chromatin in wake of such events. FACT complex and Spt6 are two important
components of this system in eukaryotes. The FACT (Facilitates chromatin
transcription) complex is highly conserved among eukaryotes and functions after
transcription initiation to allow pol II to transcribe in chromatin context and functions
independent of TFIIS and ATP (Belotserkovskaya et al., 2004). FACT is a
heterodimer composed of hSpt16 and the HMG-box protein SSRP1 (structural
specific recognition protein 1). Spt6, originally identified in a genetic screen, has been
shown to co-localize with actively transcribed region and its recruitment is
comparable to FACT and RNA pol II itself on the coding regions. Mutations in Spt6
lead to initiation of transcription from cryptic sites within coding regions (Kaplan et
al., 2003). Both FACT and Spt6 have been shown to have nucleosome deposition
activities *in vitro* (Belotserkovskaya et al., 2003; Bortvin and Winston, 1996).
Though, besides FACT and Spt6 other histone chaperones and ATP-dependent chromatin-remodeling complexes have been suggested to facilitate replication-independent exchange of histones, compelling evidence suggest that FACT and Spt6 are central to the maintenance of chromatin structure after an elongating complex has traversed a nucleosome.

Factors regulating elongation through ATP-dependent chromatin remodeling

All three classes of ATP-dependent chromatin remodeling complexes have been shown to play a role during elongation stage of transcription. Heat shock factor 1 (HSF1) has been reported to alleviate the negative effects of chromatin on transcription through recruitment of SWI/SNF activity to human Hsp70 gene (Brown et al., 1996). Genetic interactions between SWI/SNF and TFIIS, reported in yeast, and large protein complexes containing RNA pol II and SWI/SNF from yeast and human cells (Wilson et al., 1996; Neish et al., 1998) further substantiate the role of SWI/SNF family of remodeling complexes during elongation (Davie and Kane, 2000). CHD1 (chromo-ATPase/helicase-DNA-binding domain), an ATPase domain containing protein complex, functions both during elongation and termination (Tran et al., 2000). It interacts with both Set 2 histone methyltransferase and ISWI family members, both implicated in elongation (Krogan et al., 2003b; Tsukiyama et al., 1999).

Factors coordinating the downstream events

Transcription is a highly integrated process that is tightly coupled to mRNA maturation, surveillance and export (Maniatis and Reed, 2002; Orphanides and Reinberg, 2002). Addition of 5' cap, splicing, 3' end processing occurs efficiently cotranscriptionally (Proudfoot et al., 2002). An integral component in the coordination of these events is the CTD of RNA pol II, which serves as a platform for many of these factors required for mRNA maturation. Specific regions of CTD have
been shown to independently stimulate 5’ capping, splicing and 3’ end processing (Fong and Bentley, 2001). Furthermore, structural studies with CTD of pol II have suggested that the phosphorylation status of CTD, through the regulation of the structure it assumes, regulates these events (Meinhart and Cramer, 2004). The capping enzyme is recruited in part through a direct binding to the phosphorylated form of CTD (Yue et al., 1997). A direct interaction between CTD and the cleavage-polyadenylation factors, CPSF and CstF, has also been observed (McCracken et al., 1997), suggesting a link between splicing and transcription. Remarkably, the rate of elongation has been shown to regulate alternative splicing, implying that elongation critically controls selection of the exons (de la Mata et al., 2003). Furthermore, independent studies have shown that loss of serine 2 phosphorylation of CTD lead to 3’ end processing defects and decrease in steady state mRNA levels (Ahn et al., 2004, Ni et al., 2004). The transcription elongation factor Spt5 and Spt6 interact physically with the exosome providing a link between RNA surveillance and elongation. Transcript elongation has also been closely linked to transcription-dependent recombination (Prado et al., 1997) and mRNA export (Fan et al., 2001). These observations, amongst several other accumulating evidences, underscore the significance of CTD in cotranscriptional RNA processing. It has been speculated that a ‘CTD code,’ comparable to the histone code, could operate for coordination of these downstream events (Buratowski, 2003).