4. DISCUSSION

4.1 Transgenic expression in the Flp-In system

Standard gene transfer methods used in cell engineering often cause variable gene expression between different clones. This variation mostly arises from chromosomal position effects. Moreover, the tandem array of transgenes integrated into the cell genome is another long suspected culprit for unstable transgenic expression in cell lines. In contrast, site specific recombination mediated gene transfer is able to target the transgene construct to a particular genomic site with a single copy of the transgene. Therefore, gene transfer mediated by site specific recombination undoubtedly has the advantage to eliminate the position and copy number variation among clones obtained with the same transgene construct, and is ideally suited to analyze various parameters controlling gene regulation. The Flp recombinase was proved for successful gene transfer in mammalian cells (O'Gorman et al., 1991). The Flp-In system used in this study is based on site specific recombination mediated by the Flp recombinase. Thus, it enables a single copy of the transgene construct to be integrated into the predetermined single FRT site already inserted in the host Flp-In 293 cell genome (section 2.2.10). Flp-mediated targeting does not itself neutralize position effects; rather it ensures the same position effect is imposed on the integrated transgene in all clones obtained, or it applies an identical chromosomal background to all clones. By repeatedly targeting different transgene constructs to the same FRT site, I was able to compare the effects of individual factors on transgene expression at the single cell level. Time course experiments showed that the viral CMV promoter could direct the expression of the transgene $EGFP^{WT}$ or $EGFP^{CpG-}$ in culture for a relatively longer period of time compared to the house keeping gene promoter EF. The CpG dinucleotide content of the transgene coding region did not make a dramatic difference in the stability of transgene expression under nonselective conditions over time. Additionally, by applying the single cell assay and single copy integration I could accurately dissect two transcriptional regulation modes in my experimental system: the on/off and rheostat modes. Since they were inserted into the same genomic site,

the expression state of the transgenes did not show a remarkable variance between clones derived from identical transgene constructs.

4.1.1 Single copy integration is not the prerequisite for stable transgene expression

Tandem arrays of transgenes integrated into a cell's genome often result in unstable transgene expression. This is also known as repeat-induced gene silencing (RIGS). RIGS was observed in both mammalian organisms and cells (Garrick et al., 1998b; McBurney et al., 2002). Despite of RIGS, inactivation of single copy transgenes was also encountered in cell lines (McBurney et al., 2002). However, that study was done through random transfection of the lacZ-neo transgene into the host cell genome. Therefore, position effects or integration site effects were not controlled when addressing the effect of copy number on transgene silencing. In comparison, the integration site of single copy transgenes in the experimental system used here is identical in every Flp-In 293 cell. This locus was determined as a transcriptionally favorable locus (Invitrogen Co., the Netherlands). High expression level of the single copy $EGFP^{WT}$ and $EGFP^{CpG-}$ transgenes was indeed monitored under selective conditions in culture. After withdrawal of the hygromycin selection, uniform and stable expression of transgenes was detected at the early stages of cell passaging. However, progressive transgene silencing unavoidably took place over time under nonselective conditions regardless of single copy integration of the EGFP transgene. This implies that progressive transgene inactivation is overriding, if it is not a universal phenomenon, and far more complicated than what we envisaged. These results also challenged the recombinase mediated cassette exchange (RMCE) *per se*. Isogenic cells are not necessarily phenotypically identical. Some other characteristics must therefore exist to identify the single copy of the integrated transgene as a substrate for gene silencing. Likewise, other unknown mechanisms must be responsible for this silencing process.

Silenced transgene array was often found to be associated both with hypermethylation of transgenes and with the adoption of a repressive local chromatin configuration. A lower copy number of lacZ transgenes was

accompanied by a lesser degree of chromatin compaction and decreased DNA methylation in mouse genome (Dorer and Henikoff, 1994; Garrick et al., 1998c). So far, there is no report of generation-associated silencing of transgenes in transgenic mice. By contrast, in transgenic cell lines, even single copy integration of the transgene is susceptible to silencing. This led me to think about different transgenesis strategies. In transgenic animals, generally only the essential transcription unit of the transgene is integrated into the host animal genome based on an early observation by Chada et al and Townes et al. They both observed high tissue specific expression of the cloned ß-globin gene when most of the plasmid sequences were removed prior to injection; whereas most of the globin genes that were not expressed were introduced together with the vector sequences (Chada et al., 1985; Townes et al., 1985). However, this is impractical in generating transgenic cell lines because an antibiotic resistance gene must be integrated and used for stable clone selection. As a result, the whole plasmid is inclusively transfected into the host cell line. An antibiotic resistance gene and a replication origin are major components of the block of prokaryotic sequences included on expression plasmid constructs. Those prokaryotic sequences are indispensable for selecting stable clones in transfected cells and in propagating the plasmid in *E. coli*, respectively. However, these sequences are of prokaryotic origin. Remarkably, prokaryotes have a different genomic organization, with a richer CpG content compared to eukaryotes. Additionally, the DNA is predominantly unmethylated, whereas the mammalian DNA has less frequent CpG dinucleotide content and is largely methylated (Krieg, 2002). Presumably, the prokaryotic sequences integrated together with the transgene could trigger certain mechanisms, such as a host defense mechanism, to suppress expression of the alien DNA sequences (Clark et al., 1997). Therefore, it is possible that the presence of prokaryotic sequences will lead to a compacted chromatin structure either caused by or reinforced by DNA methylation. In either case, heterochromatin effects and DNA methylation will extend to the transgene region and cause suppressed expression. In Drosophila, wide spread of DNA methylation has not been found (Urieli-Shoval et al., 1982), but heterochromatin can still be independently formed. Furthermore, suppression of transgene expression in late-passage cell clones could not be reversed by treatment with 5-azacytidine (5-azaC, a DNA methylation inhibitor) or trichostatin A (TSA, a histone deacetylase inhibitor) alone, but could only be overcome by pretreatment

with 5-azaC followed by TSA (Lorincz et al., 2000). Therefore DNA methylation seems to act as a reinforcement following the chromatin compaction mediated by histone modifications. My experimental observations regarding treatment with 5-azaC and TSA also support this view. Therefore, I propose that the flanking plasmid sequences at the transgene locus will initiate and adopt a compacted chromatin configuration over time without antibiotic selection in culture. This compacted chromatin configuration can spread bidirectionally, and would eventually encompass the transgene region. Meanwhile, DNA methylation was set into place to mark the compacted chromatin region permanently in order to reinforce the compaction. This assumption certainly needs to be further tested.

Little is known specifically about chromatin compaction and DNA methylation of single copy transgene in mammalian systems. Therefore, it will be promising to determine the DNA methylation state and chromatin structure modifications on the single copy EGFP transgene, and also the sequence of their occurrences on the transgene with the experimental system used here. One can also examine if silencing of the single copy transgene is actually initiated at the flanking plasmid sequences by bisulphite sequencing and chromatin immunoprecipitation (ChIP). Moreover, repeat-induced gene silencing (RIGS) can also be further investigated in this Flp-In system. I by chance obtained a clone with multiple copy integration of the $EGFP^{CpG-}$ transgene at the FRT site in the Flp-In 293 cells. Even there will be no statistical relevance attached to data obtained through analysis of this single clone, this clone still opens up the possibilities to reveal some essential differences of the DNA methylation and chromatin compaction on the EGFP transgene locus between single copy and multiple copy integrations.

4.1.2 The CpG content of the coding region and transgene expression

CpG methylation is an asserted component of epigenetic control mechanisms over gene expression in vertebrate animals and transgenic cell lines. Already low levels of CpG methylation can exert a distinct effect on transgene expression (e.g. 7%), as shown by (Hsieh, 1994). Some experiments suggest that methylation of promoters does not lead to transgene silencing until chromatin proteins are recruited to the region (Kass et al., 1997a; Irvine et al.,

2002). In contrast, methylation of the transgene coding region is often observed to be associated with transgene suppression (Feng et al., 2001; Lorincz et al., 2004). It was demonstrated that the density of CpG dinucleotides in the transgene coding region could have a causal role in the repression of transcription in mice (Chevalier-Mariette et al., 2003). Based on this knowledge, I applied the newly synthesized EGFP gene devoid of CpG dinucleotides ($EGFP^{CpG-}$) to my cell culture based transgenic system. However, I did not detect long term stable expression of the $EGFP^{CpG-}$ transgene under the direction of either the CMV or EF promoter without antibiotic selection (Fig. 10). Compared to the wild type EGFP ($EGFP^{WT}$), this synthetic $EGFP^{CpG-}$ did not show any better stability.

EGFP, one of the most widely used reporter genes in transgene research, gives frequently an unstable and mosaic expression pattern (Hanazono et al., 1997; Gram et al., 1998). However, during humanization of the sequence, numerous CpG dinucleotides were introduced into the EGFP gene. This is thought to explain the fragile expression of EGFP due to CpG methylation in mammals. Additionally, such approaches have been successful in the past, as changing the bacterial coding sequence of the lac repressor to reflect the more AT-rich mammalian genome was the single most important factor affecting transgene expression (Scrable and Stambrook, 1997). Based on these theories, the coding sequence of EGFP was resynthesized in our lab in a way to deplete all CpGs based on codon usage (M Hampf and M Gossen, unpublished data), and named as $EGFP^{CpG-}$. However, the $EGFP^{CpG-}$ transgene used in this study did not show detectable prevention of transgene silencing conferred by the methylation incompetent coding region. Similar to $EGFP^{WT}$ transgene, $EGFP^{CpG-}$ transgenes directed by the CMV promoter were gradually suppressed after three weeks in culture without hygromycin selection. Under the direction of the EF promoter, $EGFP^{CpG-}$ transgenes were silenced in a smaller portion within the population compared to $EGFP^{WT}$ transgene clone after three weeks under nonselective conditions. This marginal difference might be due to EGFP measurement errors or the temporary antagonistic effect of the unmethylatable $EGFP^{CpG-}$ on the spreading of heterochromatinization initiated from the flanking sequences. However, the general rate and tendency of the silencing process was alike throughout the whole time course. This result implied that methylation of the

coding region of the transgene could not underlie the observed silencing of transgenes, at least not in the context of our experiments. The possibility that methylation on the promoter region caused the silencing is also an unlikely reason. For one reason, both the CMV and EF promoters contain CpG islands. CpG islands are believed to be able to escape methylation under normal conditions (Bird et al., 1985). Additionally, the EF promoter is endogenous and is known to be not methylated in the host genome. It is reasonable to assume that the EF promoter used to drive transgene expression should not be recognized as foreign through the host defense mechanism, and should escape methylation as a result. Thus, by default, the origin of silencing was assumed to reside in the backbone sequence of the plasmid introduced into the host genome by transfection. Plasmid expression vectors are widely used for gene transfer in cell lines. However, the backbone of the plasmid is prokaryotic. Upon transfection, this will introduce CpG dinucleotide patterns foreign to the host cell as discussed in 4.1.1. Most likely, flanking prokaryotic sequences introduced along with transgene sequences will be rapidly recognized and methylated by DNMT3A and DNMT3B in host mammalian cells. They can serve as methylation foci and cause methylation spreading to the promoter region and ultimately lead to transgene silencing. Alternatively, prokaryotic sequences could be heterochromatinized through the host defense mechanism. The resultant compacted chromatin can either spread to the transgene region and cause silencing, or induce gene silencing at a distance since heterochromatin blocks are able to induce silencing at distances greater than 1 Mb (Weiler and Wakimoto, 1995).

Nevertheless, several studies witnessed improved activities or stabilities of transgene expression by reducing the CpG dinucleotide content in the coding regions and even on the whole plasmid used in gene delivery in culture (Scrable and Stambrook, 1997; Shimshek et al., 2002; Yew et al., 2002). However, when quantified, the improvement was rather marginal. In transgenic mice, however, transgenes with reduced CpG content exhibited ubiquitous and high level expression (Scrable and Stambrook, 1997; Chevalier-Mariette et al., 2003). Recently, Nicolas' lab reported a systematic study on the effect of CpG content on transgene silencing in mice (Chevalier-Mariette et al., 2003). The lacZ reporter gene had a CpG content varying from high to zero, and was driven by the EF promoter. Their data revealed that the lacZ transgene with a null CpG content

abolished the strong transgene repression observed in the somatic tissues of transgenic lines with higher CpG content. Noticeably, the difference between gene transfer in cell lines and mice is the DNA fragment integrated into the host genome. In transgenic mice, only the essential transgene expression cassette is injected, instead of the whole expression plasmid as used in the transfection of cell lines. In the latter situation, the inclusion of an antibiotic resistance gene is indispensable so that stable transfectants can be selected. However, this antibiotic resistance gene is exclusively of prokaryotic origin. Thus, it is plausible to anticipate that the difference observed in transgenic mice and cell lines by using CpG-poor transgenes is possibly due to the prokaryotic plasmid backbone as I hypothesized above.

4.1.3 The promoter effect on transgene expression

It is generally accepted that the ideal promoter to be used in gene transfer should provide for both a high-level and long-term expression of the transgene. Mostly, strong promoters are therefore chosen for this demand. Strong promoters can prevail over DNA methylation and therefore overcome repression of poorly methylated genes (Boyes and Bird, 1992). In the study presented here, I used two strong promoters, CMV and EF, to direct the expression of transgenes. However, neither the CMV promoter nor the EF promoter supported long term expression of $EGFP^{WT}$ or $EGFP^{CpG-}$ in the absence of antibiotic selection in the Flp-In system. Expression of transgenes driven by either promoter underwent progressive silencing under nonselective culture conditions. A slight increase in the expression level was detected with the EF promoter compared to the CMV promoter.

Promoters of viral origin, including the CMV promoter, are currently widely used in generating transgenic cell lines and animals. It is derived from the human cytomegalovirus, a member of the herpes group. Owing to its ubiquity as a human pathogen, it is active in many cell culture systems and most tissues. The CMV enhancer is considered to be one of the strongest enhancers in vitro and has little cell type or species preference, which makes it a useful component of eukaryotic expression vectors (Boshart et al., 1985). However, although the CMV promoter allows for a very strong short-term expression of integrated genes in

vivo, it is prone to inactivation over time, and can be silenced within a few weeks after gene transfer in many cell lines and animals (Loser et al., 1998; Fitzsimons et al., 2002; Kofman et al., 2003). I also confronted the difficulty in establishing stable EGFP positive cell lines with the CMV promoter in HeLa cells. During transient transfection, the EGFP expression level was high and comparable to those driven by other promoters (i.e.: the EF promoter). But during stable clone selection, the expression level of EGFP decreased and became extinct after two weeks (W. Liu, unpublished data). The same phenomenon was also observed by other researchers (Kirsch et al., 2003). In the Flp-In system, clones showed high level of $EGFP^{WT}$ or $EGFP^{CpG-}$ expression under the direction of the CMV promoter shortly after antibiotic selection, and the expression level was comparable to clones obtained with the EF promoter. Probably, the CMV promoter could prevail over a low density of DNA methylation during early stages; however, over time, DNA methylation became so dense along the transgene that even the strong CMV promoter could not overcome the overall suppression effect. This hypothesis might also explain the failure to generate stable EGFP clones in HeLa cells, as mentioned previously. In HeLa cells, the EGFP transgene was likely integrated as an array resulted from a normal transfection procedure. Thus, in comparison to single copy integration in Flp-In 293 cells, the silencing process in HeLa cells occurred much faster due to the repeat-induced gene silencing (RIGS) mechanism. After the two-week selection course, all cells had stopped expressing EGFP.

As a consequence of the decrease in transgene expression frequently associated with viral-derived promoters, housekeeping gene promoters are increasingly employed to assess their ability to promote sustained and strong transgene expression. Housekeeping gene promoters are promising to direct transgene expression because of their indispensable cellular function in every cell of an organism. The EF promoter, a ubiquitously expressed housekeeping gene promoter, has a wide host range and has attracted great attention. It has already shown a stronger activity than viral promoters in HeLa cells upon its isolation (Uetsuki et al., 1989), as well as in transgenic cell lines, mice and Drosophila (Kim et al., 1990; Ackermann and Brack, 1996; Xu et al., 2001; Lipshutz et al., 2001). In various cell types, the EF promoter strongly supported transgene expression transiently and stably, and more efficiently than other commonly used

promoters including SV40 and the long terminal repeat of the Rous Sarcoma Virus (Kim et al., 1990). Additionally, the EF promoter significantly increased success in establishing stable cell lines with consistent expression during routine culture passage even in the absence of antibiotic selection (Gopalkrishnan et al., 1999). In transgenic mice, luciferase driven by the EF promoter was detected at an 8~18 times higher expression level than with the CMV promoter. Additionally, luciferase expression was sustained for over six months (Xu et al., 2001; Lipshutz et al., 2001). The Drosophila EF promoter was also used for ubiquitous expression of transgenes in flies with advantages over the heat shock protein (hsp) promoter and the actin5C promoter, which have negative effects due to the requirement for heat induction for strong expression, and some developmental specificity, respectively. The EF promoter, however, showed strong activity in all cells and at all stages of development. It is also able to promote the expression of a reporter gene in Drosophila cell culture (Ackermann and Brack, 1996).

Despite all the successful reports about the EF promoter, I failed to obtain EGFP transgenic cell lines with sustained expression under direction of the EF promoter in the Flp-In system. However, the expression level of $EGFP^{WT}$ or $EGFP^{CpG-}$ was very strong during the transient period, even stronger than with the CMV promoter. In the absence of antibiotic selection, the $EGFP^{WT}$ or $EGFP^{CpG-}$ transgene driven by the EF promoter could still not escape silencing over time. The EF promoter is of endogenous origin, so it is likely shielded from DNA methylation. However, prokaryotic plasmid sequences, introduced into the integration site during the recombination process, flank the EF promoter and the $EGFP^{WT}$ or $EGFP^{CpG-}$ transgene. These sequences have different genomic characteristics and are prone to be methylated in eukaryotes (Krieg, 2002). Therefore, the compacted chromatin structure and DNA methylation on the proximal prokaryotic sequence could gradually spread into the transgene region and lead to suppressed transgenic expression as discussed in the previous section.

4.1.4 The on/off vs. rheostat model

Two different suppression modes were observed in my experimental system with the single cell assay. They resemble the two reported modes in gene

expression regulation, the on/off and rheostat modes (Weintraub, 1988a; Walters et al., 1995a; Martin, 2001). The CMV promoter is associated with the rheostat mode, whereas the EF promoter showed an on/off mode in the gene silencing process.

The Flp-In system used in my study has unique properties allowing me to address the gene regulation mode. In addition to an identical integration site in the host cell genome for all transgene constructs to eliminate the position effect variegation, only single copies of transgenes are integrated. This is critical to distinguish between the overall moderate expression from an array of transgenes, and high expression from a single copy in the transgenic array (Fig. 20). Here are illustrated three copies of the EGFP gene inserted into a genome. In one case, all three copies give moderate level of expression; in the other, only one copy expressed the EGFP transgene, and at a high level. It is possible that the additive expression from all three copies in the first case comprises the higher expression from only one copy of the EGFP transgene in the second case. Therefore, the final output of EGFP expression would be indistinguishable from one another in these two situations, when assayed at the single cell level. Very clearly, multiple copies of the transgene is not capable for this distinction; whereas a single copy

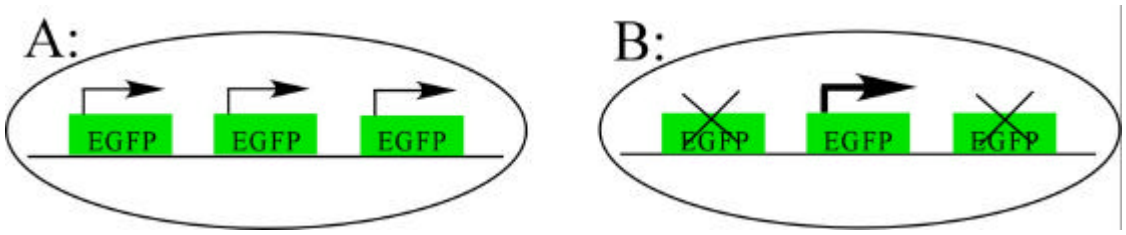

Fig. 20. Illustration for the inability of multiple copy transgenes to dissect gene regulation modes.

Three copies of the transgene EGFP were exemplified to be integrated into the host genome. A: all three copies are expressed at a moderate level. The EGFP expression level detected in each cell is the addition of expression from these three copies. B: only one out of the three copies expresses EGFP at a high level (assumably three-fold that of the moderate level to simplify the illustration), while the other two copies are silenced. The EGFP expression level detected in every cell is exclusively from the single copy, but is equivalent to the level detected in the situation A. Obviously, multiple copy integration is incapable of separating these two scenarios of gene regulation.

integration of the transgene can meet this demand. Since gene suppression is the reverse process of gene activation, I therefore shed more light on gene regulation with this unique experimental system.

The prevailing concept is that enhancer containing promoters increase the probability that a linked transcription unit establishes or maintains the active transcription state, or modulate gene regulation in an on/off mode; whereas housekeeping gene promoters regulate gene expression by controlling the transcription rate, or in a rheostat mode (Fiering et al., 2000; Martin, 2001). The CMV promoter has a strong enhancer; however, a rheostat gene regulation mode was detected during the transgene silencing process. The EF promoter is a typical housekeeping gene promoter, but the on/off mode was observed in an identical experimental setting as described for the CMV promoter. The simple explanation for this contrary observation could be that the gene activation process is not simply a reversal to the silencing process. During the silencing process, the CMV enhancer might be able to antagonize the suppressive chromatin effect and keep the suppressive chromatin structure in a metastable state. I hypothesize that in the early silencing stage, transcription factors could still bind to the CMV promoter with the aid of an enhancer. However, reinitiation of transcription by the transcription machinery was slowed down, likely due to the suppressive chromatin effect. Different cells might have different degrees of suppressive chromatin compaction. Therefore, the single cell assay would reveal different levels of EGFP expression in individual cells among the population (i.e.: the rheostat mode). With the progressive silencing process, more and more cells lower their EGFP expression level. This explanation could also apply to the on/off mode regulation associated with the EF promoter. Since the EF promoter has no confirmed enhancer element, cells were either subjected to silencing or escaped silencing in the process of silencing. This depends on whether or not the transcription machinery could bind to the EF promoter region. When assayed at the single cell level, the whole population would be separated into either EGFP expressing or nonexpressing subpopulations (i.e.: on/off mode.)

The effects of enhancers were mostly studied by transfection of a transgene driven by either enhancer containing promoters or enhancerless

promoters. When analyzed in stably transfected clones, single copy integration was generally not addressed in these studies. However, this is an important factor in determining the regulation mode, as illustrated in Fig. 20. When the transgene is integrated into the cell genome as a tandem array, which is mostly the case, the transgene may be subjected to RIGS. Not every copy of the transgene in the array behaves the same (McBurney et al., 2001). The RIGS pathway would interfere with the authentic regulation role conferred by the enhancer. The reported on/off regulation conferred by an enhancer might be an artificial product of the enhancer 'fighting' against RIGS. If this is the case, then the on/off mode can only apply for this type of experimental transgenic systems. Additionally, position effect was not considered in the aforementioned experimental systems. Different integration sites can lead to variable transgene expression. Together with a tandem array of integrated transgenes, interpretation of the enhancer effect is therefore complicated and ambiguous in those experimental systems.

4.1.5 The irreversible expression pattern of subpopulations was caused by an undetectable oscillation

I did not detect a reversion between $EGFP^{WT}$ or $EGFP^{CpG-}$ transgene expressing and nonexpressing states within FACSorted subpopulations (see Fig. 12). Oscillation has been reported in several lymphocytic and fibroblastic cell lines and likely accounts for mosaic expression patterns observed in transgene expression (Kerr et al., 1989; Ko et al., 1990; Fiering et al., 1990; Reddy et al., 1991; Feng et al., 2001). Feng et al observed that populations of mouse erythroleukimia cells (MEL) were over 99% either lacZ expressing or nonexpressing returned to a mosaic expression pattern resembling unsorted populations after a few days. Their findings demonstrated that lacZ expression at the integration locus was dynamically regulated: the integration locus was constantly being activated and inactivated (Feng et al., 1999). They further confirmed this finding by using the EGFP transgene in MEL cells by both subcloning and FACSorting approaches (Feng et al., 2001). Within 48 h of the start of culture, about 10% of the EGFP-negative cells already showed positive EGFP expression. The reversal was also detected in the EGFP-positive fraction.

Both negative and positive populations had reverted to heterocellular expression patterns similar to the parent population within several days. The oscillation of lacZ gene expression was also observed in our lab using HeLa cell lines after subcloning by limited dilution (M. Gossen, unpublished data). Isogenic clones gained the parental mosaic transgene expression pattern. Whether or not the cells express the transgene, they all have the same genomic material. Oscillations are therefore caused by an epigenetic phenomenon (Feng et al., 1999). Two factors could influence these oscillations: either the concentration of critical transcription factors varies more or less randomly in the cells or the fine chromatin structure of the transgene locus oscillates between transcriptionally competent and transcriptionally active states resulting in nonexpressing and expressing cells.

The EF-EGFP^{WT} transgene clone analyzed in this study did not show detectable oscillation between EGFP expressing and nonexpressing states (Fig. 12). Time course experiments revealed that the expression of EGFP^{WT} in this clone was progressively diminished - that is more and more cells stopped expressing EGFP^{WT} over time without the hygromycin selection (Fig. 10). Therefore, I reason that oscillations probably occurred in my expression system as well; however, the speed of oscillation changed over time, or oscillation was overwritten by the overall transgene suppression over time. During hygromycin selection and immediately after release from hygromycin selection, the EGFP^{WT} locus was in a transcriptionally active state. This could be due either to lack of oscillations in the presence of hygromycin selection or to the fact that in the presence of hygromycin selection the oscillation was so rapid that sufficient amount of EGFP proteins still remained in the cells between oscillations. Both possibilities can be explained by the selection pressure from hygromycin that forces the cells to constantly transcribe the hygromycin phosphotransferase (HPT) or to keep two rounds of transcription short enough in order to survive the antibiotic selection. As a result, the nearby EGFP transgene was also transcribed in the same manner owing to the transcriptional permissive chromatin structure and the presence of enough transcription factors in this chromatin domain. After withdrawal of hygromycin from the culture, expression of HPT was not preferentially selected any longer. As a result, oscillation was able to slow down between the two expressing states. EGFP expression was probably affected in

the same way. Thus, a heterogeneous expression pattern appeared in the initially EGFP homogeneous population over time after withdrawal of antibiotic selection. This could be explained by the growth disadvantage of EGFP positive cells or transgene load altering time of cell growth. EGFP negative cells grow faster than EGFP positive cells and take over the culture (W. Liu, unpublished data). When the oscillation became slow enough that EGFP positive cells would consequently be diluted out by overgrowing negative cells in the population. This idea might plausibly explain the irreversible expression pattern observed in Fig. 12 when cells were kept under nonselective conditions.

Different results on oscillation can also be explained by different cell lines, genomic integration sites, regulatory elements, or experimental strategies used in the studies. The aforementioned results with lacZ and EGFP reported by Feng et al (Feng et al., 2001) were based on different experimental strategies. Stable clones were obtained without antibiotic selection, even though the integration of the reporter genes was achieved by recombinase mediated cassette exchange (RMCE), an approach similar to mine. Antibiotic selection applies unnatural forces to cells (Rodolosse et al., 1998), forcing them to transcribe an antibiotic resistance gene and to adopt a corresponding chromatin state. Cells therefore behave differently with or without antibiotic selection. These differences can affect the ability to obtain stable and homogeneous transgenic clones.

More generally, expression oscillations could be the molecular mechanisms of heterogeneity and apparent randomness of the response of numerous biological systems to endogenous or exogenous signaling molecules (Feng et al., 1999).

4.2 Stable transgene expression in clones achieved by bypassing antibiotic selection

The "Sorting-Subcloning" approach newly introduced in this study undoubtedly provides an efficient way to obtain transgenic cell lines that can maintain sustained and homogeneous expression over a long period of time. Preceding attempts mostly focused on including strong native regulators such as locus control regions (LCRs) or insulators in the transgene construct (Grosveld et al., 1987; Pikaart et al., 1998; Mutskov et al., 2002), or on making the transgene so large that it is likely to include all the sequences required to establish its native epigenetic organization (Peterson et al., 1993; Porcu et al., 1997). More recently, great efforts were put into modifying the transgene expression vector in a way to reduce or deplete the CpG dinucleotides or changing the transgene sequence based on the human codon usage (Shimshek et al., 2002; Yew et al., 2002). Prokaryotic sequences have a high content of CpG dinucleotide that are not methylated. Thus, after integrated into eukaryotic genome, they would be recognized and methylated, consequently suppressed. However, neither of these strategies was fully effective. This highlights the fact that fundamental mechanisms of transgene silencing remain largely unclear even after so many years of experimental analysis.

In addition, 5-azaC and TSA treatment could not lead to complete reactivation of transgene expression in late stages of silencing (Feng et al., 2001). In early stages of silencing, 5-azaC and TSA treatment could reactivate transgene expression to a certain extent, but once the treatment was removed, the transgene returned to the initial silenced state (Gram et al., 1998). Therefore, it seems a permanent transcriptional barrier is being established along the process of transgenic silencing. Once the transgene is silenced, it is difficult to recover the transcription potency.

What inspired me to establish the "Sorting-Subcloning" approach is the gene transfer technology used in generating transgenic mice. In transgenic mice, after the transgenes are stably integrated into the genome, they can maintain the initial active expression over generations. However, in transgenic cell lines, the

initial active transgene expression is usually suppressed after a certain period of time. The noticeable difference of gene transfer between these two situations is the transgene sequences transferred into the host genome. In mice, only the essential transgene transcription units are injected, stable integrants are selected phenotypically and/or transgene sequence detection by Southern hybridization. In contrast, stable clones in cell culture are almost exclusively isolated by antibiotic selection. This is the most common method used to isolate the genetically modified cells in transgenic cell line study. It is based on the belief that co-integration of the transgene and antibiotic resistant marker gene into the host cell line genome occurs, and cells with stable integration of the transgene sequences are rescued in the presence of a cytotoxic drug, while those non-transfected cells are selectively killed. This is done practically by cotransfection of a transgene expression plasmid and an antibiotic resistant gene plasmid, or by transfection of a transgene expression plasmid on which an antibiotic resistant gene is included. Expression of the transgenes isolated under such setting is normally unstable or even completely silenced after withdrawal of the antibiotic from the culture. Stable transgene expression occurs only at an extremely low frequency (Felgner et al., 1987; Gubin et al., 1999). There are two important differences between these two situations: the first is the sequences transferred, whether or not including sequences on the plasmid of prokaryotic origin; The second one is the way stable transfectants are selected (i.e.: the antibiotics.)

In order to find out the key factor leading to unstable transgene expression in cell lines, I first transfected only the essential transgene transcription units, the expression cassette or “cass”, to HeLa cells by excluding plasmid and antibiotic resistant gene sequences. This was based on the fact that prokaryotic sequences appeared to be the problem in the early trials of generating transgenic mice that could sustain uniform and high levels of transgene expression (Chada et al., 1985; Townes et al., 1985). Because administrating antibiotics to select stable clones was impossible, I therefore used the new “Sorting-Subcloning” approach to isolate stable transfectants. Owing to the special characteristics, namely easily detectable fluorescence of the reporter gene that I used in this study, I could easily adopt FACS analysis to select those EGFP positive cells. Time course analysis revealed unchanged homogenous expression after about 100 days (Fig. 17) after isolation of EGFP homogeneous clones. Integration of the essential

transcription sequences of the transgene alone indeed ensured a stable and uniform expression in cell lines. When compared to the control transfection with linearized plasmid, transfection with "cass" alone generated almost 2-fold more EGFP homogeneous clones. More clones with heterogeneous and silenced EGFP expression were generated after transfection with linearized plasmid followed by the "Sorting-Subcloning" approach. I concluded that the prokaryotic sequences including the antibiotic resistant gene sequences were responsible for the lower rate of acquisition of EGFP homogeneous clones since they were the only difference between the transfection with "cass" and linearied plasmids. The trouble caused by the prokaryotic sequence in transgenesis can be traced back to the early trials in transgenesis of mice. However, the effect observed here is not as strong as reported for mice. Mice showed high tissue-specific expression with the hybrid ß-globin gene expression construct with most of the plasmid sequences removed prior to injection; whereas most of the cloned globin genes that were not expressed were introduced together with prokaryotic vector sequences (Chada et al., 1985; Townes et al., 1985). The DNA sequences present within the plasmid backbone were blamed for interfering with proper expression of the ß-globin transgene in transgenic mice. Ever since then transgene mice are generated by injecting only the transgene expression sequences. Later studies suggested that the prokaryotic sequences could serve as active foci for gene silencing in the mammalian genome (Clark et al., 1997). Additionally, transfer of a PCR amplified fragment containing the CMV promoter, the luciferase gene and a polyadenylation signal showed reduced inflammatory activity in mice (Hofman et al., 2001)

Besides prokaryotic sequences, a second variable was involved in dissection of the key factor leading to unstable transgene expression in cell lines. This was the antibiotic selection, normally used in isolation of the transgenic cells, but not applied in generating transgenic mice. Since I did not perform a parallel time course for clones derived from transfection with the corresponding linearized plasmid by the same "Sorting-Subcloning" approach, I could not rule out the possibility that those clones would also express the EGFP transgene stably and uniformly over time. Therefore, I performed the experiment outlined in Fig. 18. Because transfection with pCMV-EGFP$^{WT/CpG-}$ in HeLa cells failed to generate stable clones with G418 selection, pEF-EGFP^{WT} was chosen to investigate the

effect of the method used to isolate stable clones on the stability of transgene expression. Stable homogeneous EGFP clones were obtained by the same transfection, but by different isolation approaches. Time course analysis suggested that clones isolated by the “Sorting-Subcloning” approach could maintain homogeneous EGFP expression more than 80 days; in comparison, clones generated by antibiotic selection were less stable, with EGFP silenced in a small amount of cells after two months in culture. Taken together, the antibiotic selection protocol used in stable clone selection clearly accounted for unstable transgene expression over time after withdrawal of the antibiotic.

It has been noticed that stable transgene expression occurs at very low frequency in the setting of antibiotic selection. It remains unclear whether selective growth conditions as such are required for the identification of transfected cells that permit long term expression of transferred genes (Felgner et al., 1987; Gubin et al., 1999). Moreover, antibiotic selection induces some deleterious effects such as growth inhibition (Gu et al., 1992; Kim et al., 1998), mutation in the antibiotic resistant gene, cell membrane permeability (Haber et al., 1981; Assaraf and Schimke, 1987), and morphological changes (Chung et al., 2000). The concentration of the antibiotic in the culture medium even had demonstrable effects on transgene expression within those antibiotic resistant cells (Schott et al., 1996). Additionally, the prokaryotic neomycin resistance gene was reported to act as a transcriptional silencer in eukaryotic cells. Surprisingly, the suppression did not even require transcription of the neo gene (Artelt et al., 1991).

During the process of antibiotic selection for stable clones, the antibiotic might force the cells to adopt a transcriptionally permissive structure and express the antibiotic resistance gene in order to survive the selection. It is also possible that expression of the antibiotic resistant gene is down-regulated to a level which is high enough to allow the cells to survive selection, but too low to enable detection of EGFP protein expression. One might expect colonies with the opposite-EGFP expression but not antibiotic expression. The large amount of EGFP negative foci formed on the Petri dish after the antibiotic selection process could plausibly account for this assumption. Indeed, reverse transcription PCR demonstrated the presence of mRNA transcribed from the hrGFP transgene in

hrGFP negative clones after the antibiotic selection process (Kirsch et al., 2003). After the selection pressure is removed, cells could possibly sense the environmental change and adapt to a different physiological state. The passive expression of the antibiotic resistant gene and/or the transgene would be engrossed by the overwhelming heterochromatin spreading triggered by host defenses or other unknown mechanisms. Therefore, even though transgene clones with high and homogeneous expression were obtained after the antibiotic selection, expression of the transgene could be affected due to the physiological change upon withdrawal of the antibiotic in culture.

In conclusion, the "Sorting-Subcloning" approach is an excellent strategy to obtain clones with sustained transgene expression over a long period of time. In contrast to antibiotic selection, this approach provides cells a consistent culture condition. FACSorting makes it possible to selectively collect EGFP positive cells after transfection. After the first round of FACSorting, EGFP expression in the majority of cells is likely transient due to nonintegration or the episomal construct. A large portion of cells will revert to a nontransfected state by loss or degradation of the transfected DNA fragment. Thus, a second round of FACSorting is necessary. High and homogeneous EGFP expression is a rare event after transfection, especially after transfection with the whole expression plasmid. However, FACSorting is able to pick up these rare events due to both its high sensitivity and throughput. Therefore, these EGFP expressing cells are isolated under natural growth conditions, without external force. After the two consecutive FACSortings EGFP positive cells have likely integrated the transgene construct and escaped the host defense to endow a high level expression. These cells could sustain high and homogeneous transgene expression. The EGFP transgene sequences are likely integrated into a big block of euchromatin region and "melted" into this region. This is different than the forced relaxed chromatin structure when the transgene is under antibiotic selection. Interestingly, when I administered G418 to clones that were transfected with the whole pEF-EGFP^{WT} plasmid and obtained by the "Sorting-Subcloning" approach (section 3.3.3), they showed G418 resistance during the whole time course (W. Liu, unpublished data). This result suggested that the neighbor antibiotic resistant gene was also actively expressed along with the EGFP gene.

Using a sorting approach to generate stable clones was sporadically reported, but no systematic study was done to investigate the underlying principles. This study reported for the first time a systematic investigation at the cell biological level. These results will facilitate transgenic studies in cell lines. Importantly, EGFP is shown to be sufficient as a marker of durable transfection events, and as a mechanism for monitoring transgene expression level in enriched or cloned populations of mammalian cells (Gubin et al., 1997). Therefore, it might become realistic to use EGFP to establish stable clones of other transgenes with long term sustained and uniform expression. This can be accomplished by cotransfection of EGFP and any other transgene of interest because of the general cointegration phenomenon, or by cloning the transgene of interest and EGFP into the same plasmid. Designing EGFP-tagging proteins or fusion proteins are also feasible. Transgene of interest can be cloned into a bicistronic constructs with EGFP downstream of the internal ribosomal entry site (IRES). Using bidirectional promoter to express the transgene of interest in parallel with EGFP is another alternative approach. The feasibility of isolating stable clones with EGFP as a sorting marker has already been projected by the long term G418 resistance showed by clones isolated by the “Sorting-Subcloning” approach as discussed above. Subcloning by limited dilution can even be done by seeding single cells into 96-well plates mechanically by the FACSorter. As an alternative, MACS (Magnetic assistant cell sorting) can be substituted to FACSorting because of its low cost and wide availability. In this case, any transgene of interest can be tagged with a cell surface marker. An antibody against cell surface marker can be used to separate the expressing and nonexpressing cells. Finally, the newly introduced “Sorting-Subcloning” approach will greatly benefit studies requiring stable and uniform transgene expression.

Furthermore, it will be interesting to find out if the copy number of the transgene integrated into the cell genome has any correlation to the observed stable transgene expression achieved by the “Sorting-Subcloning” approach, if the steady transgene expression is indeed inversely proportional to the transgene copy number. If this is the case, strategies need to be employed in order to ensure lower copy number integration of the transgene construct. Virus infection can deliver a single copy of the transgene into various cell genomes. Moreover, transposons can mediate precise integration of discrete pieces of DNA into a cell

genome (Ivics et al., 1997; Miskey et al., 2003). *Sleepy Beauty* and *Frog Prince* are two newly activated vertebrate transposons, and have been applied to gene transfer in several types of vertebrate cells. Importantly, transposition mediated gene transfer allows only the transgene expression cassette to be inserted into the cell genome, but the plasmid backbone is excluded, and is either degraded or lost in the cell. Thus, the chances to get stable clones might be increased not only due to the active integration process, but also due to omission of vector sequences.