CHAPTER 6
GENERAL DISCUSSION

Discussion

Globin genes are arranged in the order in which they are expressed during development i.e. 5'-ε-$^{G}\gamma^{A}\gamma$-δ-β-3'. Switching in expression between the different genes occurs during development, i.e. ε→γ in the embryo and γ→β at the time of birth. The regulation of the human β-globin locus has been the subject of intense investigations in the last decades. These have led to the identification of important regulatory elements, protein factors and mechanisms and given insight into the regulation, activation and expression of the human β-globin genes. Regulatory elements upstream of the gene locus, such as the LCR, as well as elements proximal to the genes, such as promoter, silencer and enhancer sequences, are required for the correct developmental expression of the globin genes (Stamatoyannopoulos and Grosveld, 2000).

In the 1980s, a number of studies using transgenic mice with human β-globin genes, showed very low, though developmentally regulated, levels of expression (Townes *et al.*, 1985 and Kollias *et al.*, 1986). Studies on deletions giving rise to Dutch and Hispanic thalassemias (van der Ploeg *et al.*, 1980 and Driscoll *et al.*, 1989), resulted in the identification of the Locus Control Region (LCR). The LCR consists of five DNase I hypersensitive sites that have since been shown to be important for copy-number dependent, integration-site independent and tissue-specific expression of globin genes in transgenic mouse assays (Grosveld *et al.*, 1987).

The LCR is located upstream of the globin genes, at a distance varying between 5 kb, for the ε-globin gene, and more than 50 kb for the β-globin gene. The LCR therefore exhibits long-range activation properties for the expression of the globin genes. Three models have been proposed as the basis for this long-range activation: an accessibility model, a scanning/tracking model and a looping model (Martin *et al.*, 1996; Tuan *et al.*, 1992; Kong *et al.*, 1997; Stamatoyannopoulos *et al.*, 1991; Epner *et al.*, 1992; Dillon *et al.*, 1993; Grosveld *et al.*, 1993 and Orkin *et al.*, 1990). Which model(s) is actually the correct one remains to be directly demonstrated, though recent data indicate that the looping model appears to be operating *in vivo* in mediating long-range LCR-gene interactions (Tolhuis *et al.* submitted).

A great deal of research has also been done in order to gain a better understanding of the molecular basis of LCR function, the role of each DNase I hypersensitive site, the identification and the role of the proximal regulatory elements of the genes and the interplay of the LCR and these regulatory elements (see introduction for further details and references). Chapters 4 and 5 of this thesis have presented two such studies and the implication of these studies on globin gene regulation will be discussed in this chapter.

To study the elements that play a role in the regulation, activation and expression of the globin genes, it is important to do so in the context of the intact human β-globin locus and to be able to manipulate the locus in its complete context. The use of artificial vectors, like YAC, PAC and BAC, which can carry large inserts, and the use of homologous recombination methods in manipulating these inserts in host yeast and bacterial cells, are very useful tools for these purposes.

The method of homologous recombination for manipulating large DNA constructs

Several methods for homologous recombination have been developed and used in the generation of constructs to study large genes and multiple gene loci (Yang *et al.*, 1997; Zang *et al.*, 1998; Pavan *et al.*, 1990 and Duff and Huxley, 1996). For studies on the human β-globin locus methods for homologous recombination to modify both YAC and PAC/BAC inserts have been described (Narayanan *et al.*, 1999; Peterson *et al.*, 1997 (review) and chapter 3) Here, I will focus on the methods used to modify PAC/BAC since we used a construct based on a PAC vector to manipulate the human β-globin locus (see Chapter 3).

PAC and BAC vectors are artificial chromosomes based on the P1 phage and the *Escherichia* coli F-factor, which is a low copy plasmid. Both types of vectors can be used to clone and maintain large DNA fragments (100-300 kb) in *E. coli* cells. In this way, large gene constructs and gene loci can be isolated and studied in their complete context.

All homologous recombination methods for manipulating BAC/PAC vectors are based on RecA or RecE/RecT mediated recombination. Yang *et al.* 1997 were the first to publish a general, simple method for the manipulation of BAC vectors, which could also be applied to PAC vectors. The method consists of two sequential steps: an integration step of a targeting vector bearing the desired modification and an excision step that leaves behind the desired modification in the insert. A recombination plasmid containing a RecA gene, a Tet^R marker and the DNA sequence homologous to the BAC containing the desired mutation, is introduced into a $RecA^-$ *E. coli* host strain containing the BAC which is to be modified. Upon integration of the recombination plasmid by a single homologous exchange, selection using Tet^R takes place. The clones with an insertion of the recombination plasmid are then subjected to a counter-selection step, also based on Tet^R . This counter-selection results in the loss of the recombination plasmid giving rise to the desired modification staying behind in the BAC/PAC insert in a fraction of the cell clones.

A second method for homologous recombination in BAC/PAC uses the RecE/RecT mediated recombination pathway in *E. coli* host strains (Zhang *et al.*, 1998). This method differs, besides the used recombination pathway, from the other two methods in the usage of a linear DNA fragment, instead of a recombination plasmid. A PCR fragment containing homologous sequences to the wild type BAC or PAC and the desired mutation and a selection marker are transformed into the *E. coli* host strain containing RecE/RecT with the wild type BAC/PAC. After introduction of the PCR fragment, homologous recombination takes place and the BAC/PAC is modified in a single step. However, the homologous recombination also results in the integration of the resistance marker. In most cases this is not desired and the resistance marker has to be removed. This is most often accomplished by the use of FRT sites and Flp-mediated recombination. Therefore, this technique is in essence a two-step recombination method and has been called ET-recombination.

The same group that first described ET-recombination later published a revision of the method (Muyrers *et al.*, 1999 and 2000), in which the addition of a second selection marker, SacB, circumvented the FRT/FLP recombinase step. During the first step of the new protocol, a PCR fragment containing both resistance markers and the homologous sequences to the BAC/PAC insert flanking both sides of the resistance markers, is introduced into the *E. coli* host strain with the wild type BAC/PAC. After selection for the homologous recombination event, a second PCR fragment is introduced, this fragment containing the same homology sequences but also the desired mutation of the BAC/PAC, so that the final modification can take place. The presence of the same homologous sequences in both PCR fragments results in the second recombination event and in the excision of the resistance markers and the insertion of the desired mutation.

The methods described above had been published prior to the completion of the method described in chapter 3. We, however, proceeded with the development of our method for homologous recombination because of a number of disadvantages presented by the other methods. The method of Yang *et al.* is almost identical to ours, however, this method shows very low efficiency of integration and correct recombination events. This is because excision is mediated through the use of Tet^R as counter-selection, which has been reported to work inefficiently (Podolsky *et al.*, 1996). We used the rpsL gene for counter-selection, which mediates either streptomycin resistance or sensitivity depending on the correct homologous recombination event and the number of copies of rpsL genes thus introduced into the host strain. The use of this counter-selection resulted in a much higher percentage of correct recombination events.

The ET-recombination has some advantages but also disadvantages. It has the advantage of using linear PCR fragments together with the standard ET-plasmid and thus there is no need for constructing a new recombination plasmid for each recombination event. Furthermore, when the presence of the resistance marker is not interfering, the method requires just one recombination step and is very fast. However, when loss of the resistance marker is required, the method takes two recombination steps with each step requiring a new (specific) PCR fragment to be transformed into *E. coli* cells containing the BAC/PAC that has to be modified, meaning that electrocompetent cells have to be produced twice in the same experiment. This makes the ET-recombination method more laborious and time-consuming. A second disadvantage is the use of the SacB gene as selection marker. This gene has been shown to be easily mutated, resulting in inefficient counter-selection and, as a consequence, high backgrounds. In our method, which consist of two steps, the recombination vector, including the resistance marker, is lost upon counter selection and no second step of transformation is

necessary, thus making our method less laborious and faster. Furthermore our counter-selection gives rise to lower background.

The choice of a PAC vector, instead of YAC, for the manipulation of the human β-globin locus and the use of homologous recombination method in our studies, is based on the advantages that the PAC system offers compared to YACs. YACs were the first to be used for studying large genes and multiple gene loci in their full context. They have the advantage over conventional plasmids and cosmids of being able to carry very large inserts (up to 1-2 Mb) which can be easily modified because of the high efficiency of endogenous homologous recombination in yeast. However, DNA fragments cloned into YAC vectors show genetic instability and it is difficult to obtain high yields of purified YAC DNA. Finally in transgenesis there is low frequency of YAC integration into the mouse genome, presumably due to shearing during microinjection of the (larger) YAC DNA inserts (Green *et al.*, 1991 and Schedl *et al.*, 1992). PAC and BAC vectors also have a large capacity of insert DNA (up to 200-250 kb) and homologous recombination in *E. coli* can also be used to manipulate the insert DNA. Furthermore PACs/BACs circumvent the disadvantages of the YAC as they can be treated as conventional plasmids in DNA preparations, resulting in high yields of intact purified DNA and exhibit high insert stability.

YAC, BAC, PAC and cosmid-based constructs have been used to study the regulation, activation and expression of the human β-globin gene locus in its complete context in transgenic mice (Peterson *et al.*, 1993; Bungert *et al.*, 1995; Gaensler *et al.*, 1993; Porcu *et al.*, 1997; Narayanan *et al.*, 1999; Kaufman *et al.*, 1999; Strouboulis *et al.*, 1992 and Chapters 4 and 5).

Studies on elements required for globin gene expression

The homologous recombination method in *E. coli* has been used to delete hypersensitive sites 2 and 3 of the LCR and the Enh and F elements located downstream of the $^{A}\gamma$-globin gene, in the context of the complete human β-globin locus. Transgenic mice were generated with these constructs and analysed (see Chapters 4 and 5). The analysis of these mice shows that these deletions have an effect on the regulation and expression of the human β-globin genes.

Deletion of HS2 showed a significant decrease in ε expression in all lines analysed and the occurrence of PEV in the foetal liver and adult blood stages in one line. Expression of γ- and β-globin genes was also affected but to a lesser degree than ε expression. The transgene in the PEV line had integrated in the middle of a chromosome arm and is thus in a domain that is not obviously heterochromatic. Deletion of HS3 resulted in the decrease of expression of all the globin genes, with the decrease becoming worse as development proceeds. All lines analysed for the HS3 deletion showed PEV, with one line having a peri-centromeric integration site and the second line an integration similar to that of the HS2 PEV-line.

The deletion of the Enh and F elements, previously postulated to be involved in the silencing of γ gene expression in the adult stage, did not have an effect on globin gene switching and expression patterns in the foetal and adult stages. However, at the embryonic stage an elevation of both ε and γ expression levels was observed in the absence of these elements.

These results again demonstrate the interplay of multiple regulatory elements in the correct developmental regulation of expression of globin genes. The results from the hypersensitive site deletion studies show that the LCR is not acting just as a holocomplex (Ellis *et al.*, 1993; Fraser *et al.*, 1993 and Dillon *et al.*, 1993) and that its function seems to be more complex affecting differently individual globin genes. The deletions of the Enh and F elements have indicated another example of a new class of developmental stage-specific regulatory elements affecting expression of all globin genes in the locus at that particular developmental stage (see also Calzolari *et al.*, 1999 and Liu *et al.*, 1998).

Although the LCR holocomplex is necessary for the position independent, high-level expression of all the human β-globin genes (Fraser *et al.*, 1993 and Milot *et al.*, 1996), the interaction of specific hypersensitive sites with the individual globin gene promoters seems to be important for the correct gene expression during development (Navas *et al.*, 1998; Navas *et al.*, 2002; Bungert *et al.*, 1998; Peterson *et al.*, 1996 and Gui and Dean, 2001). The studies described in Chapter 5 indicate that the deletion of a hypersensitive site can have both a general effect, as in the PEV observed in the expression of all globin genes in the HS3 deletion, and/or a gene-specific effect on globin gene

transcription, as is observed in the decrease or even complete absence of ε-globin expression in all the HS2 deletion lines.

A number of other studies have also reported gene-specific effects caused by the deletion of either HS2 or HS3 (Navas *et al.*, 1998; Navas *et al.*, 2002; Bungert *et al.*, 1998 and Peterson *et al.*, 1996). HS3 has been shown in two different studies to be important for γ expression in the foetal stage (Navas *et al.*, 1998 and Bungert *et al.*, 1999). The deletion of HS3 in the context of a 248 kb human β-globin YAC construct in transgenic mice showed a decrease in γ expression levels in the foetal, but not in the embryonic stage (Navas *et al.*, 1998). The same group recently reported that the expression of the γ gene in adult transgenic mice carrying the –117 HPFH mutation (Berry *et al.*, 1992) was abolished when they deleted HS3 in the context of the 248 kb YAC construct (Navas *et al.*, 2002). Finally, a comparison between a transgenic mouse line with a deletion of HS2 and a transgenic mouse line in which HS2 was replaced by HS3 (Bungert *et al.*, 1999) showed that while in the HS2 deletion line expression of all the globin genes was affected, in the HS3-substitution line expression of all the genes, except γ, remained affected. Expression of the γ genes was partially rescued by the replacement. The HS3 for HS2 replacement therefore partially rescued γ-globin gene expression suggesting an interaction between HS3 and γ-globin potentially effected by the extra copy of HS3 in the LCR in this construct.

Questions arise related to the differences on globin gene expression observed upon deletion of HS2 between the Bungert *et al*., 1995 study and our study. These probably can be explained by differences in the sizes of the HS2 core deletions used to make the deletion constructs (for further discussion on the differences see discussion in Chapter 5).

The studies presented give some indication that a specific interaction of HS3 is important for human γ-globin gene expression at the foetal stage. The reason why our deletion of HS3 did not show such a specific effect is probably because of the position effect variegation and the subsequent decrease in expression of all the globin genes. The PEV probably masks the effect of HS3 deletion on foetal γ gene expression and the specific effect of HS3 on the foetal γ expression is therefore not observed in our studies. This indicates that an effect on the general function of the LCR as a holocomplex overrules the effects of the individual hypersensitive sites. This however, does not appear to be the case with the HS2 deletion since, in the apparent absence of PEV in the embryonic yolk sac, we are able to see a specific effect on ε-globin gene expression, as well as a more generalised reduction in expression of all globin genes in all developmental stages. HS3 has also been suggested to be the hypersensitive site necessary for ε-globin gene expression, since deletion of HS3 in the context of a complete β-globin locus resulted in a specific decrease of ε expression in three transgenic studies (Milot *et al*, 1996; Navas *et al.*, 1998 and Peterson *et al.*, 1996). This reduction in ε expression, however, could also be due to the overall reduction of the globin gene expression observed in HS3 deletion lines. Since ε-globin is already expressed at low levels in the wild type transgenic lines compared to γ-globin, a general reduction of globin expression can already make ε expression difficult to detect. Indeed, recent evidence indicates that it is HS2 that is the prime hypersensitive site necessary for ε-globin expression (Gui and Dean, 2001). Our studies clearly show a decrease and even an absence of ε expression in all the HS2 deletion lines, whereas the other globin genes are hardly affected, except for the one line showing PEV in the foetal liver and adult stages. The study in which HS2 is replaced by HS3 in transgenic mice, shows an absence of ε expression, thus suggesting that HS3 cannot functionally replace HS2 for ε-globin expression. At the same time the hypersensitive site replacement could partly overcome the decrease of the γ expression, indicating that the HS3 used to replace HS2 is functional (Bungert *et al.*, 1999). Furthermore, a specific function in histone acetylation has been recently described for the core of HS2 (Gui and Dean, 2001). This function has been proposed to be necessary for normal levels of ε expression (Gui and Dean, 2001).

Some of the hypersensitive sites within the LCR holocomplex thus seem to have different specialised affinities for different globin gene promoters. This would suggest that the LCR has to change its conformation during development, such that the right hypersensitive site can interact with the globin gene promoter that it has a higher affinity for, at the right time during development. Such a conformational change hypothesis has already been put forward by Navas *et al.*, 1998 to explain why a

deletion of HS3 had an effect on γ expression during the foetal stage, but not during the embryonic stage.

The suggestion that hypersensitive sites have different affinities for the individual β-globin gene promoters could be addressed by comparing DNase I sensitivity and histone acetylation status in the promoter regions of individual genes, for example in comparing ε-and γ-globin promoters in the HS2-versus the HS3 deleted mice. Using ChIP assays, it has already been shown that peaks of histone hyperacetylation can be observed in activated globin promoters and the LCR (review Bulger *et al.*, 2002). Furthermore, it has been shown that the deletion of the LCR resulted in the loss of peak histone hyperacetylation on the β-globin promoter (Reik *et al.*, 1998). This assay can thus be used to correlate hypersensitive site activity and gene activity or the absence of gene activity with the deletion of specific hypersensitive sites.

The deletion studies on the Enh-F elements show that, besides the interaction of the LCR and the hypersensitive sites with the individual globin gene promoters, the presence of specific gene-proximal regulatory elements is also necessary for normal globin gene expression levels. Enh and F elements are two of four elements (Enh, F, O and P) identified in the $^{A}\gamma$-δ intergenic region. This region has been implicated as being important for the γ→β switch, since naturally-occurring deletions mapping within this region in HPFH and δβ-thalassemias, give rise to elevated γ-globin expression levels in the adult (Wood, 1993).

As described in Chapter 4, deletion of the Enh and F elements resulted in an increase in ε and γ gene expression in the embryonic stage, indicating that in the wild type situation these elements have a regulatory function in repressing the embryonic genes during development. However, no effect was observed on the γ→β switch, and no expression of the γ genes was detected in the adult. This argues that the elements do not have the previously suggested silencing effect on γ gene expression in the foetal to adult switch. Other studies, in which the role of the $^{A}\gamma$-δ region in the γ→β switch has been investigated (Liu *et al.*, 1998; Zhang *et al.*, 1997 and Calzolari *et al.*, 1999) also came to the same conclusion, in that this region is not involved in the γ→β switch and that no elevation of γ gene expression in the adult was seen when this region was deleted. For example, two studies that deleted either the Enh element alone, or a 12.5 kb fragment from the $^{A}\gamma$-δ region that included the Enh and F elements as well as other putative regulatory elements reported to map within this region, did not show any observable effect on globin gene expression (Liu *et al.*, 1998 and Zhang *et al.*, 1997). By contrast, the work described in Chapter 4 reports an effect on levels of embryonic ε and γ gene expression upon Enh and F deletion. Another study in which the O and P elements, located further downstream from the Enh and F elements in the $^{A}\gamma$-δ intergenic region, were deleted showed a decrease in β expression as well as PEV (Calzolari *et al.*, 1999).

These two studies indicate, that the Enh-F and O-P as pairs of elements have regulatory functions in the expression of the β-globin genes. The data from these studies are in contrast with the observation that the deletion of all the elements from the locus does not have an effect on globin gene expression (Zang *et al.*, 1997). The latter study, however, has only been published as a meeting abstract with no real data published. Comparison of the studies on the deletion of just the Enh element and our studies on the deletion of Enh and F suggests that the intervening sequences between Enh and F and/or the F element itself are the sequences which affect the ε and γ expression in the embryo.

Although the $^{A}\gamma$-δ region does not seem to be important for the γ→β switch, the studies by Calzolari *et al*. and the work presented in Chapter 4 suggest that elements that may be important for the transcriptional regulation of globin genes at specific stages of development map within this region.

Other developmental-stage specific elements globally affecting globin gene expression, include the 5' ε-globin silencer. This region 5' of the ε gene may be quite complex in its function, since its deletion resulted in a decrease of both ε and γ expression in the embryo (Liu *et al.*, 1997), whereas mutation studies on transcription factor binding sites within the silencer affected ε silencing, however, the effect these mutations had on γ expression was not determined (Raich *et al.*, 1995).

It has become clear from these and other studies, that for the regulated developmental expression of the globin locus, the coordinate action of many elements is required, including the LCR holocomplex, as well as the action of individual hypersensitive sites and gene-proximal regulatory elements. β-

globin gene regulation is complex and still many studies need to be done to be able to understand the regulation of the locus completely.

Stochastic gene activation; general or not?

The work presented in Chapter 2 on the basis of transcriptional activation of the murine globin gene loci, showed that a large percentage of cells (23%) had an imbalance in the α versus β nuclear transcription patterns, which was maintained in the cytoplasmic mRNA levels. These data indicate a stochastic basis of transcription for the murine globin loci. Furthermore, we observed that the α-locus is activated prior to the β-locus during erythroid cell differentiation and that the transcription patterns of the globin loci, once activated, become fixed after two or three cell cycles.

One question that arises from these studies is at what stage is the stochastic decision taken for a locus to transcribe. A simple scenario would be that the decision is taken at the actual transcriptional activation step, or at a step before that e.g. in the opening up of the chromatin domain. For a triple copy human β-globin PAC transgene, as described in chapter 2 of this thesis, we showed that the three copies of the human β-globin are expressed in an all-or-none fashion, suggesting that the decision is taken at a step prior to transcriptional activation itself.

Examples of decisions being taken either before or at the stages of transcriptional activation have come from previous studies on mono-allelically expressed genes, like the Il-4 and the Ly49 gene loci (Agarwal and Rao, 1998 and Tanamachi *et al.*, 2001).

In the case of a decision being taken after the chromatin opening level, the chromatin structure in which a gene resides becomes accessible prior to the start of transcription. In this case, the number of accessible alleles that will be expressed in the end will also depend on the availability of transcription factors, which may be limiting. For example, for the Il-4 gene it has been shown that only cells competent to express Il-4 show an open chromatin conformation at the Il-4 locus. Although all these cells are all able to express Il-4, as has been shown in re-stimulation assays, only a small fraction of accessible Il-4 alleles indeed express Il-4, probably because of transcription factor limitations, resulting in mono-allelic expression (Agarwal and Rao, 1998 and Hu-Li *et al.*, 2001).

If activation of a gene is determined at the chromatin opening level, this may also be reflected upon linked genes present in the same genecluster. Those will be co-expressed and will all show the same stochastic activation pattern. This has also been observed for the Il-4 gene which is linked to the Il-13 and Il-5 genes. Il-13 and Il-5 were indeed found to be co-expressed from the same allele as Il-4 (Agarwal and Rao, 1998 and Hu-Li *et al.*, 2001).

Stochastic decisions at the level of actual transcriptional activation would predict that linked genes in the same locus will be expressed independently to each other, resulting in all kinds of allelic expression patterns. Such a stochastic activation pattern has been observed for the Ly49 locus, in which the expression of linked genes was not coordinated and many combinations of allelic expression patterns were observed (Tanamachi *et al.*, 2001).

At which level stochastic decisions for the activation (or not) of the murine α- and the β-loci are made, has not yet been addressed, but can be speculated upon. It is clear that chromatin plays a role in the stochastic activation pattern of the globin gene loci. This is reflected in the calculated probabilities for expressing either an α- or a β-allele (93% and 88%, respectively) and in the activation of α- globin prior to β-globin in the total population of erythroid cells. The globin loci reside in different chromatin environments. The α-locus is found in a constitutively open chromatin domain rich in housekeeping genes, whereas the β-locus is in a closed chromatin domain in non-erythroid cells. Because of the open chromatin configuration the α-locus is probably more accessible to transcription factors, resulting in a lower threshold for transcriptional activation and thus a higher activation probability.

A question arises therefore, as to what is the level at which the decision for transcription (or not) is taken for the α-globin locus, since it does not appear to be at the chromatin opening step. It could be investigated whether the decision is taken at the level of transcription by checking, for example, whether the linked housekeeping genes are also co-expressed in a locus expressing α-globin, or whether they remain silent in a locus not expressing α-globin.

The stochastic decision for β-globin activation is most likely to be at the chromatin level. This assumption is based on the experiments with the triple copy human β-globin PAC transgene. Since the

mouse β-globin locus resembles the human β-globin locus closely it is likely that the determinants of the stochastic decision will be the same.

A recalculation of the activation probabilities of α- and β-globin, leaving out the early expressing α-cells, shows that once the β-globin locus is accessible for transcriptional activation both loci have a more or less equal activation probability (93% and 94%, respectively). This shows that the difference in chromatin status of the both loci has a clear effect on the activation probability for globin transcription in the total cell population.

Chromatin also seems to play a role in the fixation of stochastic choices. As described in Chapter 2, the decision to express a triple copy human β-globin PAC transgene is taken prior to the actual transcriptional activation step. The all-or-none decision in transcriptional activation observed with the triple copy human β-globin locus transgene is also observed in the expression of genes influenced by PEV, which also has a stochastic basis (Lock *et al.*, 1988).

A stochastic basis for gene activation has been reported before, but only for genes involved in very specific processes like lineage commitment, differentiation decisions and for genes involved in responses to specific stimuli. The studies on the mouse globin loci, however, indicate that the mechanism of stochastic gene activation could be a more general phenomenon for gene activation than previously assumed, since the globin genes are expressed in a cell type that is fully committed and the genes are not involved in specific cellular processes. The stochastic activation of the globin alleles is even a disadvantage for the erythroid cells, since cells with an imbalance in globin chain synthesis have a shorter life time (Weatherhall, 2001). It is therefore suggested from our data that stochastic gene activation is a general phenomenon and is not restricted to the activation of genes involved in specific cellular decisions.

After publication of our work on stochastic globin gene activation, two more studies (Ozbudak *et al.*, 2002 and Elowitz *et al.*, 2002) were published addressing the question whether stochastic gene activation has a general role in gene expression. Both groups address the question whether observed variations of gene expression levels between cells in a clonal cell population were caused by stochastic gene expression. This was determined by measuring "noise" in gene expression in bacterial cells using GFP-reporter genes under the control of promoters regulated by the Lac repressor. Elowitz *et al.*, made use of a bacterial system in which they expressed two identical copies of the same gene in the same cell, thus in the same intracellular environment. This way two types of "noise" could be determined, intrinsic and extrinsic noise. Intrinsic noise was defined as the difference in gene expression between the two identical genes under the exact same conditions, and extrinsic noise as the cell-to-cell variation in expression levels of each of the reporter genes. Obzudak *et al.*, used only one gene under control of a GFP reporter and thus could only measure total phenotypic noise. Both groups indeed observe noise in gene expression and changes in noise under different cellular conditions. They both conclude that stochasticity plays a role in giving rise to overall variations in gene expression between genotypically identical cells.

If gene activation has indeed a general stochastic nature, this will have wide consequences. For example the observations that genotypically identical bacterial cells have phenotypic differences (Elowitz *et al.*, 2002 and Obzudak *et al.*, 2002), because of stochastic differences in gene activation, could also hold true for eukaryotic cells. This could lead to diversity allowing evolution and selection. Furthermore, stochastic gene activation may also allow the fine-tuning of expression levels of gene cascades in response to changes in conditions. However, the stochastic basis of gene expression as a general phenomenon also implies that, sometimes, important genes will not be expressed, or genes that should not be activated will be expressed. To circumvent these problems nature must have evolved back-up systems. For instance, if a cell fails to express an essential gene, this cell will be arrested during the cell cycle at a cell cycle check-point and undergo apoptosis. The opposite, when expression is not desired, can be solved by the evolution of gene activation as a multi-step and multi-level process. Thus the existence of stochastic gene activation as a general principle, in combination with the required back-up systems, ensures that there is order in gene expression, but also that the necessary diversity within a population of cells or even within an organism can be generated and maintained.