# 1. INTRODUCTION

Stable and uniform transgene expression over a long period of time is a prerequisite for many biological studies relying on the manipulation of genomes by introduction of additional genetic information. However, transgenes stably integrated into a cell's genome are frequently subject to progressive decrease of their expression over time, and eventually become silenced. Moreover, when the expression is assayed on the single cell level, populations of genetically identical cells often show a heterogeneous expression pattern, also known as mosaicism: some cells take on positive transgenic expression, whereas others are silenced. These phenomena have long been noticed and created great attention in both cell biology as well as the transgenic mouse field. This has also been a serious limitation to gene therapy research. There is increasing evidence that gene silencing and mosaicism are largely caused by epigenetic mechanisms acting on transgenes and foreign DNA sequences. In the following, I will introduce several biological mechanisms that have been proposed to partly explain these observations.

## 1.1. DNA methylation and chromatin modification

### 1.1.1 DNA methylation and gene silencing

#### 1.1.1.1 The properties of DNA methylation

In mammals, DNA methylation is an epigenetic mark on genomic DNA achieved by the addition of methyl groups to cytosine bases, with the methyl group projecting into the major groove. Methylation of cytosines occurs predominantly at CpG dinucleotides, but data also showed that mammalian cells possess the ability to methylate CpNpG sites as well in transfected plasmid DNA sequences (Clark et al., 1995). In vertebrates, 60-90% of all CpGs are methylated, leaving a minor part of the genome methylation free (Ng and Bird, 1999). Many of the remaining non-methylated CpGs (~15% of all CpGs in human DNA) are found in CpG islands (Antequera and Bird, 1993). CpG islands are

clusters of CpGs and first defined by Gardiner-Garden and Frommer in 1987 as a region of more than 200 base pairs (bp) long, with a high-GC content, and an observed/expected ratio for the occurrence of CpG > 0.6 (Gardiner-Garden and Frommer, 1987). The evolved CpG island criteria exclude a substantial number of small exonic regions and repetitive parasitic DNAs (Jones and Takai, 2001; Takai and Jones, 2002). Interestingly, CpG islands are found mainly in the 5' regions of housekeeping genes as well as some other tissue specifically expressed genes and usually extend from the promoter region into the first exon and sometimes into the first intron (Gardiner-Garden and Frommer, 1987; Larsen et al., 1992; Takai and Jones, 2002). Analysis of the distribution of DNA methylation in normal adult somatic tissues reveals that, whereas the majority of cytosines in the context of the CpG dinucleotide are methylated, CpG islands typically remain methylation-free. However, there are certain situations where these sequences become methylated and as part of gene regulation (Bird et al., 1985). The majority of CpG islands on the inactivated X-chromosome are methylated (Goto and Monk, 1998), as are the CpG islands of many non-essential genes in long term culture and immortal cell lines (Jones et al., 1990; Antequera et al., 1990).

#### 1.1.1.2 DNA methyltransferases (DNMTs)

The process of DNA methylation in mammals is carried out by at least three catalytically active DNA methyltransferases (DNMTs). The mammalian family of DNMTs consists of five known members, DNMT1, DNMT2, DNMT3A, DNMT3B, and DNMT3L (Bestor, 2000b). These five proteins are placed in the DNMT family based on sequence homology, however, only DNMT1, DNMT3A, and DNMT3B have been shown to exhibit catalytic activity (Okano et al., 1998; Leonhardt et al., 1999; Hsieh, 1999; Okano et al., 1999; Bestor, 2000b; Hata et al., 2002). Of these three enzymes, DNMT1 functions as the maintenance DNMT to ensure that the DNA methylation patterns are faithfully transmitted to the newly synthesized DNA strand following replication (Bestor, 2000b; Yokochi and Robertson, 2002). For this reason, the DNMT1 is found in an enzyme complex located at the replication fork together with proliferating cellular nuclear antigen (PCNA) (Chuang et al., 1997). In contrast, DNMT3A and DNMT3B are believed to be the primary *de novo* methyltransferases. They establish new DNA methylation patterns (Okano et al., 1999). Both enzymes are probably

responsible for the establishment of methylation patterns during embryogenesis because the process of *de novo* methylation is especially critical during embryonic development and these two enzymes are highly expressed in undifferentiated embryonic stem cells and at low levels in somatic tissues (Okano et al., 1999). Additionally, the *de novo* DNMTs are thought to be responsible for the methylation of repetitive elements (Jones and Takai, 2001).

#### 1.1.1.3 DNA methylation and gene silencing

DNA methylation is an epigenetic process involved in controlling many cellular functions, such as gene expression, genome stability, X chromosome inactivation, and chromatin structure (Jones and Takai, 2001; Jones and Baylin, 2002; Geiman and Robertson, 2002). A large body of evidence accumulated over the years suggests that DNA methylation negatively interferences with transcriptional activity. A strong correlation between DNA methylation and gene inactivity was documented about 25 years ago (Razin and Riggs, 1980). The loss of expression from artificially methylated gene constructs upon transfection into tissue culture cells strongly supported this hypothesis (Stein et al., 1982). Furthermore, the administration of a DNA demethylating agent, 5-azacytidine (5-azaC), was able to reactivate the previously methylated, silenced, endogenous genes and retroviruses (Groudine et al., 1981).

The most direct mechanism by which DNA methylation could interfere with transcription would be to prevent the binding of the basal transcriptional machinery and of ubiquitous transcription factors to promoters. However, this is not an universal mechanism because some heavily methylated genes can be transcribed effectively as naked DNA templates (Tate and Bird, 1993; Kass et al., 1997b) and the binding of Sp1 is indifferent to methylation state (Smith et al., 2004). The alternative possibility is that specific transcriptional repressors recognize methyl-CpGs and, either independently or together with other components of the chromatin, turn off transcription. The current concept is that the DNA methylation-dependent repressors work in a chromatin context because DNA methylation concurs with a compacted chromatin structure. Thus, transcriptional suppression is independent of the DNA sequence itself, which makes this mechanism more plausible.

### 1.1.2 Chromatin modification and gene silencing

#### 1.1.2.1 Chromatin structure

The building block of chromatin is the nucleosome. This particle consists of a histone octamer comprised of a histone H3 and H4 tetramer and two histone H2A and H2B dimmers. This is wrapped twice by 147 bp of DNA to form the nucleosome core (Fig. 1) (Kornberg, 1974; Finch et al., 1977). A variable length of DNA, associated with histone H1, acts as a linker to connect the nucleosome cores. Histones are basic proteins that consist of a globular domain and a histone tail. The histone tails protrude out of the nucleosome, and are targets for covalent post-translational modifications.

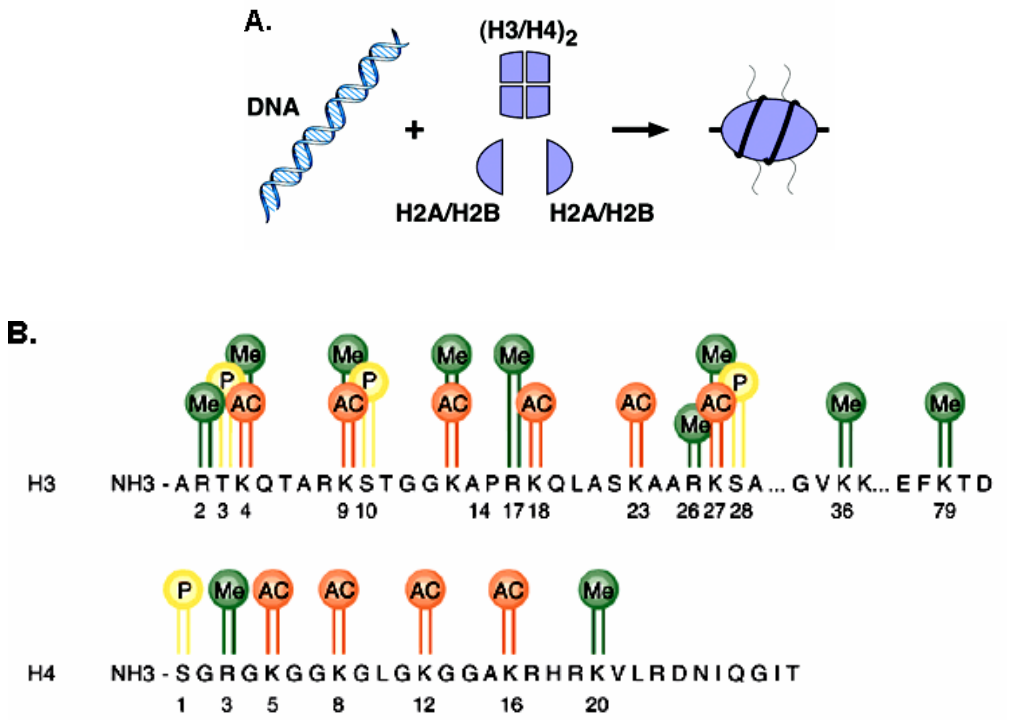


**Fig. 1. Diagrammatic representation of the nucleosome and potential amino acid residues of modification on histone H3 and H4.**

(A) Nucleosome. The building block of chromatin consists of 147 base pairs of DNA wrapped twice around the histone octamer complex with the histone tails extending out. Adapted from Grewal and Moazed, Science 2003. (B) Sites of posttranslational modifications on histone H3 and H4 amino termini. Many modification patterns have been closely linked to unique biological outcomes. Me: methylation, Ac: Acetylation, P: Phosphorylation. Adapted from Jaskelioff and Peterson, Nat. Cell Biol. 2003.

Chromatin is generally grouped into two distinguished types according to cytological staining, heterochromatin and euchromatin. Heterochromatin is condensed and dark stained chromosome regions found throughout the cell cycles. Heterochromatin appears most frequently, but not exclusively, at the centromere and telomeres of the chromosome. Heterochromatin has long been regarded as sites of relatively gene-poor and inactive genes. Recent molecular studies have shown that most of the DNA in heterochromatin is highly repeated DNA that is late replicating and very seldom transcribed (Weiler and Wakimoto, 1995; Henikoff, 2000). However, not all inactive genes and non-transcribed regions of DNA are visible as heterochromatin. By contrast, the light stained and less condensed chromosome portions are called euchromatin. Euchromatin is in general transcriptionally permissive.

#### 1.1.2.2 Chromatin modification and gene expression

Gene transcription in mammalian cells does not occur on naked DNA, but instead occurs in the context of chromatin. Therefore, chromatin structure plays an important regulatory role, and modifications to the chromatin and the histones themselves impact gene expression. Although histone proteins themselves come in generic or specialized forms (Wolffe and Pruss, 1996), a diversity of posttranslational modifications, e.g. acetylation, phosphorylation, and methylation, of the histone tail domains allow regulatable contacts with the underlying DNA (Jenuwein and Allis, 2001). The enzymes transducing these modifications are highly specific for particular amino acid positions (Strahl et al., 1999; Jasencakova et al., 2003). Among these modifications, de/acetylation and methylation of lysine residues on histone H3 and H4 received the lion's share of attention. Acetylation of lysine 9, 14, and 27 on H3 is associated with a loose chromatin structure and an active transcription state, as does acetylation of arginine 3 and lysine 16 on H4. On the contrary, lysine 9 and 27 deacetylation and/or methylation, and lysine 14 deacetylation are associated with gene suppression (Rea et al., 2000; Muller et al., 2002; Cao et al., 2002). In addition, methylation of lysine 4, 36 and 79 on H3 leads to heterochromatin formation and gene silencing (Jaskelioff and Peterson, 2003). The relationship of chromatin modification and transcriptional control is far from being clarified; researchers are still prying their way into this mystery.

### 1.1.3 DNA methylation and chromatin modification

DNA methylation is tied to the modifications in the chromatin. The most popular concept is DNA methylation directs the assembly of a specialized repressive chromatin structure, thus the transcription machinery has no or less access to such a structure, resulting in gene silencing. The breakthrough in understanding how methylation-mediated repression worked was the finding that the methyl-CpG binding protein MeCP2 interacts with a co-repressor complex containing histone deacetylases (HDACs) (Nan et al., 1998; Jones et al., 1998). Methyl-CpG binding proteins can read and bind to the methylated CpG dinucleotides, recruit other repressive proteins including HDACs, and promote the formation of compacted chromatin structure. This process leads to the final gene silencing. However, scientists started rethinking this epigenetic information flow upon the exquisite finding in filamentous fungi, which indicated that the methylation of H3 could control DNA methylation. This finding for the first time suggested that histone modifications can impact both *de novo* and maintenance DNA methylation (Tamaru and Selker, 2001). A similar sequence of epigenetic events was supported by other subsequent studies (Yates et al., 2003; Bachman et al., 2003; Mutskov and Felsenfeld, 2004). Apparently, cytosine methylation is not the primary cause in the process of gene silencing; instead, it acts as a sensor or reinforcement upon the gene already inactivated in other ways, such as by chromatin modifications. Evidence for an equivalent dependence in mammalian systems is eagerly in need. The causal relationship between gene silencing and DNA hypermethylation still remains controversial, although it is clear that gene silencing is associated with DNA methylation. Comprehensive research in more species with different approaches has yet to be done until we can draw the final conclusion.

## 1.2 Position effect variegation

### 1.2.1 Position effect variegation in transgenic animals

Genes transferred out of their native context are often subject to epigenetic effects. This phenomenon is described as position effect variegation (PEV), first discovered in Drosophila by Muller about 75 years ago (Muller, 1930). Basically,

PEV occurs when genes normally located and expressed in areas of euchromatic chromatin are translocated by chromosomal rearrangement or transposition to chromosomal areas with heterochromatic structures, resulting in variegated or mosaic expression patterns. PEV was viewed as restricted to Drosophila until the first case was reported in mouse about the brown coat color gene by Russell and Bangham (Russell and Bangham, 1961). A translocation between the X chromosome and autosomal genophore 8 (codes for the brown coat color) caused the variegated brown phenotype in XX female mice. Subsequent studies showed further that variegations in mice and in Drosophila share many common characteristics. The position effect on transcription is also implicated in human diseases with either strong epigenetic or imprinting components.

Numerous studies have genetically and cytologically documented that PEV can influence gene expression both *in cis* and *in trans* via chromatin environment. Most early work on PEV focused on the effects *in cis*: of chromosomal rearrangement on gene expression within the vicinity and on the same chromosome (Paro, 1990; Shaffer et al., 1993; Howe et al., 1995; Weiler and Wakimoto, 1995). When a normally euchromatin gene is juxtaposed with heterochromatin, transcription will be repressed or silenced. However, PEV can act over great distance. Research on PEV *in trans* was best exemplified by the dominant brown eye gene mutation in Drosophila. The resultant mutated allele by an insertion of a large block of heterochromatin can bring in the wild-type allele to associate with the repressive chromatic complex at the nuclear periphery (Henikoff et al., 1995; Dernburg et al., 1996; Csink and Henikoff, 1996; Csink and Henikoff, 1998). Taken together, chromosomal rearrangements impact nuclear architecture and consequently impact nearby gene expression *in cis* and *in trans*.

### 1.2.2 Position effect in transgenic cell lines

Position effect variegation (PEV) also expands its influence into the construction of transgenic cell lines. Stably integrated transgenes are often poorly expressed because of position effects that are caused by the influence of the site of chromosomal integration (Bestor, 2000a). In cultured cells, two categories of position effects have been recognized: stable and silencing (Walters et al., 1996; Pikaart et al., 1998). Stable position effects are characterized by pancellular

expression of transgenes, which are really rare events. In contrast, silencing position effect is the most common phenomenon encountered in the cultured cells. It is characterized by progressive silencing of the transgene at a rate associated with the site of integration. Additionally, this silencing effect has some similarity to those observed in Drosophila and mammals (Graubert et al., 1998; Wakimoto, 1998), which is characterized by clonally inherited silencing of expression in a fraction of the cells of a given tissue. PEV thereby results in heterogeneous expression of the transgenes. It was recently reported that the position effect could even be achieved by either orientation of the transgene at the identical genomic locus in a mammalian cell line (Feng et al., 2001). The expression of the transgene was found to be influenced by the flanking chromatin. These data on PEV make sense in light of our understanding of the relationship between chromatin assembly and gene transcription because gene transcription requires a loose chromatin environment in order to get access to the transcription machinery.

Position effect variegation is also the major obstacle to accurately dissect the functions of many elements within the transgene because of the unpredictable integration sites. Obviously, targeting different transgene constructs to a defined genomic locus becomes the prerequisite for those purposes. Undoubtedly, the advent of site specific recombination mediated by the Cre or Flp recombinase meets this end. Transgenes can be targeted to a predefined genomic locus, namely the lox P or the FRT site. These sites as such are products of random events, but through them, transgenes of interest are able to be repeatedly targeted to those defined genomic loci. Therefore, position effects are eliminated among clones obtained from the same transgene construct.

## 1.3 Repeat-induced gene silencing

When exogenous DNA is transferred into mammalian cells, chromosomal inserts almost always consist of long tandem arrays of the transfected transgenic DNA (Palmiter and Brinster, 1986). Sometimes transgene arrays can comprise hundreds or even thousands of copies. The transgene arrays often show unpredictable levels of expression and are very susceptible to gene silencing.

The transgenic suppression triggered by the presence of multiple copies is thus termed as repeat-induced gene silencing (RIGS).

60 years ago, Pontecorvo put forth a bold conjecture: a heterochromatic segment should arise every time that a minute euchromatic region undergoes repeated reduplication in the genotype and the replicas remain adjacent to each other in the chromosome (Pontecorvo, 1944). He was obviously inspired by the notion that heterochromatin is repetitive, whereas euchromatin is not. During the last decades, it was amply confirmed that tandem repeats presented in abundance in heterochromatin but not euchromatin. In Drosophila, heterochromatin consists primarily of simple sequence repeat arrays, whereas euchromatin consists primarily of single-copy sequences. Direct evidence is available that repetitive sequences induce heterochromatin formation, possibly because these sequences can pair with one another (Dorer and Henikoff, 1994). Another fact is the hypermethylation of the transgenes within the inactive tandem array reported in many studies with mammalian systems. However, it is unclear whether methylation precedes or is the consequence of the observed chromatin compaction associated with repeated transgene arrays.

As a matter of fact, repeat-induced gene silencing was first documented in plants. An inverse correlation between transgene number and gene activity was observed (Jorgensen, 1990). However, the significance of this finding was not realized until the description of trans-inactivation by Matzke et al. (Matzke and Matzke, 1991) and other homology-dependent gene silencing phenomenon in Nicotiana tabacum (Hobbs et al., 1990), Petunia hybrida (Linn et al., 1990) and Arabidopsis thaliana (Scheid et al., 1991). Although a molecular mechanism for repeat-induced silencing is yet to be fully elucidated, the correlation between the silenced state and the adoption of a less accessible chromatin configuration observed both in mammalian cells and in other lower eukaryotes is consistent with a model in which homologous pairing within the array induces heterochromatinization at the transgenic loci.

The distinguished evidence of RIGS in mammals is the work done by Garrick and his colleagues (Garrick et al., 1998a). They applied the lox/Cre system to integrate different numbers of a transgene, lacZ, at the identical

chromosomal location in the mouse genome. They found that the reduction in copy number of the transgene resulted in a marked increase in the expression level, accompanied with decreased chromatin compaction and decreased methylation at the transgene locus. A follow-up study in a cultured mammalian cell line also showed that tandem repeats of the transfected lacZ gene were also silenced mediated by a compacted chromatin structure (McBurney et al., 2002). Therefore, the presence of multiple homologous copies of a transgene within a concatameric array indeed has a repressive effect upon gene expression in mammalian systems.

Noticeably, the arrangement of endogenous loci, such as the rRNA, tRNA and histone genes as high-copy concatamers of a repeat unit suggests that multiple copy arrays need not always be subject to repeat-induced gene silencing. Therefore, other characteristics must exist to distinguish between the natural cellular repeat and transgene tandem arrays.

## 1.4 Host defense mechanism

Gene silencing is due not only to the endogenous rearrangements, but also as a consequence of the integration of exogenous DNA into a host cell or organism. Throughout evolution and still today, cells deploy elaborate host defensive systems to protect their genomic structure, to oppose the expression of abnormal transcription units, and to extinguish the harmful invasive genes (Xu and Bestor, 1997; Yoder et al., 1997). Increasing evidence supports the idea that various transgene silencing phenomena reflect the active host defense unleashed upon natural invasive DNA or parasitic DNA sequences, such as transposable elements, retroviruses, bacterial DNA, but also by artificially introduced plasmid DNA sequences.

### 1.4.1 Restriction-modification system in prokaryotes

The best studied host defense mechanism is in prokaryotes, and is called the restriction-modification system. In prokaryotes, the cytosine or adenine is modified by the DNA methyltransferases with corresponding sequence specificity,

and prohibits the action of several endogenous restriction endonucleases on the host genomes. By contrast, extraneous DNA is cleaved sequence-specifically by the host's restriction endonucleases. The ability to excise foreign integrated DNA could be considered the most effective defense mechanism of the host against a very incisive mutagenic event.

### 1.4.2 Host defense in eukaryotes

Gene transfer is commonly used to generate transgenic cell lines, transgenic plants and mice. However, in numerous instances, the transgenes have been rapidly or gradually inactivated and their sequences were extensively methylated (Lichtenberg et al., 1988; Orend et al., 1991). Gene silencing has been very well known to compromise a number of gene transfer efforts and to be a barrier to many forms of gene therapy currently under development. Transgenes could activate a genome defense system that is normally triggered by transposable elements and other natural invasive DNA. This mode of gene silencing is thus considered as a further cellular defense mechanism against foreign gene expression. Since no occurrence of restriction endonucleases were found in mammals, and the resultant mechanism of degradation or elimination of extraneous DNA taken up in the cells is not functional. Therefore, methylation of the integrated foreign DNA is currently considered to be the main host defense mechanism to inactivate invasive genetic materials. DNA methyltransferases are apparently the only remnants of the prokaryotic restriction-modification system (Doerfler, 1992). Fungi and plants have similar methylation-based systems and also employ diverse post-translational mechanisms to reduce or eliminate the expression of alien transcription units. Other mechanisms might also exist to operate in organisms that apparently lack the capacity to methylate their DNA.

### 1.4.3 Possible DNA feature defined as foreign in host

How does the host distinguish a fragment of DNA sequence as foreign? It can be explained by the DNA sequence feature of the transgene or the vector used in all forms of gene transfer. Dinucleotide frequencies in the human genome differ markedly from those of other organisms. In particular, human DNA has low

frequencies of CpG dinucleotides except in CpG islands. Other genomes, such as bacteria, tend to be relatively rich in CpG dinucleotides, even in regions of moderate GC content. The lacZ gene of *Escherichia coli* is normally inactivated by the methylation pathway in transgenic mice, but a synthetic version of the gene that followed human codon usage was relatively resistant to methylation and silencing (Scrable and Stambrook, 1997). Stretches of prokaryotic vector sequences widely used in gene transfer technique are not well tolerated by higher eukaryotic genomes (Matzke and Matzke, 1998). Vector sequences often gain spontaneously dense methylation that can usually spread into neighboring transgenes and result in transgene inactivation (Jakowitsch et al., 1999). These regions of DNA with different dinucleotide frequencies than the host genome perhaps lack the ability to bind eukaryotic nuclear proteins and thus are conspicuous to *de novo* methyltransferases (Matzke et al., 2000).

## 1.5 The property of the promoter on gene expression

### 1.5.1 Promoter and gene expression

Promoters play the most important roles in the control of gene expression because the binding of the transcription machinery to the promoter sequence is the prerequisite for gene transcription. Promoters inherently contain information to distinguish functional specificity (Lemon and Tjian, 2000). Promoter often contains many modules, such as the core promoter, enhancers, repressor and boundary elements. They are also shown to be involved in the formation or maintenance of different chromatin states with respect to a gene expression state. This reflects on the fact that gene expression is normally different from tissue to tissue and in different developmental stages, although the genetic material is the same in almost every single cell of all the tissues throughout the entire life. It was actually proposed very early on that there were promoter components on the extent of the epigenetic effects. Among them are the active enhancers. They are known to be able to recruit chromatin modeling factors through enhancer-binding factors to disrupt chromatin structure (Lemon and Tjian, 2000). Recent studies with human ß-globin transgenes in mouse cells have found that an intact enhancer is required to suppress the silencing effects through

the exclusion of these transgenes from heterochromatin and furthermore correlated histone acetylation of the transgenes with this exclusion (Francastel et al., 1999). Therefore, promoters differ not only in their strength and developmental or tissue specificity, but also in their susceptibility to epigenetic events.

### 1.5.2 Different promoter types in eukaryotes

Different types of promoters exist in eukaryotes. The normal transcription of many, perhaps all, eukaryotic genes requires activators. In yeast and other simple organisms, upstream activating sequences (UASs) were identified that is sufficiently to accelerate transcription initiation. They are usually located about 250 base pairs upstream of the transcription start site (Guarente et al., 1982). UAS fails to work when positioned at a distance greater than ~600-700 base pairs upstream (Keegan et al., 1986), or anywhere downstream of the gene (Struhl, 1984; Guarente and Hoar, 1984). In contrast, mammalian genes contain complex arrays of specific DNA sequences that combine more commonly shared core promoter elements with broadly diverse gene-specific enhancer elements and cooperate to define specific expression patterns (Dynan, 1989). The highly conserved sequence called the TATA box is required for sufficient transcription initiation, and located in a fixed position about 25-35 bases upstream of the transcription start site. Enhancers are normally found in the promoter components of tissue specific genes, and are generally associated with facultative chromatin. A large number of genes have been studied in which no TATA box is evident by sequence analysis. Many of these genes are not transcribed at high rates. Noticeably, most of them contain a GC-rich stretch of 20-50 nucleotides upstream of the transcription start site. SP1 factors can recognize these sites. Therefore, genes lacking of the TATA box might rely on these SP1 binding sites to initiate sufficient transcription.

### 1.5.3 Promoter choices in transgenic technology

The choice of the appropriate promoter is critical for stable (trans)gene expression. The most widely used promoters in transgenic experiments are

strong viral promoters or enhancer containing promoters. Human cytomegalovirus (CMV) promoter, simian virus 40 (SV40) promoter, and Rous Sarcoma Virus (RSV) promoter are of viral origin and commonly used in transgene delivery. Among them, the CMV and SV40 promoters contain enhancers. These promoters enable reasonable transgenic expression in short-term transient studies, particularly in a given cell type specific context. Recently, housekeeping gene promoters and ubiquitination promoters have been pointed out as strong non-viral promoters and can direct transgene expression to relatively high levels. Human elongation factor-1 alpha promoter (EF-1a or EF) is one of them and showed its promising utility in many transgenic studies. More importantly, due to its indispensable housekeeping function in all cells, the EF promoter expression is consistent from a temporal viewpoint, relatively insulated from changes in cell physiology and is cell type independent (Kim et al., 1990; Wakabayashi-Ito and Nagata, 1994; Goldman et al., 1996).

## 1.6 Gene regulation modes: the on/off vs. the rheostat model

In higher eukaryotes, gene transcription depends heavily on transcriptional enhancer elements located up- or downstream of the core promoter. Gene transfer experiments thereby commonly apply enhancers to activate the transgene expression. Enhancers are well known to increase the transcription level in a variety of assays. However, the meaning of 'increase the transcription level' continues to be a problem and has attracted a lot of attention. Questions were raised on whether enhancers determine the transcription state of a gene (the on/off model) or they regulate the amount of RNA transcribed from an active gene (the rheostat model) through which the enhancers increase the transcription level of the gene (Fig. 2).

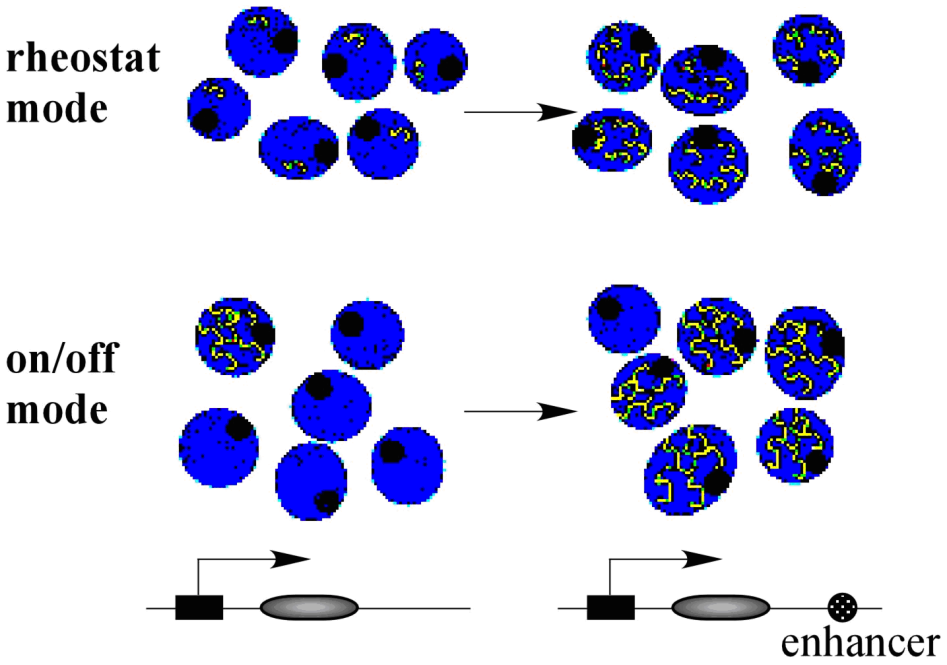


**Fig. 2. Rheostat versus on/off models for transcriptional activation by enhancers.**

The yellow wavy lines represent mRNA transcripts from the construct outlined below the blue cells. The rheostat mode involves an increase in the amount of mRNA transcribed by the gene in each cell when the gene is linked to an enhancer. The on/off mode involves heterocellular expression of the gene in the absence of an enhancer, and an increase in the proportion of expressing cells when an enhancer is linked to the gene. If each population of cells is analyzed for total mRNA expression, the two modes are indistinguishable: in each case the enhancer increases the amount of RNA. When a single cell assay is used, however, the difference is apparent. Adapted from Fiering, BioEssays, 2000.

Enhancers of pol III-transcribed rDNA have long been thought to act solely by the on/off mode (Trendelenburg and Gurdon, 1978; Bakken et al., 1982; Morgan et al., 1983), but it has never been widely accepted for enhancers of promoters that use pol II. Early interpretations of enhancer action relied on experiments from prokaryotic system and yeast (Ptashne, 1986). Different than higher eukaryotes, however, cell differentiation is negligible in unicellular organisms. Yeast genes are constantly in the active state, thus relatively simple UASs are sufficient to accomplish changes in transcription rate. In eukaryotes, the 'classical' effect of enhancers had been proposed following early work with

the SV40 enhancer. The enhancer was detected to increase the expression level of the gene, and this increase resulted from the elevated transcription rate (Weber and Schaffner, 1985; Treisman and Maniatis, 1985). Obviously, the experimental assays used in those studies are critical to dissect the aforementioned two different transcriptional regulation models. A typical transient assay involves transfection of a plasmid to a dish of cells; after a period of time some or all of the cells are rendered into an extract which is assayed for mRNA or a certain gene product (such as chloramphenicol transferase, luciferase), or a product secreted into the medium is assayed (such as human growth hormone). Such 'bulk' biochemical assays do not permit the distinction between the on/off and the rheostat modes because either mode results in an increase in the total assayed product. The same limitation holds true for *in vitro* transcription reactions. The increase of the gene transcripts could be due either to an increased polymerase loading (by rheostat mode) or to recruitment of otherwise inactive templates (by on/off mode). Likewise, bulk biochemical assays of expression from stably integrated transgenes in cell lines or mice can not distinguish the two modes either. Current textbook discussions of enhancer activity do not address the veracity of the rheostat model, and it has permeated thinking on transcriptional control to the extent that many are unaware that there is any alternative. However, the shortcoming of the bulk assay was noticed before long and the single cell assay was inevitably introduced into the dissection of the enhancer effect on the transcriptional regulation mode. Weintraub first re-examined the effect of the SV40 enhancer using a single cell assay and concluded that the enhancer had only a slight effect on the transcription rate (rheostat mode), but greatly increased the probability of a transfected template becoming active (on/off mode) (Weintraub, 1988b). Other studies using the lacZ reporter gene and fluorescence-activated cell scanning (FACS) technique also supported this result (Walters et al., 1995b). *In vivo* experiments also contributed to the understanding of the regulation modes of enhancers in their native context (Gu et al., 1993; Chen et al., 1993; Hug et al., 1996; Bouvier et al., 1996; Sleckman et al., 1997; Epner et al., 1998). These experiments suggest that at least some enhancers have both on/off and rheostat effects, but that their influence on transcription rate is not large. Apparently, both on/off and rheostat modes are applied to endogenous and transgene regulation.

Recent reports have provided more detailed evidence of stochastic gene regulation (Ross et al., 1994; Newlands et al., 1998), however, they do not provide a clue to the regulatory elements controlling this stochastic process. The enhancers are suggested to likely have a key role since they have the probability to alter the gene expression (Magis et al., 1996; Francastel et al., 1999). Further experiments to confirm this hypothesis are necessary. Nevertheless, the current knowledge does allow us to speculate on the biological utility of a stochastic function for enhancers.

## 1.7 Aim of the study

I am interested in analyzing the regulation of foreign gene expression at the single cell level in our transgenic expression system in order to shed more light on the expression regulation of endogenous genes, and if possible, apply the knowledge gained here to achieve a stable transgenic expression system in a general sense. Therefore, the study presented here initially aimed:

- To test whether a single integration of the transgene unit into a predefined and transcriptionally permissive locus using Flp-In system in the host cell line genome could bring about long term stable transgene expression.
- To study whether promoters from different origins affect the stability of transgene expression by comparing the human elongation factor 1 a promoter (EF-1a) and the commonly used human cytomegalovirus promoter (CMV).
- To determine whether the CpG content of the transgene sequence is inversely related to the stability of the transgenic expression through introducing the widely used enhanced green fluorescent protein (EGFP) reporter gene into cultured cells.
- To define the gene regulation mode through detecting the gene expression level at the single cell level by fluorescent activated cell scanning (FACS) analysis in the transgenic system used in this study.
- To investigate the effect of foreign DNA sequences on transgene expression by transfecting only the essential transcription units and obtaining stable clones with fluorescence-activated cell sorting (FACSorting).

Largely by chance, these experiments led me to address different methods of selecting transgene integration events, which turned out to be a

decisive factor determining the stability and homogeneity of transgene expression.