Université de Montréal

Study of β-globin locus polymorphisms in hemoglobin switching using the YAC system

par
Gillian Louise Drury

Programme de biologie moléculaire
Faculté des études supérieures

Mémoire presenté à la Faculté des études supérieures
en vue de l'obtention du grade de maître ès sciences (M.Sc.)
en biologie moléculaire

Février, 2002

Université de Montréal
Faculté des études supérieures

Ce mémoire intitulé:
Study of β-globin locus polymorphisms in
hemoglobin switching using the YAC system

présenté par:
Gillian Louise Drury

a été évalué par un jury composé des personnes suivantes:

Dr. Lee Wall et Dre. Anne-Marie Mes-Masson
président-rapporteur

Dre. Marie Trudel
directrice de recherche

Dr. Pierre Belhumeur
membre du jury

Resumé:

Chez l'humain, le locus β-globine est composé de cinq gènes fonctionnels qui sont exprimés de façon séquentielle en commençant avec le gène embryonnaire en 5', suivi par les gènes foetaux, et en suite, les gènes adultes en 3'. Cette commutation des gènes β-globine est régulée au niveau transcriptionnel par une région de contrôle du locus (LCR) situé en 5' et par les séquences qui environnent les gènes. Il existe un individu où la commutation des gènes foetaux à adultes n'est pas survenue donc il exprime seulement les gènes foetaux avec un taux d'hémoglobine normale au stade adulte. Cet individu s'est avéré homozygote pour des polymorphismes au niveau du deuxième site d'hypersensibilité à la DNaseI du LCR (HS2) et dans le deuxième intron du gène foetal $^{G}\gamma$ ($^{G}\gamma$IVSII). Étant donné que des souris transgéniques porteuses de YAC contenant le locus β-globine humain entier démontrent une régulation normale des gènes globine humaines, l'introduction du HS2 et $^{G}\gamma$IVSII polymorphiques dans ce YAC pourrait élucider le mécanisme de commutation. L'expression des gènes globine dans des souris transgéniques produites à partir des YAC polymorphiques pourrait mettre en évidence l'importance du HS2 et du $^{G}\gamma$IVSII dans la commutation des gènes foetaux à adultes.

Mots clés: locus β-globine, HS2, gènes foetaux, commutation d'hémoglobine, YAC, recombinaison homologue.

Abstract:

In humans, the β-globin locus contains five functional genes which are sequentially expressed starting with the 5' embryonic gene, then the fetal genes and finally, the 3' adult genes. This process of gene switching is transcriptionally regulated by the locus control region (LCR) and β-like gene proximal sequences. There exists an individual who expresses fetal hemoglobin at normal levels and has not undergone the fetal to adult switch. The β-globin locus of this individual is homozygous for polymorphisms in the second DNaseI hypersensitive site (HS2) of the LCR and the second intron of the $^{G}\gamma$ fetal gene ($^{G}\gamma$IVSII). To elucidate the switching mechanism, attempts were made to introduce the polymorphic HS2 and $^{G}\gamma$IVSII into YACs containing the human β-globin locus. The wild-type YAC has been used to generate transgenic mice which display correct developmental regulation of the human β-like genes. Similarly, when these mutant YACs are used to produce transgenic mice, their patterns of human β-like globin expression may indicate whether HS2 and $^{G}\gamma$IVSII impact on transcriptional switching.

Keywords: β-globin locus, HS2, fetal genes, transcriptional switching, YAC, homologous recombination

Table of Contents:

List of Tables:

List of Figures:

Liste des sigles et abréviations:

BP-1	beta-protein 1
CAT	chloramphenicol acetyl transferase
CSBP	conserved sequence binding protein
EKLF	erythroid Krüppel like factor
FKLF	fetal Krüppel like factor
5-FOA	acide 5-fluoroorotique
HbA	hémoglobine adulte avec 2 chaines α et 2 chaines β
HbA2	hémoglobine adulte avec 2 chaines α et 2 chaines δ
HbF	hémoglobine foetale
HPFH	persistence héréditaire d'hémoglobine foetale
HS	site hypersensible à la DNaseI
HS1	premier HS en 5' des gènes β-globine
HS2	deuxième HS en 5' des gènes β-globine
HS3	troisième HS en 5' des gènes β-globine
HS4	quatrième HS en 5' des gènes β-globine
IVSII	deuxième intron
LCR	région de controle du locus
NF-E2	nuclear factor- erythroid 2
PFGE	électrophorèse en champ pulsé
PGK-neo	phosphoglycerate kinase- neomycin resistance gene
RBC	erythrocyte
SATB1	special AT binding protein-1
SSCP	polyporphisme de conformation d'ADN simple brin
SSP	stage selector protein
TAL1/SCL	stem cell leukemia factor
YAC	chromosome artificielle de levure
YIp	plasmide d'intégration dans la levure

List of Abbreviations:

BP-1	beta-protein-1
CAT	chloramphenicol acetyl transferase
CSBP	conserved sequence binding protein
EKLF	erythroid Krüppel like factor
FKLF	fetal Krüppel like factor
5-FOA	5-fluoroorotic acid
HbA	adult hemoglobin with 2α and 2β chains
HbA2	adult hemoglobin with 2α and 2δ chains
HbF	fetal hemoglobin
HPFH	hereditary persistence of fetal hemoglobin
HS	DNaseI hypersensitive site
HS1	first HS 5' of β-like genes
HS2	second HS 5' of β-like genes
HS3	third HS 5' of β-like genes
HS4	fourth HS 5' of β-like genes
IVSII	second intron
LCR	locus control region
NF-E2	nuclear factor- erythroid 2
PFGE	pulsed field gel electrophoresis
PGK-neo	phosphoglycerate kinase- neomycin resistance gene
RBC	red blood cell
SATB1	special AT binding protein-1
SSCP	single strand conformational polymorphism
SSP	stage selector protein
TAL1/SCL	stem cell leukemia factor
YAC	yeast artificial chromosome
YIp	yeast integrating plasmid

Dedication:

To my Grandparents...

Acknowledgments:

I would like to thank my supervisor, Dr. Marie Trudel for her invaluable advice and support through out my Masters course. I am also grateful to the members of the Trudel laboratory for their technical assistance and contribution to the environment in which I studied. In addition, I greatly appreciated and benefited from the advice provided by Dr. Martine Raymond and her laboratory for yeast work. Finally, I would like to thank family and friends for their tolerance and support over the past two years. It should also be noted that I am extremely grateful to the FCAR who have financially supported my Masters studies.

1. Introduction:

Section A: Genes and regulatory elements of the β-globin locus

1.1 β-like globin gene expression and arrangement in the human.

In humans, hemoglobin is the major protein constituent of RBCs at approximately 15g/dL. Hemoglobin is a heterotetramer composed of two α-type globin chains and two β-type globin chains. Different α-type and β-type chains are produced in a developmentally regulated fashion to give rise to embryonic, fetal and adult hemoglobins. During the 3rd to 8th weeks of gestation primitive erythropoiesis occurs in the yolk sac of the embryo. At this stage the erythrocytes are nucleated and the hemoglobin produced is composed of two ζ chains (α-like) and two ε chains (β-like), $\zeta_2\varepsilon_2$. Subsequently, at approximately 6 weeks of gestation, definitive erythropoiesis begins and the major site of RBC production shifts to the fetal liver. The hemoglobin produced is called HbF or $\alpha_2\gamma_2$ in the fetal liver environment. Prior to parturition, erythropoiesis occurs in the spleen in addition to the fetal liver. Finally at birth, the major site of erythropoiesis changes to the bone marrow and adult hemoglobin is expressed: HbA, $\alpha_2\beta_2$, is the main form of hemoglobin produced (98% of total hemoglobin); HbA2, $\alpha_2\delta_2$, is also expressed in the adult to a lesser extent, representing approximately 1-2% of adult hemoglobin.

The beta globin gene locus in humans is located on chromosome 11q15.5 and is composed of five homologous genes and one pseudogene (ψβ) which is not expressed at any stage of development. These genes are organized 5' ε-$^{G}\gamma$-$^{A}\gamma$-ψβ-δ-β 3' in the order they are expressed during development. Approximately 6 to 18kb upstream of the gene cluster lay four DNaseI hypersensitive sites which are only detectable in

erythroid cells, known individually 5' HS1 through 5' HS4 and known collectively as the locus control region (LCR). Increased DNaseI sensitivity is often associated with increased accessibility to the DNA due to transcription factor binding. And Indded, the β-globin LCR is important for high level expression of all β-like globin genes. Further upstream of the LCR are more DNaseI hypersensitive sites which are not exclusive to erythroid cells. There is also an evolutionarily conserved erythroid specific DNaseI hypersensitive site situated 20kb 3' to the β–globin gene. Finally, the entire β-globin gene locus is imbedded in a region encoding a set of evolutionarily conserved odorant receptor genes (ORGs) called human olfactory receptor (HOR) genes (9).

1.2 β-like globin gene expression and arrangement in the mouse.

In the mouse, a similar developmental regulation of the α and β globin genes occurs as in humans. According to Whitelaw *et al.* (83), primitive erythropoiesis occurs in the blood islands of the yolk sac in the mouse embryo from days 7.5 to 11.5. From day 9.5 onwards, the expression of x_2y_2 hemoglobin, composed of the embryonic α-like chain, ζ, and the embryonic β-like chain, βH1, diminishes. However, throughout this period the expression levels of the adult α gene and the ε^y gene product, y, actually surpass the expression levels of ζ and βH1. Definitive erythropoiesis begins in the fetal liver at approximately day 11.5 and the major hemoglobin produced is still α_2y_2. Erythropoiesis also begins in the fetal spleen during this period. Then at day 15.5, the spleen becomes the major location of erythropoiesis and the expression of the adult β–like genes, β^{min} and β^{maj} increases leading to the production of $\alpha_2\beta^{min}{}_2$ and

$\alpha_2\beta^{maj}{}_2$, or hemoglobin minor and hemoglobin major. Finally, the major site of erythropoiesis shifts permanently to the bone marrow just prior to birth.

The mouse β–globin locus is organized in a similar fashion to the human locus. It is composed of five functional genes (ε^y, βH0, βH1, β^{maj}, β^{min}) and two pseudogenes (βH2, βH3). The order of these genes is 5' ε^y-βH0-βH1-βH2-βH3-β^{maj}-β^{min} 3' and like the human locus, the 5' genes (ε^y and βH1) are expressed prior to the adult genes located in 3'. The mouse locus also has a 5' LCR which confers high level expression of the β–like globin genes. In fact, the mouse LCR is composed of four erythroid specific DNaseI hypersensitive sites which have substantial sequence homology to the 5' HSs in the human locus (33). The erythroid specific 3' DNaseI hypersensitive site is also present in the mouse locus as well as the non erythroid specific DNaseI HSs upstream of the LCR. Like the human locus, the mouse locus is also surrounded by ORGs named mouse olfactory receptor (MOR) genes. These ORGs are well conserved between the mouse and human (9).

1.3 Study of β–globin locus regulation:

The β–globin gene locus is a prototypical model for transcriptional regulation. Both the LCR and the β–like globin genes have been intensely studied. The LCR has been dissected to determine which elements confer enhancer effects, chromatin opening capabilities, developmental specificity etc., while the genes and their flanking sequences have been analysed to find enhancers, silencers, important promoter elements etc.. Following are the key studies using basic constructs which were used to extensively characterize the β–globin locus.

1.3.1 Plasmid constructs.

Many plasmid constructs have been used to study β–like globin gene regulation. Since the LCR measures approximately 20kb and is therefore too large to be studied in its entirety using plasmids, truncated LCR constructs have been generated. In particular, there are two major truncated LCRs used in these constructs, in order to obtain sufficient expression of reporter gene(s) in cell lines and in transgenic animals. There is the 8kb miniLCR composed of the HS4 core and sequences in 3', the HS3 core and sequences in 5', the HS2 core and sequences in 5', and the HS1 core and large amounts of flanking sequence on both sides. Secondly there is the 2.5kb microLCR containing the cores of all four erythroid specific HS with very little flanking sequence. These constructs are usually linked to a β–like globin gene promoter and either a reporter gene and/or a β–like globin gene (22). In addition, there is the β-globin minilocus developped by the Grosveld group. This construct contains all of the HSs as well as their flanking sequences and measures 6.5kb. In addition, the β-globin and neo genes are linked to this truncated LCR construct (13.3kb total length) (14). All of these constructs were initially believed to confer position independent expression of linked transgenes (14, 22)

1.3.2 The 150kb and 248kb YACs.

The 150kb YAC contains approximately 144kb of human genomic DNA, including the entire β–globin locus, in a pYAC4 vector. The 150kb YAC contains approximately 30kb of sequence 5' of the LCR and 40kb of flanking sequence 3' of the

adult β–globin gene. It has been well characterized to ensure that no rearrangement events had occurred within the 144kb segment (25). Transgenic mice carrying this 150kb YAC express the human β–like globin genes at levels similar to the endogenous murine β–like globin genes. In addition, the human genes are appropriately regulated during the fetal to adult switch. The mouse βH1 to β^{min} and β^{maj} and the human γ to β switches occur at approximately day 12 in the 150kb YAC transgenic mice (26). The fact that the human fetal and adult genes switch at the same time as the endogenous murine genes indicates that globin gene switching is regulated by factors in trans which act similarly in the human and in the mouse. The temporal regulation of these factors in the mouse causes the human locus to switch at the same time as the endogenous murine locus. Unfortunately, the characterisation of the ε^{y} to βH1and the ε to γ switches have not been well documented in this 150kb YAC transgenic mouse model. Although the human embryonic gene is silenced prior to the fetal genes, fetal genes are very well expressed compared to the ε gene during erythropoiesis in the yolk sac of these transgenic mice rendering it impossible to determine when the embryonic to fetal transcriptional switch occurs (26). A second YAC, measuring 248 kb and containing the β–globin locus was also isolated and characterized. Again, the vector used was pYAC4 and the insert size was approximately 240kb. This YAC has a similar 5' breakpoint to the 150kb YAC but has 140kb of flanking sequence 3' of the β–globin gene. This YAC has been shown to confer onto the human transgenes a similar level of expression to the endogenous genes in the mouse locus and allow for the same developmental regulation of the β–like globin genes as the 150kb YAC when used to generate transgenic mice (62). However, there is at least a two fold variation in

expression levels of the human globin genes per transgene copy between transgenic mouse lines (26, 58). This renders it difficult to assess subtle differences in β-globin gene expression in mutated or deleted YAC constructs.

1.3.3 The linked 70kb cosmid.

Two 35kb cosmids, one containing the LCR and the ε gene, the second containing the $^{G}\gamma$ through β genes, were ligated. In the process a 15 base pair linker was introduced at the joint between the two cosmids. The ligated cosmids were used to generate transgenic mice which displayed normal developmental regulation of the β–like globin genes. However, the relative expression of the human fetal genes was greater than embryonic genes in the yolk sac as was previously demonstrated in the YAC transgenic mice. In addition, the transgene was expressed at best at approximately 60% of the endogenous murine locus and some lines were subject to position effect variegation of the transgene. This may indicate that some of the 5' and 3' flanking sequences (especially the 3' DNaseI HSs) included in the YAC constructs, but not present in the ligated cosmid construct are required for an appropriate expression level of the human β–like globin genes (73).

1.4 Study of the LCR as a unit and as DNaseI HS subunits.

The β-globin locus has been studied using the previously mentioned systems as well as other experimental models. Most studies attempt to dissect the function of the individual components of the LCR and the gene proximal elements. In particular, the DNaseI hypersensitive sites upstream of the ε gene were found in 1984 by Tuan *et al.*.

At first, the HS1 and some minor DNaseI hypersensitive sites between HS1 and ε were described in K562 cells, a human leukemic cell line which expresses the embryonic and fetal globin genes (80). Subsequently the three other erythroid specific DNaseI HS upstream of HS1 as well as the non erythroid specific HS5 were discovered (78, 81). The importance of the four erythroid specific 5' HSs for high level expression of the β–like globin genes led to this region being named the LCR. The importance of the LCR in globin gene expression was supported by the observation that Hispanic (γδβ) thalassemia patients who have a deletion encompassing HS2 through 5 cannot produce high levels of any of the β–like globin genes (21). In addition, constructs containing the HS components could drive the high level, copy number dependent, stage specific expression of globin transgenes in erythroid cell lines and in transgenic animals. Subsequently the roles of the individual HSs and flanking sequences were investigated thoroughly. The results from the experiments described below are listed in table 1.1.

1.4.1 HS1.

The 5' HS1 is located approximately 6kb upstream of the ε gene. Its effect on β-like globin gene transcription is subtle. In fact, humans with deletions encompassing HS1 have normal β-like globin gene regulation. HS1 has evolutionarily conserved binding sites for GATA-1 which constitutes the core. GATA-1 and 2 are erythroid specific transcription factors containing two zinc fingers which serve in DNA binding. It also now seems quite likely that in the erythroid environment, the GATAs require cofactors for transcriptional activation (77). There are also several blocks of well conserved sequence which extend 500bp 5' of the core and 200bp 3' of the core (33).

Based on a study performed in transgenic mice containing deleted miniLCR constructs linked to the human β-globin gene, HS1 has a minimal impact on the expression of the adult gene. The removal of HS1 from the construct had little effect on the expression of the transgene while the sole presence of HS1 only allowed negligible expression of the transgene (14). However, a study using the linked cosmid construct described previously, found that deletion of a 1.7kb fragment, containing the core as well as all conserved 5' flanking sequences and all but one conserved 3' sequence, gave variable expression of the human β-like globin genes in transgenic mice. Transgenes which integrated close to centromeres, heterochromatic regions, were found to have very weak expression of all β-like globin genes per transgene copy. The 5' HS1 deleted cosmids which integrated elsewhere in the genome were not subject to this decrease in transcription (52). However, the deletion of a 2.3kb fragment of HS1 encompassing all conserved sequences in the endogenous murine locus does not significantly alter the level or developmental specificity of β-like globin gene transcription (6). All of the previous results would be compatible with HS1 playing an important role in allowing position independent expression of the β-like globin genes but that this activity is unnecessary due to the euchromatic location of the endogenous locus in mice and humans.

1.4.2 HS2.

The 5' HS2 of the LCR resides approximately 10kb upstream of the ε gene. The core of this HS site measures 375bp and includes binding sites for many erythroid specific and more broadly expressed transcription factors, including: Sp-1, Ap-1, NF-

E2, GATA-1 and 2, USF, YY1, EKLF or FKLF and FKLF-2, TAL1/SCL and HS2NF5 reviewed in (33) and (13). Ap-1, Sp-1, USF, YY1 and TAL1/SCL are not uniquely expressed in erythroid tissues while EKLF, the FKLFs and GATA-1 are erythroid specific. It is interesting to note that NF-E2 and Ap-1 are related to one another and EKLF, the FKLFs, YY1 and Sp-1 belong to the Krüppel family of zinc finger proteins reviewed in (33). However, HS2 has extremely well evolutionarily conserved sequences extending over a 700bp region which starts approximately 200bp 5' of the core and ends 100bp past the core in the 3' direction. It is known that HOXB2 binds the region 3' of the core. Therefore, it is possible that transcription factors could bind conserved regions outside of the core and contribute to HS2 function. The HS2 also appears to be more sensitive to DNaseI digestion during embryonic and fetal stages (84) implying that open chromatin structure and increased transcription factor binding occur during these stages. This observation also raises the possibility that HS2 would be more involved in the regulation of the ε and γ genes than the rest of the LCR.

1.4.2.1 Enhancer function of HS2.

Most importantly HS2 is the only 5' HS to have a demonstrated classical enhancer function. This was tested by placing HS2 upstream, downstream, in different orientations and at different distances from a CAT reporter gene driven by an ε promoter in K562 cells. The transcription observed is initiated at HS2 in the direction of the ε promoter regardless of orientation, position and distance (36). It is possible that the enhancer activity of HS2 is linked to transcription in this assay since it would allow increased accessibility of the ε promoter region to transcription factors. In

this experiment, most of the enhancer activity of the HS2 depends upon two immediately adjacent Ap-1/NF-E2 sites. Truncation of HS2 at the tandem Ap-1/NF-E2 sites or the replacement of these sites by three GATA-1 binding sites completely abolishes HS2 enhancer function in K562 cells and in MEL cells (a mouse erythroleukemia cell line which expresses the adult β–globin genes) (13). In the same study, it appears that the truncation of HOXB2 binding sites 3' of the core results in less enhancer activity in embryonic-fetal stage K562 cells but has no effect in adult stage MEL cells. Surprisingly, the truncation of the USF and YY1 sites in addition to the HOXB2 sites allows for very good expression of the CAT reporter gene in K562 cells but not in MEL cells. With the caveat that this system may not include the sequences required for the proper arrangement of HS2 and that the cellular environment may not be equivalent to the human *in vivo* situation, this may indicate that HOXB2 may be an activator and USF and/or YY1 may act as a repressor of the ε gene in the fetal stages of erythropoiesis.

More recently, it has been found that the tandem NF-E2 sites, as well as other regulatory elements in HS2, are responsible for impeding the silencing of HS2 linked transgenes in stable transfection assays. In the K562 fetal stage erythroid cell line, the stable expression of a HS2 enhanced $^{G}\gamma$ gene occurred more frequently if the NF-E2 sites in the HS2 core were functional. The stable expression of the integrated $^{G}\gamma$ transgene in K562 cells was also linked to a lack ofphysical association between the transgene and heterochromatic regions of genomic DNA. This nuclear organisation was also dependent on the tandem NF-E2 sites in the HS2 core (23).

1.4.2.2 Deletion of HS2 in the linked cosmid model.

The expression of a cosmid construct with a 742bp deletion in the HS2 region, which encompassed all conserved sequences 5' of the HS2 core but not the HOXB2 sites and GATA sites 3' of HS2, was analysed in transgenic mice. The expression of the β-like globin genes was developmentally regulated but was expressed at a level less than 25% of the normal expression per transgene copy for all the genes. This would indicate that HS2 is critical for high level expression of all β-like globin genes at all stages of development. It was also found that the transgenic mouse lines which expressed the lowest levels of the β-like globin genes had transgenes integrated near centromeres, suggesting that HS2 is also important for position independent expression of the β-like globin genes (52).

1.4.2.3 The YAC HS2 constructs.

In the YAC system, several deletions and replacements of HS2 have also been performed. Unfortunately, most of the results in this system are contradictory. One first deletion was carried out in the 248kb YAC and encompassed a 1.9kb HS2 fragment which has the same 5' boundary as the core but which extends a further 1.6kb in the 3' direction, obliterating all of the GATA-1 and HOXB2 binding sites 3' of the core. When this HS2 deleted YAC was used to generate one single copy transgenic mouse line, there was very little effect observed. The transcriptional switches were normal but the expression of all the β-like globin genes was reduced to 60% of wild type levels per transgene copy. These results are not entirely convincing since there is up to a two fold variation in the expression of the YAC β–like globin genes in the two

control mice produced (58, 59). YACs missing components of the LCR may be more vulnerable to such position effects. Since only one deleted HS2 YAC line was generated, it cannot be concluded that the HS2 deletion is the cause of reduced β-like globin gene expression.

The other HS2 deletion was performed in the 150kb YAC and in this case, only the 375bp core was deleted. Surprisingly, YAC transgenic mice generated with the smaller HS2 core deletion, had a much greater reduction of the β–like globin gene expression. Transgenic mouse lines carrying a single copy of the HS2 core deleted YAC have a pronounced reduction in β–like globin gene production. The expression of the embryonic gene is virtually undetectable while the γ and β genes do not attain more than 30% and 20% of wild type expression levels per transgene copy, respectively. The authors suggest that such a small deletion may act as a dominant negative and thus have a stronger effect on β–like globin gene expression than a larger deletion. A small deletion could have dominant negative properties by maintaining the structure of the HS2 and interactions with other components of the LCR and the β–like globin genes without recruiting the transcription factors required for its enhancer function (11).

The second type of experiment performed in the YAC system is the replacement of the HS2 core by the HS3 core. More specifically, in the 150kb YAC, the 375bp HS2 core was replaced by the 220bp HS3 core and the resulting YAC was used to generate three single copy transgenic mouse lines. These mouse lines had much the same pattern of expression of the human β–like globin genes as the HS2 core deleted mice: There was very little expression (at most 30% of wild type per transgene copy) of the

embryonic and adult globin genes. However, the fetal genes were expressed at approximately 60% of wild type levels per transgene copy, but in the day 9.5 embryonic yolk sac only. In contrast to the HS2 core deletion where HS formation is abolished, the formation of the other DNaseI hypersensitive sites of the LCR is unaffected in the transgene of mice where the core of HS2 is substituted by HS3. This indicates that HS3 can functionally replace HS2 to increase the chromatin accessibility of the remaining HSs of the LCR and can partially rescue the expression of fetal genes during primitive erythropoiesis in the yolk sac (11).

1.4.2.4 HS2 deletion of the endogenous mouse locus.

The HS2 has also been deleted from the mouse endogenous locus. This was done by replacing a 1.1kb fragment of HS2, which encompasses the core and 3' flanking sequences that bind GATA-1 and HOXB2, with a *PGK-neo* cassette in ES cells. The *PGK-neo* cassette was removed by FLP recombinase, and the resulting ES cells were used to generate mouse lines. These mice have a reasonably normal developmental pattern of expression of the endogenous β–like globin genes. A reduction in expression is not perceptible for the ε^y and βH1 genes, and the β^{maj} and β^{min} are transcribed at about 70% of normal levels (20). The similarity of the mouse and human HS2 sequence would suggest that this region would also have a conserved function. The results obtained by deleting a large segment of HS2 in the mouse are reasonably similar to those obtained for the larger deletion of HS2 done in the 248kb

YAC. This indicates that the HS2 may also have a conserved function in mice and humans. The mild changes in expression of the globin genes also point to larger deletions of HS in the LCR being less disruptive than core deletions.

1.4.3 HS3:

The HS3 of the 5' LCR, like HS2 extends over a large region and binds many transcription factors. Most of these transcription factor binding sites are located in a 225bp core region, approximately 13kb 5' of the ε gene, which is extremely well conserved between mouse and human and which functions as the entire HS3 in *in vitro* assays. However, there is substantial homology over a 900bp region which surrounds the core. Located immediately 5' to the core is a Ap-1/NF-E2 binding site. In addition, further 5' there are two YY1 binding sites and a CSBP-2 binding site. In the region 3' of the core, there is a YY1 site which is immediately adjacent to the core, as well as three YY1 sites and a Sp-1 site which are positioned further downstream. There are also several very well conserved sequences 3' of HS3 which could potentially be bound by transcription factors and contribute to HS3 function (33). It has been suggested that the HS3 contributes predominantly to adult globin gene expression since the core is more sensitive to DNaseI during this stage of development (84).

There is also a minor HS approximately 1kb upstream of the HS3 core. This HS has several well conserved sequence segments including a conserved consensus sequence for GATA-1. This site has not been deleted in any YAC constructs or in the mouse locus and has not been subject to extensive study (33).

1.4.3.1 Deletion of HS3 in the linked cosmid model.

A 1.4kb fragment encompassing the HS3 core as well as 3' flanking sequences was deleted from the linked cosmid construct. Sequences 5', known to bind CSBP-2 and YY1, were not deleted from the cosmid. This construct was used to generate four transgenic mouse lines which had reasonably similar phenotypes. Globin gene switching was unaffected by the deletion of HS3 but the level of expression of all the genes was approximately a third of normal levels per transgene copy. One of the lines expressed the globin genes at lower levels, but this was easily explained since the transgene had integrated in a centromeric region (52). This indicates that the HS3 has a role in increasing the expression of all β-like globin genes through out development. Surprisingly, these results are very similar to those obtained with the HS2 deleted linked cosmid. Unlike HS2, HS3 does not have a demonstrated classical enhancer capacity in *in vitro* systems (78), but seems to be as important as HS2 *in vivo*.

1.4.3.2 HS3 deletion of the YACs.

The HS3 has been studied using the YAC system in much the same manner as HS2. First, there was a large, 2.3kb deletion of HS3 performed in the 248kb YAC. This deletion removes all known transcription factor binding sites as well as all known conserved sequences surrounding HS3. This HS3 deleted YAC was used to generate two transgenic mouse lines containing 2 and 3 transgene copies each. The previously mentioned studies of HS deletions have all reported β-like globin gene expression in at least one single copy transgenic mouse line. It is possible that insertion of more than one transgene copy could influence expression by the presence of multiple LCRs in the

same region and by titrating regulatory factors present in exceedingly small amounts. These transgenic mouse lines expressed the human ε gene at no more than 30% of normal levels per transgene copy in the yolk sac but γ production in the yolk sac increased by approximately 80%. The γ expression does not seem affected during fetal stages but the adult β–globin gene expression is reduced by 30% (59). This would suggest that HS3 is absolutely necessary for adequate expression of the embryonic genes but that γ and β do not require HS3 for high level expression or transcriptional switching.

The HS3 core deletion was done in the 150kb YAC and the 248kb YAC. In both cases, this deletion spans the 225bp fragment which was previously defined as the core. This HS3 core deleted 248kb YAC was used to generate transgenic mice carrying between one and three copies of the transgene. These mice had severely impaired expression (less than 20% of wild type per transgene copy) of ε in the embryonic yolk sac, however γ expression in the yolk sac was normal. In contrast, γ expression in the fetal liver was almost absent at 12.5% of wild type levels per transgene copy, and adult β–globin gene expression was on average 25% of wild type levels per transgene copy (55). The HS3 core deleted 150kb YAC produced two lines of transgenic mice with one transgene copy. These mice had very little expression (less than 15% of wild type levels per transgene copy) of the human β–like globin genes at all stages of development including the expression of γ in the yolk sac (10). The differences in expression of γ in the yolk sac between the 150kb and the 248kb HS3 core deletion may be due to error in measurement since expression was evaluated by less precise RT-PCR in the 150kb YAC in comparison to RNase protection assays for the 248kb YAC

study. The drastic effect of the HS3 core deletion on γ and β expression is not observed with the larger HS3 deletion which is explained, like with the HS2 deletions, by dominant negative interactions of the HS3 core deleted LCR. However, ε expression is severely diminished with all HS3 deletions in the YAC system, indicating that this HS is required for ε expression. The variation in expression levels were significant in the HS3 core deleted mouse lines in both studies. This would indicate that HS3 is also important for position independence of transgene expression for the fetal and adult genes as was demonstrated in the cosmid model. It is also interesting that in the YAC model, the fetal and adult genes are not as severely affected by HS3 deletion as in the linked cosmid model which may be due to different deletion sizes.

The HS4 core has been used to substitute the HS3 core in the 150kb YAC. This construct was used to generate three mouse lines containing one transgene copy. Most interestingly, the presence of HS4 in this system confers poor expression of all globin genes at all stages of development, similar to the HS3 core deleted 150kb YAC except for the γ genes during erythropoiesis in the yolk sac (10). Coincidentally, this phenotype matches the HS3 core deletion in the 248kb YAC exactly but it still remains clear that HS4 cannot completely substitute HS3 function.

1.4.3.3 HS3 deletion of the mouse endogenous locus.

The HS3 has also been deleted from the endogenous mouse locus. In ES cells, a *PGK-neo* cassette was used to replace a 2.3kb fragment of HS3 which included the 5' flanking YY1, CSBP-2 and Ap-1/NF-E2 sites, the entire core and some of the 3' YY1 binding sites. It should be noted that there are sequences 3' of the deletion which are

known to bind YY1 and Sp-1, as well as sequences which are well conserved evolutionarily and may contribute to HS3 function (33). When the *PGK-neo* cassette was removed with FLP recombinase, the ES cells were used to generate HS3 mutant mouse lines. These mice were found to have very subtle phenotypes since the expression of the ε^y and βH1 genes was only reduced only by approximately 10% compared to wild type levels and the expression of the adult β genes was reduced by approximately 30% (34). These results are in contrast to the YAC experiments where HS3 was indispensable for ε production. If the human and mouse HS3 have similar functions, then perhaps sequences 3' of HS3, removed in the large YAC deletion but present in the HS3 deletion of the mouse locus, are important for ε expression.

1.4.4 HS4:

The extreme 5' erythroid specific DNaseI hypersensitive site is located approximately 18kb 5' of the ε gene. Unlike the other HSs, HS4 has a 280bp core which is not flanked by well conserved sequences or transcription factor binding sites. The only proteins known to bind the core are GATA-1, NF-E2, Sp-1, TEF2 and Ap-1. However there are well conserved sequences within the core which probably bind transcription factors which have not yet been identified (33). Similarly to HS1 and 3, HS4 does not increase transcription when placed upstream of a CAT reporter gene driven by an ε promoter (78). This result was confirmed by deletion experiments using β-globin minilocus transgenic mice where HS4 impacted minimally on transgene expression (14). However, other groups have found and that HS4 has a significant

enhancer function using the same β-globin minilocus transgenic mice (24, 63). These conflicting results are unfortunately not clarified by further studies in the YAC and cosmid systems.

1.4.4.1 The HS4 deletion of the linked cosmid construct

A 875bp fragment which included the entire 280bp core was removed. Three transgenic mouse lines were obtained with this transgene and all of them had completely normal expression of the human β-like globin genes per transgene copy (52). This result is concordant with the *in vitro* studies where HS4 did not impact on transcription of the β-like globin genes. Unfortunately, the location of transgene integration was not verified, so no conclusions can be drawn with respect to the effect of HS4 on position independent expression of the transgene. It could be that all of these mouse lines had transgenes inserted in euchromatic regions which would allow expression of the human β-like globin genes regardless of position effects caused by the deletion of HS4.

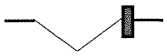
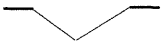
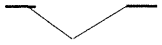
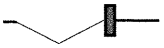
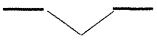
1.4.4.2 HS4 deletions in the YAC system.

The HS4 has been deleted in both the 150kb and 248kb YACs. These YACs both used the same deletion of the 280bp core and were used to generate two lines of single copy transgenics in the case of the 150kb YAC and four lines containing between 1 and 4 copies of the transgene for the 248kb YAC. The expression of all of the β–like globin genes was abolished (less than 10% of wild type levels per transgene copy) in the case of the HS4 core deleted 150kb YAC (10). This indicates a

requirement for HS4 through out development. In the case of the 248kb YAC, the deletion of HS4 allows for normal expression of ε and γ in the yolk sac, but causes a severe decline (to approximately 30% of wild type expression per transgene copy) in expression of γ and β in the fetal liver and adult blood (54) indicating that HS4 is only required during definitive erythropoiesis. The only features which were different between these two experiments were the number of integrated transgenes, the measurement of RNA levels by RT-PCR for the former and by RNase protection assays in the latter experiment, and the 3' flanking region in the YAC constructs. Thus the role of HS4 in β–like globin gene regulation seems to be important during definitive erythropoiesis but remains ambiguous in the context of primitive erythropoiesis.

Nevertheless, the YAC results are in sharp contrast to the results obtained in the linked cosmid model. The main difference between these experiments is the size of the deletion. In support of this notion, another study involving a large deletion (2.8kb) of the HS4 in the endogenous mouse locus found a 20% reduction in β-like globin gene expression (6). Though HS4 is not flanked by conserved sequences, maybe there are important elements of conserved function contained in this region which have not been identified. This could explain the large dominant negative effect observed in HS4 core deletions of the YAC and the relatively normal expression of the β-like globin genes if the flanking sequence is also removed as was seen with the mouse and cosmid HS4 deletions.

Table 1.1: Summary of LCR deletion experiments.

		ε or βH1 embryonic stage (YS)	γ or ε^y embryonic stage (FL)	β or β^{min} and β^{maj} adult stage
HS1				
Δ1.7kb linked cosmid:		100%	100%	100%
Δ2.3kb endogenous murine:		100%	100%	100%
HS2				
Δ742bp linked cosmid:		25%	25%	25%
Δ375bp 150kb YAC:		0%	30%	20%
Δ1.9kb 248kb YAC:		60%	60%	60%
Δ1.1kb endogenous murine:		100%	100%	70%
HS3				
Δ1.4kb linked cosmid:		30%	30%	30%
Δ220bp 150kb YAC:		15%	15%	15%
Δ220bp 248kb YAC:		20%	100% in YS 13% in FL	25%
Δ2.3kb 248kb YAC:		30%	180% in YS 100% in FL	70%
Δ2.3kb endogenous murine:		90%	100%	70%
HS4				
Δ875bp linked cosmid:		100%	100%	100%
Δ280bp 150kb YAC:		100%	100% in YS 30% in FL	30%
Δ280bp 248kb YAC:		10%	10%	10%
Δ2.3kb endogenous murine:		80%	80%	80%

Expression levels of β-like globin genes are indicated for appropriate developmental stage when it is not specified otherwise. YS and FL indicate expression in the embryonic yolk sac and fetal liver respectively. The HS cores are represented by dark shaded boxes while flanking sequences are in light shaded boxes.

1.5 Regulation of the individual β-like globin genes.

It should be noted that all of the β-like globin genes share some similarities. For example, since these genes have arisen through gene duplication, they all have the same arrangement of three exons and two introns. The exons are relatively similar in size, ranging from 91 to 222 nucleotides. However, the intron sizes are quite different: The first intron of the β-like globin genes is relatively small (on average 126 nucleotides) while the second intron spans almost 900 nucleotides. The size of the second intron enables it to accomodate regulatory sequences. The β-like globin genes all have regulatory sequences upstream of their promoters and some have 3' enhancers as well.

The LCR was once thought to confer high level expression and developmental regulation of the β-like globin genes. Early studies which generated transgenic mice with human β-like gene constructs lacking the LCR displayed invariably low expression of the transgene for example (49). However, the developmental patterns of expression of these same constructs was relatively normal. It has also been demonstrated using the endogenous mouse locus and in the 150kb YAC, that the removal of the 5' LCR diminishes globin gene expression drastically (to 1-4% of wild type) but does not impact on the transcriptional switches or on the histone acetylation and chromatin accessibility of the locus (5, 19, 68). Thus the LCR serves to increase transcription of the β-like globin genes but the developmental specificity of globin gene expression is determined by gene proximal sequences. Even though, it is still possible for the LCR to participate in this developmental control of the β-globin genes, it would appear that the gene proximal sequences are of primary importance.

1.5.1 The epsilon gene.

1.5.1.1 Factors which bind upstream of ε.

The region 5' to the epsilon gene has been phylogenetically footprinted like the LCR to discern regions which are evolutionarily conserved and thus likely to be important to ε expression. In a 2kb fragment 5' to the ε cap site (start of ε transcript), there are 21 conserved regions of which 10 bind YY1. GATA-1 and CP1 were found to bind many conserved non-consensus sequences in this 2kb region. Oct-1 also binds several consensus sites in this 2kb region while, Sp-1 appears to bind two adjacent locations approximately 100bp upstream of the cap site. In addition, 5 conserved sites contain the consensus sequence for stage selector protein (SSP) which was first thought to be involved in fetal to adult gene switching (32). It is evident from the number of phylogenetic footprints and the dense transcription factor binding sites up to 2kb 5' of the cap, that ε is very tightly regulated by gene proximal sequences and is not only regulated by its proximity to the LCR.

1.5.1.2 Sequences which affect ε expression.

The roles of the previously determined sequences as well as other sequences further upstream of ε were assayed by transfecting K562 and HeLa cells with constructs containing different truncated 5' ε sequences, the ε promoter and a luciferase reporter gene. These constructs revealed that negative regulatory sequences reside between HS1 and GATA-1 sites, 6kb and 4kb upstream of ε respectively. There are two more negative regulatory elements located approximately 3kb upstream of ε that contain five regions which are extremely well conserved among primate species and which were

found to bind protein by DNaseI footprinting. Of these five sites, three contain GATA motifs and one contains a progesterone receptor consensus sequence. The subsequent mutation of these GATA motifs using the same ε promoter and luciferase reporter gene, resulted in significant increases in transcription indicating that GATA factors can act as repressors of ε expression. Finally a third site 1.7kb upstream of ε, has also been found to negatively impact transcription in this system. In this region, the DNaseI footprints correspond to GATA-1, Oct-1, and YY1 binding sites identified by the phylogenetic footprinting experiments descibed previously. Sequences with positive effects on transcription have also been identified at -2.9kb and -1.6kb and have not been well characterized (44).

It also appears that two direct repeat sequences upstream of ε may have an important role in ε regulation. These direct repeat sequences are located from –70 to –105 from ε and have been shown to bind the COUP-TFII nuclear receptor during the primitive stage of erythropoiesis. During definitive erythropoiesis, COUP-TFII is no longer expressed and a factor named DRED then binds to the repeats. The binding of DRED is thought to prevent the binding of EKLF and other Krüppel like factors to the adjacent CACCC sequence which in turn down regulates ε expression at the definitive stage of erythropoiesis (75). Furthermore, there is a silencer located approximately 275bp upstream of the cap site. This region binds both GATA-1 and YY1. A deletion performed in the 150kb YAC removed the silencer and 100bp of 3' sequence, which includes another GATA-1 binding site at -177. This construct was used to generate three single copy YAC transgenic mouse lines, all of which had profound deficiencies (less than 10% of wild type expression per transgene copy) in the expression of ε and γ

in the embryonic yolk sac. In addition, both γ and β expression in the fetal liver were moderately reduced to approximately 50% of wild type levels per transgene copy (45). This demonstrates that the ε silencer has an important effect on the regulation of the other β-like globin genes. More importantly, the deletion of a silencer led to an unexpected decline in ε expression. The authors suggest the removal of positive regulatory elements 3' of the silencer, perhaps the GATA-1 site 177bp upstream of the cap which was deleted in addition to the silencer, could have caused the decline in ε expression. However, the result that transcription factor binding to the ε promoter is required for fetal gene expression in the yolk sac would suggest that the ε promoter could either interact with the fetal genes or contribute to the generalized opening of chromatin structure which would extend to the fetal genes.

1.5.2 The fetal genes.

1.5.2.1 Factors which bind upstream of $^{G}\gamma$ and $^{A}\gamma$.

The fetal genes of the human locus are thought to have evolved through gene duplication. Therefore, the genes are quite similar; in fact the $^{G}\gamma$ and $^{A}\gamma$ chains differ by only one amino acid. In addition, the promoter regions and intronic sequences of these genes are very well conserved. These genes have also been subjected to phylogenetic analysis like the ε gene which has revealed 13 very well conserved regions extending up to 1250bp 5' of the cap site. In fact the phylogentic footprints occur in two distinct regions, namely the first 200bp and from 800bp to 1250bp upstream of the cap. These conserved regions were then tested for their ability to bind transcription factors. In the first region, SSP binds two sites immediately adjacent to Sp-1 binding sites. In

addition, there is another Sp-1 binding site in this region located next to one of the three CP1 sites found in this region. There is also a GATA-1 site and Oct-1 site which are again situated side by side. In the region of conservation located further upstream there are two neighbouring GATA-1 and Oct-1 sites. As was seen in the region 5' of ε, many of the regions which bind GATA-1 do not contain the consensus motif. There are also sites which bind YY1, CP1 and unidentified proteins in this region. Subsequent phylogenetic analysis of the region between ε and $^{G}\gamma$ also found conserved GATA-1 and Oct-1 binding sites 2.7kb upstream of the $^{G}\gamma$ cap site (31). The large amount of transcriptional factors which bind the region up to 1.3kb upstream of the fetal genes would suggest that the control of these genes is quite tightly regulated. In fact, transcription factor binding in the embryonic and fetal genes is both dense and extends quite far upstream of the genes.

1.5.2.2 Upstream sequences which affect γ expression.

In addition, truncations of the region upstream of the fetal genes were performed to locate functional elements within this region. These experiments involved the generation of transgenic animals using constructs containing the μLCR and the $^{A}\gamma$ gene truncated at either 141, 201, 382 or 730 bp upstream of the cap site. These experiments indicated that the region contained in the –141 truncation participates in silencing the fetal genes during adulthood. The –201 and –382 truncated constructs were found to have increased fetal gene expression as well as a lack of fetal gene silencing during the adult stage. This would indicate that the region between –141 and –201 contains positive regulatory elements potent enough to overcome the

silencing properties of the region downstream. Interestingly, the Sp1, GATA-1 and Oct-1 factors are known to bind in this region and may be responsible for this effect. The inactivity of the -382 to –201 region corresponds with a lack of protein binding in this region. Surprisingly, the –730 truncated $^{A}\gamma$ construct displayed normal regulation of the transgene. This would indicate that elements required for adult silencing of the fetal genes reside in this area (71). Unfortunately, the phylogenetic analysis did not indicate that proteins bound this region, indicating that non-evolutionarily conserved sequences contribute to the function of the fetal genes.

1.5.2.3 Downstream sequences which affect γ expression.

It was also strongly suspected that the fetal genes were controlled by sequences 3' of $^{A}\gamma$. Several population studies and analyses using plasmid constructs had situated an enhancer immediately 3' to $^{A}\gamma$. This lead to the identification of GATA-1, Sp-1 (40), SATB1 and HOXB2 binding sites contained in this AT rich region. It was subsequently discovered that an activating protein complex termed γPE bound to SATB1 and HOXB2 and could contribute to the observed enhancer function of this region (12). However, when this region was removed in a 750bp deletion of the 150kb YAC, transgenic mice generated with this construct showed normal regulation of all the human β-like genes. Therefore, if this region has any enhancer function *in vivo*, it can be compensated for by other regions of the locus (46).

1.5.3 The adult genes.

1.5.3.1 The δ and β promoter regions.

Like the fetal genes, the two adult β-like globin genes, δ and β, are thought to have arisen through gene duplication. However, this duplication event probably occurred earlier than that of the fetal genes since the δ and β sequences differ more substantially than the fetal gene sequences. In fact, the promoter regions only display a great deal of homology up to 70bp 5' to the cap site. This region is known to bind CP1 and GATA-1 in the case of the β gene but these sites have degenerated upstream of δ. In addition, sites for EKLF, GATA-1, Sp-1, CDP and NF1 located in the first 200bp upstream of the β cap site are required for the high level expression of β and have been abolished in the δ promoter reviewed in (17). The lack of these sites in the δ promoter may be responsible for its much weaker expression as compared to β. Whereas the embryonic and fetal genes have been shown to abundantly bind factors 1-2kb upstream of their cap sites, the β gene binds very few factors past -200bp. In fact, only BP-1 and BP-2, homeobox repressor proteins bind this region. In accordance, experiments involving the truncation of the β promoter reveal that the distal region does have repressor function (17).

1.5.3.2 Sequences upstream of δ.

The distal regions upstream of δ are much different than those of β. Between –275 and –350 lie a GATA-1 site and a BP-1 binding site. This region has been shown to possess negative regulatory activity on the δ-globin gene (17). However, there also

exist important regulatory sequences further upstream of δ. There are GATA-1 and PYBF binding sites in a 250bp pyrimidine stretch located roughly 1 to 1.2kb upstream of δ. Based on transfection of MEL cells with plasmid constructs deleted for this region, this pyrimidine stretch augments the expression of the β-globin gene (1). It was subsequently found that a SWI/SNF type complex named PYR is recruited to this region. A cosmid construct containing the miniLCR and the 3' portion of the β-globin locus from $^{A}\gamma$ through β, was deleted in the PYR binding region and used to generate transgenic mice. These mice displayed a 3 day delay in switching from fetal to adult genes (56) indicating that sequences 5' of δ may not directly impact on δ expression but are of importance to the regulation of the locus as a whole.

1.5.3.3 The β IVSII and 3' β enhancers.

The β-globin gene also contains two enhancers, one located in the 3' portion of the second intron and another located 3' of the gene. The 3' region of the β-globin IVSII has been found to contain sites bound by SATB1, GATA-1 and Oct-1 in K562 cells (35). Oct-1 binding at some sites may compete with other homeobox proteins such as HoxB6 and Antennapedia (35). This agglomeration of transcription factor binding suggests an important role for this enhancer in the expression of not only the β-globin gene itself but the other β-like globin genes as well. Interestingly, all the transcription factors which bind this region, except for GATA-1, are homeobox proteins, which may be involved in long distance enhancement reviewed in (8). The second β-globin enhancer, located 3' of the gene, has been deleted in the 150kb YAC

and used to generate two lines of transgenic mice each carrying four transgene copies. These mice display the correct developmental expression pattern for the human β-like genes. Although the expression level of the embryonic and fetal genes was normal, the adult β-globin gene was expressed at less than 10% of normal levels per transgene copy (45). This suggests that the 3' β enhancer acts specifically on the β-globin gene.

Section B: Regulatory mechanisms of the β-globin locus

1.6 Models of LCR interaction with the β-like globin gene proximal sequences

Several models of LCR interaction with the β-like globin genes have been proposed in the past decade. Although none of these models has been proven, each model can explain some but not all aspects of β-like globin gene regulation. It should be noted that these models are not mutually exclusive and therefore it is possible that the LCR interacts with gene proximal sequences *via* multiple mechanisms.

1.6.1 The LCR holocomplex.

It has been proposed that the different HSs of the LCR associate to form a holocomplex which then interacts with the β-like globin genes. Support for this model comes from the layout of the LCR: The DNaseI HSs are very evenly spaced by approximately 3kb intervals. It is also quite possible that the assembly of this holocomplex and/or the recruitment of factors to the HSs and flanking regions is promoted by the core sequences (11). In addition, it may explain why the larger YAC deletions were less detrimental to globin gene expression than core deletions. It would be possible that a large deletion would allow for the correct assembly of the other HS in

a holocomplex, while core deletions would allow for binding of transcription factors to flanking sequences and interfere with the assembly of this holocomplex.

1.6.2 Gene competition.

The regulation of the β-like globin genes is often explained by gene competition for the enhancer activity of the LCR. This model is also compatible with the distance model in which genes located closest to the LCR are transcribed earliest in development since proximity to the LCR may in itself competitively favour interaction. In the 150kb YAC, the order of the globin genes was inverted with respect to the LCR. Single copy transgenic mouse lines had no expression of ε in the embryonic yolk sac but instead β, now the gene closest to the LCR, was expressed in the yolk sac and through out development (74). So it is possible that the ε gene is activated first in the embryonic yolk sac due to its proximity to the LCR. Subsequently, the ε gene would be permanently silenced by factors in trans and the fetal genes, now the closest to the LCR, would be activated. In turn, fetal gene expression wanes but is never completely silenced. A small amount of HbF is produced normally in adulthood and there is a wide variety of point mutations and large deletions in the β-globin locus which allow for higher level HbF expression in adults. Therefore, a competitive model was also proposed to explain the transcriptional switch from fetal to adult genes. The switch from fetal to adult hemoglobin would be caused by a change in the LCRs preference for interaction with the fetal genes to the adult genes due to the binding of developmentally regulated transcription factors. For instance, transgenic mice containing HS2 and HS3 driven $^{A}\gamma$ and β globin genes with mutated CACCC or TATA motifs (-140 and –30

respectively from the $^{A}\gamma$ cap site) were found not only to have lowered expression of the fetal gene but increased expression of the β gene. If the fetal and adult genes were not competitively regulated the diminished expression of γ, due to promoter mutations, should not have impacted adult gene expression (67). The -140 region has been previously described as a positive regulatory region for the fetal genes. Furthermore, the -140 CACCC sequence is likely to bind one of the Krüppel like transcription factors expressed during fetal stages. However, the β-globin locus was not intact in this experiment and the HS2 and HS3 sequences may not normally interact this way with the fetal and adult promoter sequences.

In parallel with this result, it has also been found that EKLF null 150kb β-globin YAC transgenic mice do not undergo fetal gene silencing when they enter the adult stage. At embryonic day 15, the human fetal genes in these mice are 5 fold over expressed as compared to YAC transgenic mice, while the β-globin gene which normally accounts for the vast majority human β-like globin at this stage, is not expressed at all. This would imply that, in adulthood, EKLF disrupts LCR interaction with the fetal genes and promotes LCR interaction with the β-globin gene proximal sequences. This result also indicates that EKLF may be an important mediator of fetal and adult β-like globin gene competition (57). It has been argued that the autonomous silencing of the ε and γ genes during development is not compatible with the competitive model of β-like globin gene regulation (8). However, it can easily be countered that the silencing process simply vastly decreases the ability of the promoter to compete for LCR interaction and is an intrinsic part of globin switching through gene competition.

1.6.3 The looping model.

The next two models attempt to explain the physical mechanism through which the LCR interacts with β-like globin gene proximal sequences. The looping model of LCR-promoter interaction involves the looping of the LCR and its bound transcription factors to directly associate with gene proximal elements reviewed in (8). This model proposes that the SWI/SNF complex remodels the LCR and/or β-like gene promoter sequences such that the bound transcription factors are correctly positioned with respect to each other for optimal interaction and enhancement of transcription. Interestingly, a SWI/SNF related complex is known to be recruited by EKLF sites upstream of the β-globin gene (42) and this complex could theoretically assist in the looping process.

In addition, it has been shown that the β-globin promoter impacts on the recruitment of EKLF to the HS3 of the LCR (41). The looping model could easily explain this by direct interaction between the LCR and the promoter. However, it is difficult to reconcile the looping model with certain aspects of β-like globin gene activation: In yeast, enhancers which act by a looping mechanism are not capable of exerting this activity over long distances such as those present in the β-globin locus (approximately 70kb) reviewed in (8).

1.6.4 Linking model.

The linking model of LCR mediated enhancement of globin gene transcription would have protein complexes form from the LCR, along the β-globin locus, to the different β-like globin genes reviewed in (8). This protein link would allow for the

developmentally regulated transcriptional activation of the different genes. Support for this model mostly derives from the shortcomings of the looping model, in particular, the inability of looping processes to allow enhancer activity to be transmitted over large distances. It is clear that portions of the HS cores cannot increase transcription from a distance. For example, Langdon *et al.* (38), found that the tandem NF-E2 sites of HS2 placed upstream of a γ gene promoter and luciferase reporter gene allowed for 20 fold enhancement of transcription in K562 cells. However, when this sequence was placed 150bp downstream of the reporter gene, the enhancer activity of the tandem NF-E2 sites was not transmitted to the reporter gene. This position difference was not observed when the entire HS2 region was used as an enhancer so it was suggested that the regions which flank the HS2 core may be important for transmitting the enhancer effect over a distance. In particular, in the case of HS2, the 3' flanking sequence which binds the HOXB2 homeodomain protein may be important in transmitting enhancer activity to the downstream globin genes. Many homeodomain proteins have been found to bind a protein called Chip which is involved in long distance promoter-enhancer interaction by linking and in stabilizing enhancer bound transcription factors in *Drosophila* (76). Theoretically it is possible that HOXB2 is responsible for the long distance enhancer functions associated with the HS2 and the LCR by allowing the formation of a protein link from the LCR holocomplex to the β-globin gene proximal sequences. Though very little direct experimental evidence supports this model, it provides an alternative to the looping mechanism and could explain LCR enhancer activity over the distances observed in the β-globin locus. Most importantly, it should be noted that these models are not mutually exclusive and that it is likely that a

combination of distance effects, gene competition, linking and looping occur to allow for the sequential enhancement of the β-like globin genes.

1.7 Intergenic transcription of the β-globin locus.

1.7.1 Transcription of the LCR:

The regions between the β-like globin genes are transcribed in erythroid cells. Transcription from the HS2 towards ε was described by Tuan and was linked to the enhancer function of this region (36, 79). Other transcribed regions include the region 5' of the LCR (61), the LCR, and the regions surrounding all of the β-like globin genes (2). Transcription upstream of the LCR commences at ERV-9, an endogenous retrovirus, long terminal repeat. Intergenic transcriptional initiation of δ begins at a precise point within a 2.5kb fragment located 2kb upstream of the cap site (30). However, transcription of the LCR and surrounding ε and γ does not have a precise initiation site (61). The transcription of the LCR occurs throughout development (61) and the disruption of the transcription of the LCR is thought to be severely detrimental to globin gene expression. Particularly severe declines in β-like globin gene production are associated with the introduction of transcribed genes into the LCR. For instance, the HS deletions in mice were accomplished by targetting a *PGK-neo* cassette by homologous recombination into ES cells. The *PGK-neo* was then excised from some ES cell lines using FLP recombinase. Interestingly, mice generated from ES cells containing *PGK-neo* inserts in the LCR had much more severely affected globin gene expression than mice generated from ES cells which had excised the *PGK-neo* gene (6, 20, 34). It has also been observed that the insertion of the Friend virus near the human

HS2 causes a dramatic decline in globin gene expression. Thus it is possible that appropriate transcription of the LCR is critical for high level expression of the β-like globin genes.

1.7.2 Intergenic transcription of the β-like genes.

Experiments using nuclear run on assays have found that the ε-γ and δ-β transcripts are produced through out development (2), while experiments using RNA-FISH have found that the ε-γ transcripts are produced exclusively during fetal stages while the δ-β transcripts are produced during adult stages. The latter results are supported by deletion experiments in the 248kb YAC involving the intergenic δ-β start site which abolishes the fetal to adult globin switch of the human β-like genes in transgenic mice (30). Since intergenic transcripts are associated with accessible chromatin structure, these results would suggest that the chromatin structure of the β-like globin genes is developmentally regulated such that only the ε and γ genes are available for interaction with the LCR during embryonic and fetal stages while the δ and β genes become available for interaction with the LCR during the adult stage. This is indicative of an important role for intergenic transcripts in globin gene switching.

1.7.3 Models of LCR enhancement and intergenic transcription.

It now becomes necessary to reconcile the intergenic transcription results with the previously discussed models for β-like globin regulation. Theoretically, if the transcripts are responsible for β-like globin gene switching by increasing chromatin

accessibility of the appropriate genes, then the competitive model of promoter interaction with the LCR could only work for the ε and γ genes since they are the only two sets of genes to be available for transcription at the same time. The looping and linking models would simply allow for interaction of the LCR with the genes in most accessible chromatin regions. However, it is possible that the intergenic transcripts are simply a consequence of the chromatin structure of the locus and that they do not directly influence β-like globin gene expression.

Section C: Anomalies in hemoglobin expression

1.8 Hereditary persistence of fetal hemoglobin.

The transcriptional switch is not always completed in humans. There are a relatively large proportion of people (approximately 4% in the USA) which express higher than usual levels of fetal hemoglobin (43). This benign increase from 1% HbF to 5 to 20% HbF has been termed hereditary persistence of fetal hemoglobin (HPFH). There are two types of genetic lesion which cause HPFH, namely deletions encompassing δ, β, and 3' HS1 and point mutations in the fetal gene proximal sequences. These naturally occurring point mutations coincide with important regulatory sequences upstream of $^{G}\gamma$ and $^{A}\gamma$, and for the most part they either abolish binding of negative regulatory factors or increase the binding of transcriptional activators to allow for increased expression of the fetal genes. For example, -117 $^{A}\gamma$ G to A reduces GATA-1 binding, -202 C to G increases Sp-1 binding, -198 T to C creates a Sp-1 binding site etc. reviewed in (28, 29, 38). In fact, the -117 $^{A}\gamma$ mutation has been introduced into the 248kb β-globin YAC. Transgenic mice containing this

YAC show normal expression of the human β-like genes except that the expression of the fetal genes remains elevated through adulthood (60). Studies have also been performed in cell culture with a cosmid containing $^{G}\gamma$ through β. This cosmid happened to contain the -117 $^{A}\gamma$ mutation and not surprisingly, adult cell lines were unable to silence the fetal genes (53). This supports the idea that the transcriptional switch is regulated by factors in trans which are similar in mice and humans.

1.9 β-thalassemia and Sickle Cell Disease.

HPFH is a modifier of other hemoglobinopathies such as β-thalassemia and sickle cell disease. In the case of β-thalassemia, larger amounts of the α-globin chains are produced compared to the β-like globin chains. This is often due to heterozygosity for a β-globin allele which expresses poorly (β^{+}) or in more severe cases, for a β-globin allele which is not expressed at all (β^{0}). The reduction in β-globin chains leads to a reduction in the number of $\alpha_2\beta_2$ tetramers and to an increase in α_4 tetramers which have a lesser affinity for iron. Thus these individuals are anemic due to the lack of β-globin chains and have problems with iron deposition in the tissues. The production of γ-globin chains during adulthood can remedy this disease by decreasing the amount of α-globin which tetramerises as well as increasing the amount of functional hemoglobin. In sickle cell disease, both alleles of the β-globin gene have a point mutation in the 6th codon, resulting in the wild type glutamic acid residue being replaced by a valine. This mutation causes the Hb to polymerise leading to the distortion and lysis of the RBCs. Interestingly, there are several β-globin locus haplotypes of sickle cell chromosomes,

each of which contains its own set of polymorphisms. Some haplotypes (19 and 20 or Benin and Bantu) cause extremely severe sickle cell disease and are associated with very little production of HbF, while haplotypes 3 and 31 (Senegalese and Saudi Arabian/Indian) cause only moderate disease due to high HbF production reviewed in (37). The increased production of HbF can improve SCD severity by raising the amount of functional hemoglobin. It is also thought that these polymorphisms can improve SCD severity by increasing fetal gene competition for the LCR which in turn decreases the expression of the mutant β-globin gene and diminishes the amount of hemoglobin polymerisation in RBCs.

1.10 An exceptional case of HPFH.

A Sicilian study on thalassemia found an individual who had normal levels of hemoglobin (12.1g/dL), of which 98.6% was HbF. Aside from being homozygous for a number of polymorphisms in the β-globin locus, this individual was also found to be homozygous for a β^0-thalassemic allele containing a G to A mutation in the first nucleotide of the second intron (splice donor site) of the β-globin gene. This individual was in good health except for a large splenomegaly. The parents and siblings of this individual were heterozygous for the thalassemic allele and also had high levels of fetal hemoglobin (64).

1.10.1 HS2 polymorphism.

As mentioned previously, this Sicilian individual with the exceptional case of HPFH was homozygous for a number of polymorphisms. The first of these is in a

microsatellite found 3' of the 5' HS2 of the LCR. The wild type sequence for this region is $(AT)_9N_{12}(AT)_{11}$, while the sequence found in the HPFH individual is $(AT)_9N_{12}(AT)_{10}$. A number of studies have been done on human populations with HPFH and many show that this microsatellite region can impact on fetal gene expression. For example, one study has established a link between low HbF levels in β-thalassemic individuals and the 5' HS2 sequence $(AT)_9$ or $_{10}N_{12}(AT)_{12}$ (39, 65). Yet another study has found that the HS2 sequence $(AT)_9N_{12}(AT)_{10}$ is associated with higher numbers of cells producing HbF in heterozygotes for the sickle cell trait (51).

This region has also been implicated in HOXB2 binding. The *hoxb* cluster of homeobox proteins notably *hoxb2*, 3, 4, and 5 are well expressed in the erythroid lineage. *Hoxb2* in particular, is not expressed in lymphoid cell lines but is well expressed in all erythroid cell lines tested and in some granulocytic cell lines (48, 50). However, *hoxb2* seems to be expressed in the later stages of differentiation in the erythroid lineage and more weakly expressed during early granulopoiesis (27). Although the knock out of *hoxb2* in the mouse has been completed, a defect in erythropoiesis was not reported (4, 15). This may be due to the fact that the groups performing these studies were interested in *hoxb2* involvement in hindbrain and skeletal development rather than hematopoiesis.

HOXB2 was first proposed to bind to the putative $3'^{A}\gamma$ enhancer as a result of a screen which tested the ability of a K562 cDNA library to specifically bind this region. One clone from this library, containing a HOXB2 fragment, was found to bind the $3'^{A}\gamma$ probe, but more interestingly, sequences derived from the 3' flanking sequence of the 5'

HS2 competed for interaction with this factor. The same group showed that HOXB2 could footprint regions 3' of $^{A}\gamma$ and 3' of HS2 (40, 69).

It was also found that a GATA-1 site was essential for *hoxb2* expression in erythroid cell lines, since mutating the GATA-1 binding site or the addition of GATA-1 antisera abolished *hoxb2* expression (82). Further investigations into the region 3' of $^{A}\gamma$ found that another homeodomain protein, SATB1, competes with HOXB2 and that both proteins can recruit a activating protein complex termed γPE (12, 47). Theoretically, though it has not been tested, SATB1 could compete for HOXB2 binding sites in 5'HS2 and either of these homeodomain proteins could recruit the γPE complex to the LCR.

1.10.2 The $^{G}\gamma$ polymorphism.

The second polymorphism in this HPFH individual was found in a microsatellite in the second intron (IVSII) of $^{G}\gamma$. This microsatellite has been found to have several different arrangements in the different sickle cell haplotypes. For example, the severe Benin-19 haplotype has $TC(TG)_9AG(TG)_2(CG)_2$ while the less severe Senegalese-3 and Saudi Arabian-Indian-31 haplotypes have $(TG)_{11}(CG)_3$ at this region (37). It so happens that the HPFH individual has the microsatellite sequence which is associated with the milder form of sickle cell disease and higher HbF levels. Unfortunately, only one individual was used to determine the sequence at this IVSII $^{G}\gamma$ microsatellite for the Bantu haplotype 20. The sequence was found to have the $(TG)_{11}(CG)_3$ sequence but this individual also atypically expressed higher levels of fetal

hemoglobin (17% of total Hb) (37). This is an interesting coincidence, and suggests a role for this sequence in regulating the expression of the fetal genes.

1.10.3 The $^{A}\gamma$ polymorphism.

The third polymorphism occurs in the IVSII of the $^{A}\gamma$ gene which contains a similar microsatellite at roughly the same location as the $^{G}\gamma$ IVSII. In the study on sickle cell haplotypes and HbF expression mentioned previously, there was also a correlation between the $^{A}\gamma$IVSII microsatellite sequence $(TG)_{13}$ found in the Senegalese and Saudi-Arabian-Indian haplotypes and increased HbF production. The Bantu haplotype microsatellite $(TG)_{10}(CG)_{3}(TG)_{9}$ and the Benin haplotype microsatellite $(TG)_{10}(CG)_{4}(TG)_{7}$ were associated with low levels of HbF production. The HPFH individual has the former microsatellite associated with higher HbF expression. One other experiment to ascertain the role of $^{A}\gamma$IVSII in the expression of the fetal genes has been performed in cell culture. It involved transfecting series of chimeric globin gene constructs containing different segments of $^{A}\gamma$ and β-globin genes. When these chimeric constructs were expressed using fetal stage K562 cells, the control $^{A}\gamma$ gene construct was well expressed (10-40% of endogenous $^{A}\gamma$ per transgene copy) while the control β-globin gene construct was barely expressed. Most of the chimeric constructs behaved similarly to the wild type constructs with two notable exceptions: The construct containing the β-globin gene with its IVSII replaced with that of $^{A}\gamma$ showed increased transcription which was initiated at several non-specific sites. In contrast, the construct containing the $^{A}\gamma$ gene containing the βIVSII was expressed in only one of

six stably transfected cell lines (16). This would indicate that the IVSII of $^{A}\gamma$ is required for the expression of its gene. In addition, the expression of non-specifically initiated β-globin-$^{A}\gamma$IVSII transcripts suggests that the IVSII of $^{A}\gamma$ can render this construct more accessible to transcriptional machinery but that the β-globin promoter is not properly recognized in a fetal environment. Therefore it cannot be assumed that the $^{A}\gamma$IVSII acts in the same manner as the βIVSII enhancer. A study on thalassemic alleles also found that the $^{A}\gamma$ $(TG)_{13}$ microsatellite was associated with milder disease but more importantly, they found that this specific sequence bound specifically to an unidentified protein. In support of a role for IVSII in the expression of the β-like globin genes, Bank and colleagues, have previously shown that the IVSII of the β-globin gene binds many transcriptional factors and regulates mRNA stability of the β-globin gene (35). It would be possible for the IVSII of $^{G}\gamma$ and $^{A}\gamma$ to have a conserved evolutionary function with the β-globin IVSII even if there is no evident homology between these sequences. Nonetheless, it is probable that the IVSII of $^{G}\gamma$ and $^{A}\gamma$ impact the level of fetal hemoglobin expression.

1.10.4 The -530 β polymorphism.

The last polymorphism found in the HPFH individual was found approximately 0.5kb upstream of the β-globin gene. This site is another microsatellite of the type $(AT)_xT_y$ which has also been associated with HbF levels in sickle cell haplotypes. The Indian haplotype and the HPFH individuals sequence is $(AT)_9T_5$ and binds to a distaless family of homeobox protein called BP1 more strongly than the Bantu and Benin

haplotype sequences $(AT)_8T_4$ and $(AT)_6T_9$ (18). BP1 has a reputed repressor function on the β-globin gene and thus may skew competitive LCR interactions with the γ and β-globin promoters in the favor of the fetal genes.

2. Materials and Methods.

2.1 YIP constructs.

2.1.1 PCR amplification.

Polymerase chain reactions were carried out in a Perkin Elmer (Norwalk, CT) thermocycler using a modified version of the technique mentioned in (3). Oligonucleotide pairs (Life Technologies, Burlington, ON) used for PCR amplification are listed in table 2.1. Typically, each reaction was performed in a volume of 20μL, containing the appropriate buffer, described in table 2.2, in addition to 1mM of each oligonucleotide, 125μM of each dNTP (Amersham Pharmacia Biotech, Baie D'Urfé, QC), 4 units of Taq polymerase, as well as plasmid or genomic DNA.

Table 2.1: PCR primers.

Name	Sequence (5' to 3')	Description (Genbank #)
99-17	GCTGACCACCTGACTAAAACT	5' HS2 (8784-8804)
99-18	TATAGAGGCCACCTGCAAGAT	3' HS2 (9007-8987)
00-2	AGGTAGTTGTTCTTCTTGCA	3' $^{G}\gamma$IVSII (35701-682)
00-3	AGGGCTGGGCATAAAAGTCAGGGC	5' β exon 2 (62612-62590)
00-4	AGGGCTGGGCATAAAAGTCAGGGC	3' β promoter (62096-62119)
00-5	GCTGGACACATATAAAATGC	5' $^{G}\gamma$ IVSII (2) (35405-35424)
01-09	TGAGAAAGGAGAGAGGAGAAAG	5' HS2 (7699-7720)
01-10	TTCCCCACAAGAGTCCAA	5' HS2 (8331-8313)
01-11	CGCGTAATACGACTCACTATAGGG	T7

Table 2.2: Buffer conditions for PCR amplification.

Forward primer	Reverse primer	Buffer used (final concentrations)
99-17	99-18	1.0mM $MgCl_2$, 50mM KCl, 10mM Tris pH 8.3
00-2	00-5	1.0mM $MgCl_2$, 50mM KCl, 10mM Tris pH 8.3
00-3	00-4	0.75mM $MgCl_2$, 50mM KCl, 10mM Tris pH 8.3
01-09	01-10	1.5mM $MgCl_2$, 50mM KCl, 10mM Tris pH 8.3
T7	01-10	1.5mM $MgCl_2$, 50mM KCl, 10mM Tris pH 8.8

All reactions were held at 94°C for 3 minutes prior to the amplification cycles. Cycling conditions were dependent on the oligonucleotide pair and are described in

table 2.3. This was followed by a final incubation at 72°C for 10 minutes and the reactions were then stored at 4°C. Generally amplification reactions performed for cloning were run for 25 cycles to reduce the introduction of mutations by Taq polymerase. PCR amplification reactions used to generate fragments for cloning were cycled 25 times to reduce Taq mediated mutagenesis. PCR reactions used to generate sequencing templates were cycled 30 times while amplifications directly from whole yeast were used for preliminary screening and underwent 35 cycles.

Table 2.3 Cycling conditions for PCR amplification.

Forward primer	Reverse primer	Denaturation	Annealing	Elongation
99-17	99-18	94°C, 60s	57°C, 45s	72°C, 30s
00-2	00-5	94°C, 60s	61°C, 45s	72°C, 30s
00-3	00-4	94°C, 60s	64°C, 60s	72°C, 60s
01-09	01-10	94°C, 60s	55°C, 45s	72°C, 40s
T7	01-10	94C, 60s	55C, 30s	72C, 75s

2.1.2 Enzymatic digestions and modifications.

Restriction endonucleases and modification enzymes were obtained from New England Biolabs (Missisauga, ON), Amersham Pharmacia Biotech (Baie D'Urfé, QC), Life Technologies (Burlington, ON) and Promega (Madison, WI). These enzymes were used in accordance with the conditions recommended by the manufacturer. In certain cases, one or more of the following techniques were used to modify digested DNA fragments. Blunting of sticky ended fragments was achieved by an incubation of 20 minutes at 25°C with 1U Klenow fragment of *E. coli* DNA polymerase I per μg DNA and 33μM of each dNTP in the digestion buffer. It should be noted that the Klenow fragment of *E. coli* DNA polymerase I and many other modification enzymes are active in all commonly used digestion buffers. This eliminated the need to adjust

salt concentrations or precipitate the DNA prior to the modification of the digested fragments. Dephosphorylation of DNA fragments was done by incubation of 0.3 units of calf intestinal alkaline phosphatase per μg of DNA fragment. This reaction was completed in 60 minutes at 37°C in the digestion buffer. Phosphorylation of DNA fragments was accomplished by incubating with 10 units of T4 polynucleotide kinase per μg of DNA and 20μM each dNTP, in the digestion buffer, for 20 minutes at 25°C.

2.1.3 Purification of DNA.

All of the techniques in this section are slightly modified forms of techniques described in (66).

2.1.3.4 Ethanol precipitation: Ethanol precipitation of DNA was done by adding 0.1M NaCl to the aqueous DNA solution and then ethanol was added to a final concentration of 70% (v/v). This mixture is gently mixed by inversion, incubated for at least 20 minutes at -20°C and centrifuged at 16000×g for 10 minutes at 4°C. The supernatant was then removed and the DNA pellet was dried by vacuum before being resuspended in ddH_2O, for sequencing templates or TE buffer (10mM Tris pH 8, 1mM EDTA pH 8).

2.1.3.2 Isopropanol precipitation: Isopropanol precipitations were performed mostly on PCR fragments used as sequencing templates since traces of ethanol in the template can yield poor sequencing results. Again, NaCl was added to the DNA solution to a final concentration of 0.1M. Then an equal volume of isopropanol was added (ie 50% v/v) and the mixture was incubated at -20°C or -80°C for at least 20 minutes. Both the

temperature and duration of precipitation were dependent on the size of DNA fragment to be precipitated. The mixture was then centrifuged at 16000×g for 10 minutes at 4°C and the supernatant was removed. The remaining DNA pellet was dried by vacuum before being resuspended in ddH_2O or TE buffer.

2.1.3.3 Agarose gels: DNA fragments were separated on 0.8-1.0% (w/v) agarose gels (Life Technologies, Burlington, ON) depending on the size of the fragment, and run at approximately 100V in 1X TBE buffer (89mM Tris, 89mM boric acid, 2.4mM EDTA) with 0.2mg/L ethidium bromide. Gels were subsequently visualized and photographed using UV illumination.

2.1.3.4 Electroelution from agarose gels: The desired band was excised from the gel and placed in a 12-14000MWCO dialysis membrane (Spectrum Medical Industries, Houston, TX) with 1X TBE buffer (the same as mentioned previously but without ethidium bromide). The gel fragment was then electroeluted at 100V for 60 minutes in 1X TBE buffer. The buffer from the dialysis membrane was then recuperated, verified under UV light for the presence of nucleic acids and precipitated with ethanol or isopropanol.

2.1.3.5 Polyacrylamide gels: Smaller DNA fragments were separated on 8-20% (w/v) polyacrylamide gels (EM Science, Gibbstown, NJ) with a 29:1 acrylamide to bisacrylamide ratio (w:w). Acrylamide percentage varied according to DNA fragment size but all gels were run at approximately 200V in 1X TBE buffer. The gel was

subsequently stained in a 0.2mg/L aqueous ethidium bromide solution and visualized and photographed under UV light.

2.1.3.6 Electroelution from polyacrylamide gels: Electroelutions from polyacrylamide fragments were done as previously mentioned for agarose gels, except the dialysis membrane was left to migrate 2 hours at 100V. The increased duration of the electroelution was required since the polyacrylamide fragments were so large that they folded tightly in the dialysis membrane, impeding the migration of DNA into the buffer.

2.1.3.7 Phenol chloroform extractions: Phenol chloroform extractions were performed on fragments up to 4kb by adding an equal volume of phenol (buffered in 1M Tris pH 8.0) to the aqueous DNA solution, vortexing and centrifuging at 16000×g for 5 minutes at RT. This process was repeated twice on the superior phase, first with 50% (v/v) phenol, 50% (v/v) chloroform and then with chloroform alone. The superior phase was then ethanol precipitated.

2.1.4 Ligation:

2.1.4.1 Blunt ended ligations: Blunt ended ligations were carried out using a 10:1 insert to vector molar ratio. Typically, these reactions were done in a volume of 5μL, containing approximately 200ng of DNA, the 1X ligase buffer (without ATP) supplied

by the manufacturer, 1U T4 DNA ligase (Amersham Pharmacia Biotech, Baie D'Urfé, QC), and 125μM ATP. This mixture was incubated at 4°C for at least 16 hours

2.1.4.2 Sticky ended ligations: Sticky ended ligations were done in the same manner as blunt ended ligations except that the insert to vector ratio was lowered to 5:1, the final concentration of ATP used in the ligation reaction was 1mM as recommended by the manufacturer for this type of ligation reaction, and the incubation was carried out for 4 to 16 hours at 16°C.

2.1.5 Transformation:

2.1.5.1 Media and selection: TYM broth was composed of 2% (w/v) bactotryptone, 0.5% (w/v) yeast extract, 0.1M NaCl and 10mM $MgSO_4$. Luria Bertani (LB) broth was composed of 1% (w/v) bactotryptone, 0.5% (w/v) yeast extract, 1% NaCl (w/v) and LB agar had an additional 15g/L of agar. All selective media contained ampicillin at 100mg/ml.

2.1.5.2 Preparation of competent DH5α *E. coli* cells: A 2mL culture of DH5α cells were grown shaking overnight at 37°C in TYM broth. Approximately 100 μL of this culture was used to inoculate 20mL of TYM broth, and was grown in the same conditions as previous, to an optical density at 600nm (OD600) between 0.2 and 0.8. Then the 20mL culture was expanded to 100mL by adding 80mL of prewarmed TYM. This culture was kept at 37°C, shaking until an OD600 of 0.5 to 0.9 was obtained. An additional 400mL of warmed TYM was added to the culture, which was grown, until an

OD600 of 0.6 was obtained. The cells were subsequently cooled in an ice-water bath and centrifuged at 1600×g for 10 min at 4°C in sterile bottles. The supernatant was decanted and the pellet of cells was gently resuspended in 100mL of cold sterile TFBI (10mM $CaCl_2$, 30mM KOAc, 50mM $MnCl_2$, 100mM KCl, glycerol 15%(v/v)) and centrifuged again at 3000×g for 10 min at 4°C. The buffer was decanted, the pellet was resuspended in 20mL of cold sterile TFBII (10mM NaMOPS pH 7.0, 75mM $CaCl_2$, 10mM KCl, glycerol 15% (v/v)) and then aliquoted into 200μL samples and quick-frozen in an ethanol and solid CO_2 bath. These aliquots were stored at -80°C for future use. All competent cells were streaked on LB-amp plates to detect contamination both prior to and subsequent to this procedure and typically, a competence of more than 10^7 transformants/μg was obtained using pBluescript SK (+) as a control vector.

2.5.1.3 Transformation: A 200μL aliquot of competent cells was incubated on ice for 30 minutes with 100μL of 0.1M Tris pH 8 and 0.5 to 5 μL of ligation mix or 10-100pg of supercoiled DNA. The cells were subsequently heat shocked for 5 minutes at 37°C and kept on ice briefly before adding 500μL of LB broth. The cells were allowed to recover for 1 hour with shaking at 37°C and were then plated on LB ampicillin plates. In the case of ligations which allowed for blue/white screenings to distinguish clones containing religated vector from clones containing vectors with inserts, the cells were plated with 150μL of 20mg/mL X-gal (Vector Biosystems Inc., Toronto, ON). All plates were incubated approximately 16 hours at 37°C.

2.1.6 Mini-preparation of plasmid DNA:

Plasmid DNA was prepared by a modified alkaline lysis method (66). Isolated colonies from transformation plates were used to inoculate 2mL LB ampicillin cultures. These cultures were left to grow at 37°C, shaking for 16 hours and subsequently transferred into 1.5mL tubes. The tubes were centrifuged at 6200×g for 1 minute and the pellet was resuspended in 100µL of lysis solution (25mM Tris pH 8, 10mM EDTA pH 8, 50mM glucose, 2mg/mL hen egg lysosyme (Sigma, Oakville, ON)). The tubes were incubated for 30 minutes at RT and then 200µL of a 0.2M NaOH and 1% SDS solution was added. The tubes were again incubated on ice for 5 minutes prior to neutralisation with 150µL of 3M NaOAc pH 4.8. The tubes were again incubated 30 minutes on ice and centrifuged for 10 minutes at 16000×g. The supernatant was then ethanol precipitated and resuspended in 100µL of TE.

2.1.7 Colony hybridization:

2.1.7.1 Membrane preparation: Isolated colonies from transformation plates were streaked on LB ampicillin plates and grown 16 hours at 37°C. Pre-wetted Hybond-N membranes (Amersham Pharmacia Biotech, Baie D'Urfé, QC) were applied to the streaked plates. In accordance with the protocol in (3), these membranes were placed sequentially on filter paper soaked in 0.5M NaOH, 1M Tris pH 7.5 and a 0.5M Tris pH 7.5 and 1.25M NaCl solution for 5 minutes. The membranes were then washed in 0.5M Tris pH 7.5 and 1.25M NaCl for 2 minutes, dried and crosslinked with 1200µJ of ultra-violet light.

2.1.7.2 Nick translated probe: Nick translated probes were prepared with a protocol similar to (66). A reaction mixture containing 100ng of probe DNA in 20μL of 50mM Tris pH 7.5, 5mM $MgCl_2$, 10mM 2-mercaptoethanol and 100μg/mL BSA with 10μM each d(ATG)TP, 3.3μM ^{32}PαdCTP (30μCi total), 10U *E. coli* DNA polymerase I (Amersham Pharmacia Biotech, Baie D'Urfé, QC) and 100ng/mL DNase I (Roche, Laval, QC) was prepared. This reaction was incubated for 30 to 90 minutes at 16°C and stopped when an incorporation of 30-60% of ^{32}PαdCTP was obtained with 100μL Tris pH 7.5, EDTA pH 8, 5% (w/v) SDS. The probe was then denatured by boiling for 5 minutes and kept on ice.

Table 2.4: Nick translated probes.

Probe	Location (Genbank #)
HS2	2.3kb β-locus KpnI-NcoI fragment (7768-10075)
ε	0.9kb β-locus ClaI-PvuII fragment (18638-19506)
$^{G}\gamma$	2.6kb β-locus StuI-SphI fragment (34094-36686)
β	1.7kb β-locus BamHI-PstI fragment (62613-64301)
LYA	2.7kb pBR322 BamHI-PvuII fragment
RYA	1.7kb pBR322 BamHI-PvuII fragment
URA3	0.8kb pRS406 PstI-NcoI fragment
HS1-4	7.8kb miniLCR composed of:
	1.2kb β-locus SacI-SacI fragment (955-2203)
	0.8kb β-locus SacII-PvuII fragment (4278-5124)
	1.5kb β-locus KpnI-BglII fragment (7768-9218)
	4.2kb β-locus EcoRI-EcoRV fragment (10947-15182)

2.1.7.3 Hybridization: The colony lift membranes were prepared by the method of Sambrook (66), with the following modifications: Membranes were incubated for 4 hours at 65°C in hybridization solution (10X Denhart solution 2% (w/v) ficoll 400, 2% (w/v) polyvinylpyrrolidone, 2% (w/v) BSA all obtained from Sigma, Oakville, ON),

5X SSC (0.75M NaCl, 0.075M sodium citrate, pH 7.0), 0.1mg/mL salmon sperm DNA, and 0.1% (w/v) SDS). Afterwards, the denatured nick translated probe was added to the hybridization solution and incubated with the membranes for 16 hours at 65°C. Membranes were then washed in 0.1% (w/v) SDS, 0.1X SSC for 1-2 hours and the radioactivity remaining on the membrane was visualized by autoradiography film (Eastman-Kodak Company, Rochester, NY)

2.1.8 Maxi-preparation of plasmid DNA on CsCl gradient:

A 20mL culture of LB ampicillin broth was inoculated with an isolated colony and incubated with shaking at 37°C for 8 hours. Half of this culture was transferred into 500mL of LB ampicillin broth and was incubated at 37°C, with shaking for 16 hours. The culture was then centrifuged at 1600×g for 10 minutes at 4°C and resuspended in 10mL of a 25% (w/v) sucrose, 50mM Tris pH 8 solution. Then 2mL of a 10mg/mL hen egg lysosyme in TE solution was added to the bacterial suspension and incubated for 10 minutes on ice in order to degrade the bacterial cell wall. A further 2mL of 0.5M EDTA pH 8.5 was added to the mixture and incubated 10 minutes on ice this destabilizes the bacterial plasma membrane. Finally, 16mL of a 50mM Tris pH8.0, 62.5mM EDTA, 0.1% triton X-100 (Sigma, Oakville, ON) solution was added to lyse the cells and incubated a further 15 minutes on ice. The bacterial lysate was then centrifuged at 80000×g for 1 hour at 4°C.

Purification on CsCl gradient was performed by a method similar to (3). The non-viscous supernatant was removed, leaving behind much of the genomic DNA and recentrifuged at 80000×g for 16 hours at 20°C over 3mL of TE saturated with CsCl to

remove RNA. Then, 4.5g of CsCl and 0.4mL of 4mg/mL ethidium bromide was added to the bottom most 7.6mL of the recentrifuged mixture. This was centrifuged at 250000×g for 24 hours at 15°C after which, the supercoiled plasmid DNA band was removed from the gradient while the linearised plasmid DNA was left behind. The ethidium bromide was then removed from the DNA with several extractions with an equal volume of isopropanol. This was followed by two ethanol precipitations to remove residual CsCl. The final DNA solution was quantitated and verified for purity by spectrophotometry. Using this method a yield of 0.5 to 5mg of DNA is obtained depending on the copy number of the plasmid.

2.1.9 Preparation of DNA for sequencing:

2.1.9.1 Plasmid DNA: Plasmid DNA already purified by CsCl maxipreps were diluted in ddH_20 according to the IRCM (Montreal, QC) sequencing service specifications (25fmol/μL). Plasmid DNA was otherwise prepared in triplicate by the previously described miniprep method but resuspended in a total volume of 90μL. After digestion for 1 hour at 37°C with 10μL 5mg/mL RNaseA (Roche, Laval, QC), these DNAs were purified on spin miniprep columns (Qiagen, Mississauga, ON). Purified DNA was eluted using ddH_20, quantitated and verified on an agarose gel using DNA standards and diluted appropriately for either the IRCM or Université Laval sequencing services.

2.1.9.2 PCR products: Typically, 9 or 10 identical PCR reactions and controls were performed in the conditions mentioned above. These reactions were run on 8% polyacrylamide gels, after which, the PCR product was excised from the gel,

electroeluted and precipitated with isopropanol. Again, PCR products were quantitated and verified on agarose gels and diluted in ddH_20 appropriately for sequencing at the IRCM sequencing service or the Université Laval (Québec, QC) sequencing service. Sequencing was performed with the appropriate primers as listed in table 2.4.

Table 2.5: Sequencing primers.

Name	Sequence (5' to 3')	Description (Genbank #)
99-17	GCTGACCACCTGACTAAAACT	5' HS2 (8784-8804)
99-18	TATAGAGGCCACCTGCAAGAT	3' HS2 (9007-8987)
00-1	CATTTGGGCTGGAGTAG	5' $^{G}\gamma$IVSII (1) (35505-35521)
00-53	CAGATGGGGGCAAAGTATGTCCAG	5' $^{G}\gamma$ IVSII (3) (35462-35485)

2.2 Homologous recombination in yeast:

2.2.1 Media and selection:

2.2.1.1 Rich media: YPD (10g/L yeast extract, 20g/L peptone, 20g/L dextrose) was used to culture yeast strain AB1380 ($trp1^-$ $lys2^-$ $ura3^-$) containing the A201F4 YAC at 30°C from frozen stocks. The strain was subsequently restreaked on selective dropout media.

2.2.1.2 Minimal media: trp^-, lys^-, ura^- dropout media were composed of 20g/L dextrose, 540mg/L amino acid mixture (2g adenine-HCl, 2g L-arginine-HCl, 1g L-histidine-HCl, 3g L-isoleucine, 3g L-leucine, 1g L-methionine, 2.5g L-phenylalanine, 10g L-threonine, 2.5g L-tyrosine), 1.7g/L yeast nitrogen base (without amino acids and ammonium sulphate), and 5g/L ammonium sulphate. This media was supplemented with 20g/L bactoagar for solid plates and 20mg/L uracil for trp^-lys^- dropout media.

2.2.1.3 5-FOA media: 5-FOA plates were of similar composition to minimal media except that 1g/L 5-FOA in addition to 2g/L of a richer amino acid mix were added (2g adenine, 1g uracil, 2g L-tryptophan, 1g L-histidine, 1g L-arginine, 1g L-methionine, 1.5g L-tyrosine, 1.5g L-isoleucine, 7.5g L-valine, 1.5g L-lysine, 2.5g L-phenylalanine, 5g L-glutamic acid, 1.8g L-aspartic acid, 10g L-threonine, 1.8g L-serine, 3g L-leucine).

2.2.2 Yeast transformation: Yeast transformations were performed by the lithium acetate method described in (3). Yeast cells are kept in the mid logarithmic phase of growth and subsequently treated with 0.1M LiOAc and 1 and 10μg (typically 5μg) of linearised plasmid DNA was added. The cells were then heat shocked and plated on appropriate dropout media.

2.2.3 Selection of transformants on 5-FOA: Screened yeast transformants were cultured in 3mL $trp^{-}lys^{-}$ dropout broth and grown for 12-16 hours, with shaking at 30°C. 200μL aliquots of this culture were spread on each of 2 5-FOA plates and incubated for 72hrs at 30°C.

2.2.4 Minipreparation of yeast genomic DNA: Yeast minipreps were done according to (3). Yeast cells were spheroplasted with lyticase (Sigma, Oakville, ON) and subsequently lysed with SDS. After lysis, the proteins were precipitated with KOAc and the DNA recovered by ethanol precipitation followed by isopropanol precipitation.

2.2.5 Yeast plugs: Intact yeast chromosomal DNA was prepared in agarose plugs. A 3mL aliquot of the appropriate media was inoculated with a yeast colony and incubated for 24 hours at 30°Cwith shaking. The cells were spun down at 900×g for 5 minutes and rinsed by resuspending in 5mL of ddH_2O and recentrifugation at 900×g. This was followed by rinses with 5mL of 50mM EDTA and 4mL SCE (1M sorbitol, 100mM trisodium citrate, 50mM EDTA, pH 7.0). The cells were then spheroplasted by incubation in 150μL of a solution composed of 2mL SCE, 50μL 2-mercaptoethanol and 100U lyticase for 90 minutes at 37°C. Then 200μL of pre-warmed 1.5% (w/v) low melting point agarose (GibcoBRL, Gaithersburg, MD) in 0.125M EDTA was added to the spheroplasts and left to solidify in tubing. The plugs were expelled from the tubing and incubated for 1 hour at 37°C in LDS solution (1%(w/v) lithium dodecyl sulphate, 10mM Tris,1M EDTA, pH 8.0). The LDS solution was changed and the plugs were incubated 16 hours at 37°C. The plugs were then rinsed and stored in TE_{50} (10mM Tris, 50mM EDTA, pH 8.0).

2.2.6 Pulsed field gel electrophoresis: Yeast chromosomes were separated on 1% (w/v) agarose gels in 0.5X TBE. The voltage was pulsed for separation of DNA fragments measuring 8 to 500kb on a Q-Life (Kingston, ON) PFGE system. After migration, the gel was coloured in ddH_2O containing 0.5mg/mL ethidium bromide for 30 minutes followed by destaining for 10 minutes in ddH_2O.

2.2.7 Southern blotting:

On yeast genomic DNA prepared on a miniprep scale: Enzymatic digests of 10μL of yeast miniprep DNA were run on a 1% (w/v) agarose gel as previously described. The gel was then placed in a denaturing solution (0.5M NaOH, 1M NaCl) for 30 minutes and subsequently neutralised for 1 hour in 0.5M Tris, 3M NaCl, pH 7.4. The agarose gel was then placed upside down on a support lined with 3M paper (Mandel, ON). A nylon membrane and 2 layers of 3M paper were placed on top of the gel and air bubbles between the layers were eliminated. A 10cm pile of absorbent paper towel was placed on top of the gel and membranes, and weighted down for a 16 hour incubation. The nylon membrane was recuperated and UV crosslinked with 1200μJ of radiation (Stratagene, LaJolla, CA). These membranes were hybridized with nick translated probes as described in the section on bacterial colony lifts.

On PFGE Gels: These gels were transferred onto nylon membranes as previously described with the following modification. Prior to the incubations in denaturing and neutralising solutions, depurination with 0.25M HCl for 15 minutes was performed and was followed by 3 rinses in ddH_2O.

2.2.8 SSCP: The DNA fragment used for SSCP came from the HS2 PCR product obtained with primers 99-17 and 99-18. Cycling for 35 cycles, using conditions previously described, directly on yeast colonies yielded enough material for this analysis. Samples were denatured by incubating 10μL of PCR reaction in 1.5X TBE, 5% deionized formamide and 20mM NaOH in a final volume of 20μL at 95°C for 5 minutes and immediately placing the samples on ice afterwards. These samples were

loaded onto a prechilled 10% polyacrylamide gel (29:1 w:w acrylamide to bis acrylamide ratio) containing 5% glycerol (v/v) and run in 1.5X TBE at 20V/cm for 5 hours at 4°C. The gel was then stained for 5 minutes in 50ng/mL ethidium bromide and photographed.

3. Results:

The major objective of this project is to determine the regions of the β-globin locus which impact fetal to adult hemoglobin switching. Understanding the fetal to adult globin switch may allow modulation of this process and to delay or prevent the down regulation of the fetal genes. This would have a tremendous impact on the clinical treatment of hemoglobinopathies which affect the adult genes, such as sickle cell disease and β-thalassemia.

Naturally occuring mutations and polymorphisms have been associated with elevated HbF levels in adulthood. Since one such HPFH individual (previously described in the introduction) expresses almost 100% HbF in adulthood, the polymorphisms found in the β-globin locus of this individual were chosen for study. In order to assess the roles of the polymorphisms found in the β-globin locus of the HPFH individual in β-like globin gene regulation, the HS2 and the $^{G}\gamma$ IVSII polymorphisms were selected for study in the 150kb β-globin YAC system. These polymorphisms were to be cloned into yeast integrating plasmids (YIps) for introduction into the 150kb YAC by homologous recombination to generate an HS2 HPFH polymorphic YAC, a $^{G}\gamma$ IVSII polymorphic YAC and a compound polymorphic YAC for both the HS2 and $^{G}\gamma$ IVSII HPFH sequences. These YACs will subsequently be used to generate transgenic mice. A comparison of β-like globin gene expression and regulation between these transgenic mice and mice with the wild type YAC transgene could provide evidence for the importance of these sequences in fetal to adult gene switching. Further investigation of the functions of these sequences may elucidate the mechanisms involved in the fetal to adult globin switch.

3.1 Generating the HS2 HPFH polymorphic YAC:

A yeast integration plasmid (YIp) with an HS2 HPFH polymorphic insert was generated to introduce the HS2 HPFH polymorphism into the 150kb β-globin YAC. The wild type HS2 YIp was generated by inserting a 2.3kb KpnI to NcoI fragment of HS2 (genbank # 7768 to 10075) obtained from pII (a gift from Q. Li) into the KpnI and NotI sites of the pRS406 vector (70). This 2.3kb fragment contains the HS2 core and all flanking sequences. More importantly, this fragment contains approximately 1kb of flanking sequence on either side of the HPFH polymorphism which should allow for efficient homologous recombination into the homologous region of the β-globin YAC.

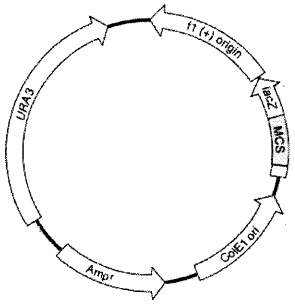


Fig 3.1: Representation of pRS406 (72).

The pRS406 YIp vector contains an origin of replication for *E. coli*, an ampicillin resistance gene, for selection of pRS406 containing *E. coli* clones, a lacZ gene, permitting blue/white selection on X-gal media to easily distinguish clones with inserts, and a *URA3* gene for selection of yeast clones containing pRS406 on media lacking uracil (fig. 3.1). Of note, sequences required for autonomous replication in yeast are absent. Thus for pRS406 to confer prototrophy to uracil in transformed yeast, it must integrate into the genome. Subsequently, the HS2 HPFH polymorphic sequence was amplified from genomic DNA by PCR with primers 99-17 and 99-18 (fig. 3.2), and this PCR fragment was then digested with XbaI and MfeI

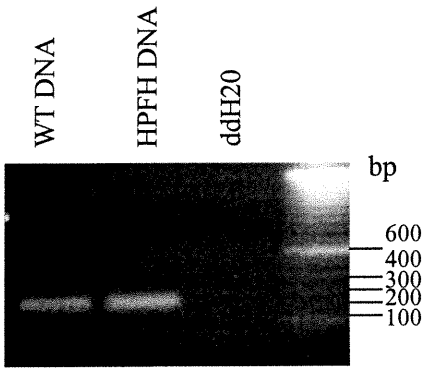


Fig 3.2: PCR amplification of HS2 polymorphic region with primers 99-17 and 99-18. This amplifies a 224bp fragment corresponding to genbank # 8784-9007.

restriction enzymes which cut on either side of the polymorphic sequence. This restriction digest fragment was used to replace the homologous XbaI to MfeI of the HS2 wild type sequence by the HS2 HPFH polymorphic sequence in the YIp. This YIp was verified by restriction digest for integrity of the HS2 insert (fig. 3.3) and sequencing of the polymorphic region to ensure that the sequence was $(AT)_9N_{12}(AT)_{10}$ and not the wild type $(AT)_9N_{12}(AT)_{11}$ and also to ensure that no PCR artefacts were introduced (fig. 3.4). In total, 4 HS2 HPFH polymorphic YIp clones were generated. These YIps were used to introduce the HPFH polymorphisms into the β-globin YAC

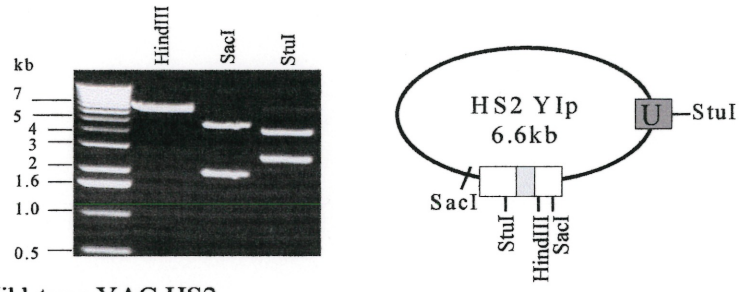


Fig 3.3: Diagnostic restriction digest of HS2 YIp: HindIII linearizes the 6.6kb YIp while SacI digest yields bands at 2.0 and 4.6kb and StuI digest yields bands at 2.6 and 4.0kb.

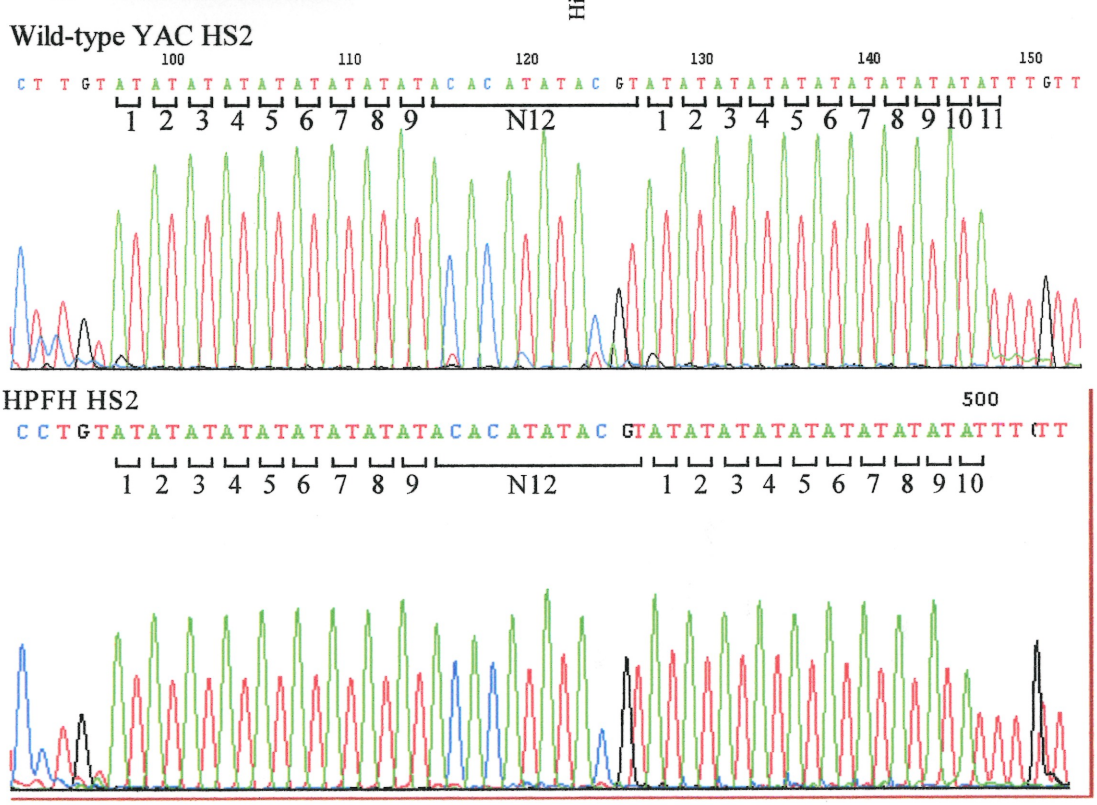


Fig 3.4: Wild type YAC and HPFH polymorphic YIp sequences. The HPFH polymorphic region sequenced with primer 99-18 in the wild type YAC (top) and YIp clone 1.7.16 (bottom).

by the pop-in and pop-out method as illustrated in fig. 3.5. The pop-in involves the integration of the YIp into the YAC by homologous recombination and selecting for markers contained in the YIp. Correct integration events can be favored by introducing dsDNA breaks in the YIp insert. This increases recombination between homologous sequences in the YIp insert and the YAC and avoids recombination events which use YIp vector sequences.

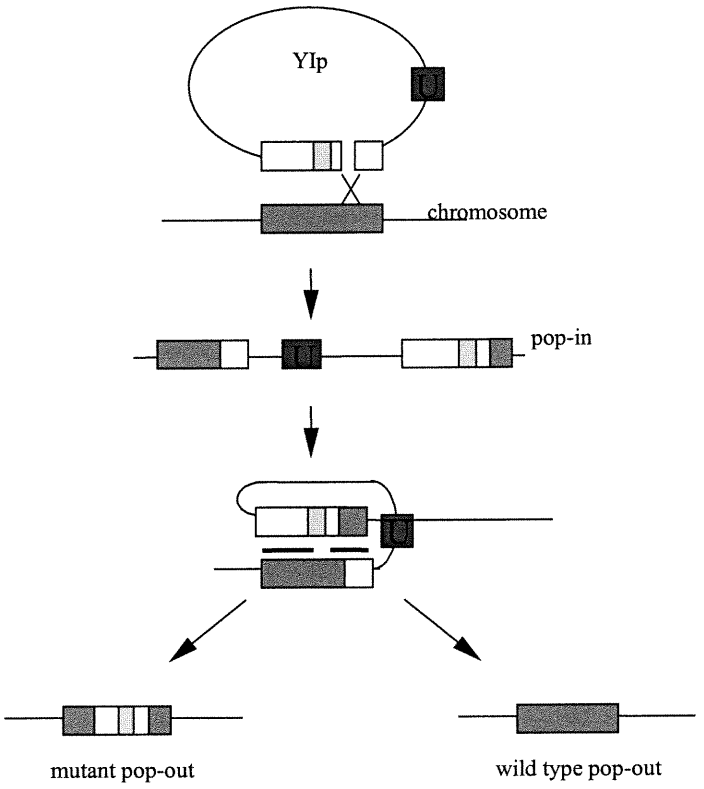


Fig 3.5: Homologous recombination in yeast to introduce targeted mutations. When selected for *URA3*, the digested YIp inserts into a homologous region of the yeast chromosome. Subsequent selection against *URA3* leads to excision of the YIp by homologous recombination. This event can leave the yeast chromosome with the YIp insert sequence or the wild type sequence depending on the location of the cross over indicated by the solid bar.

In this particular case, the yeast strain AB1380 containing the A201F4 YAC, the 150kb β-globin YAC previously described in the introduction, was used for all manipulations. This yeast strain requires tryptophan, lysine and uracil. Therefore the 150kb β-globin YAC can be maintained by selecting for the TRP1 and LYS2 genes located on the left and right YAC arms respectively. (Actually, the β-globin YAC was

originally maintained in a trp$^-$ura$^-$ yeast strain and the right YAC arm contained a *URA3* marker. When the modification of the YAC by the pop-in and pop-out method became feasible, it was transferred into the trp$^-$lys$^-$ura$^-$ AB1380 yeast strain and the right YAC arm containing the *URA3* gene was subsequently disrupted by retrofitting with the LYS2 gene. It should also be noted that the endogenous AB1380 *URA3* gene is disrupted by a transposable Ty element.) This yeast strain was transformed by the LiAc method with the SmaI (genbank # 8224) linearised HS2 HPFH polymorphic YIp. This restriction site occurs approximately at the one third mark of the 2.3kb HS2 insert. The pop-in clones were selected on trp$^-$lys$^-$ura$^-$ media to select for the YAC arms as well as the integration event with the *URA3* containing HS2 YIp. The appropriate integration events in the pop-in clones were screened for by PCR using primers 01-09, 01-10 and the T7 primer in the YIp vector sequence and Southern blotting using several digestions with HS2 and *URA3* probes (figure 3.6 and 3.7). Several clones with the appropriate integration events were obtained from the SmaI linearised YIps namely: 1.7.13.1, 1.7.13.3, 1.7.16.1, 1.7.16.16, 1.7.16.17, 1.7.25.5, 1.7.31.1, and 1.7.31.2.

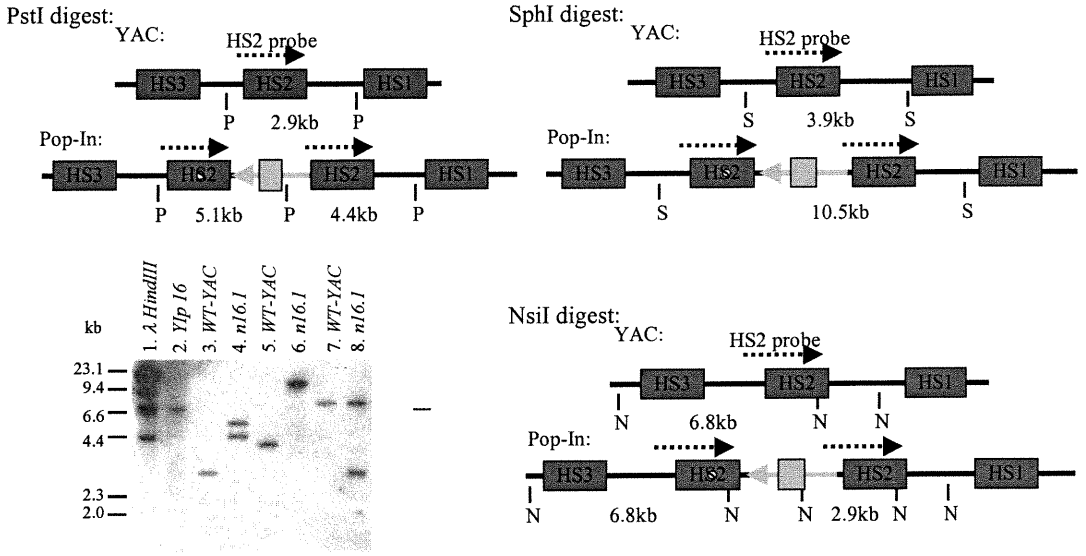


Fig 3.6: Southern blot on HS2 pop-in clone 1.7.16.1 using the HS2 probe. Diagrams indicate expected band lengths based on restriction fragment length and probe location. Note that the HSs are darkly shaded while the *URA3* gene and pRS406 vector sequence are lighter. The HPFH polymorphisms appear as white spots. HindIII digested λ phage (lane 1) and linearized YIp (lane 2) were used as molecular weight markers. Yeast genomic DNA was digested with: PstI (lanes 3 and 4) to yield the expected bands at 2.9kb in the wild type YAC and 4.4 and 5.1kb in the pop-in clone, SphI (lanes 5 and 6) to yield expected bands at 3.9kb in the wild type YAC and 10.5kb in the pop-in clone, NsiI (lanes 7 and 8) to yield expected bands at 6.7kb in the wild type YAC and 2.9 and 6.7 in the pop-in clone.

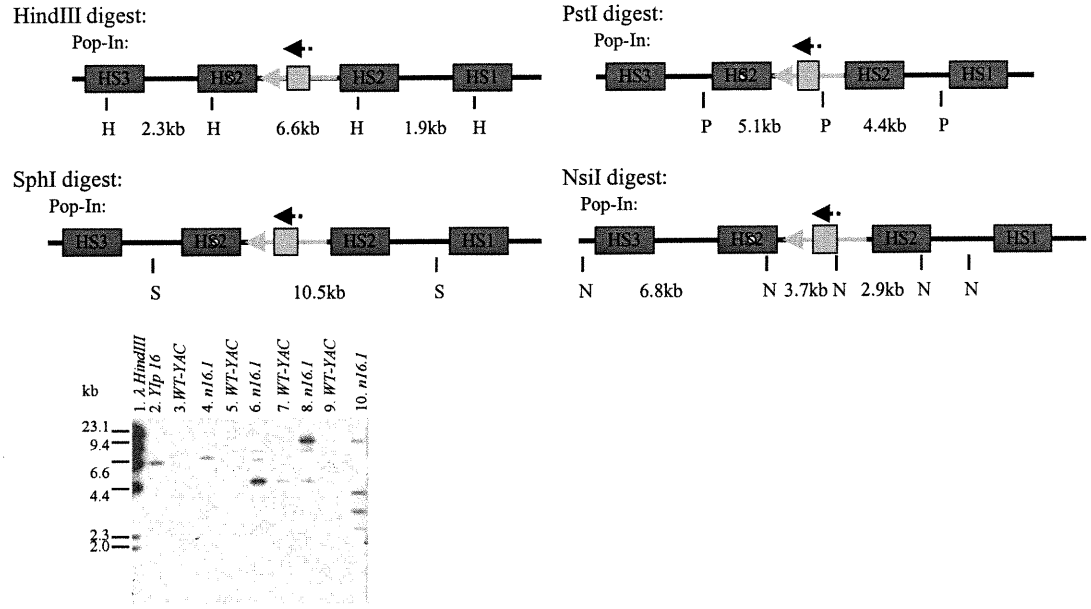


Fig 3.7: Southern blot on HS2 pop-in clone 1.7.16.1 using the *URA3* probe. It is of note that bands in both the wild type YAC and the pop-in clones correspond to the disrupted endogenous *ura3* gene and fragments of *ura3* remaining in the YAC arms after LYS2 retrofitting which hybridise weakly. In contrast, intense specific bands were observed in pop-in clones. Yeast genomic DNA was digested with: HindIII (lanes 3 and 4) yielding the expected band at 6.6kb in the pop-in clone. PstI (lanes 5 and 6) to yield the expected band at 5.1kb in the pop-in clones, SphI (lanes 7 and 8) to yield the expected band 10.5kb in the pop-in clone, NsiI (lanes 9 and 10) to yield the expected bands at 2.9 and 3.7kb in the pop-in clone.

These pop-in clones were subsequently selected for the excision of the YIp. When the selection marker in the integrated YIp is selected against, the YIp is excised by a second round of homologous recombination. This excision event is random and therefore can remove either the wild-type sequence of the YAC or the mutated sequence initially introduced with the YIp. The proportion of excision events which leave the mutated YIp sequence is proportional to the ratio of the amounts of flanking sequence on either side of the mutation contained in the YIp construct. In this case, the excision of the YIp was done by selecting against the *URA3* gene contained within the HS2 YIp by plating the pop-in clones on 5-FOA media (7). Clones obtained after 5-FOA selection were screened by PFGE and hybridized with an HS2 probe in order to assess whether gross deletions, inserts or other rearrangements had occurred during the series of recombination events (fig. 3.8). These clones were then assessed for appropriate excision events by the same Southern blot procedure used on the pop-in clones. In the case of the HS2 pop-out YACs, this involved the shift of bands back to those expected for the wild type YAC since the HS2 HPFH polymorphism deletes a dinucleotide and does not introduce or remove any restriction sites (fig. 3.9).

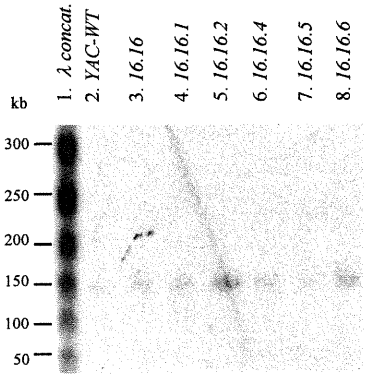


Fig 3.8: PFGE and Southern blot on wild type YAC (lane 2), pop-in clone (lane 3) and pop-out clones (lanes 4 to 8) using an HS2 probe. Yeast plugs were run on PFGE to yield the expected band at 150kb indicating that no gross insertions or deletions had occurred

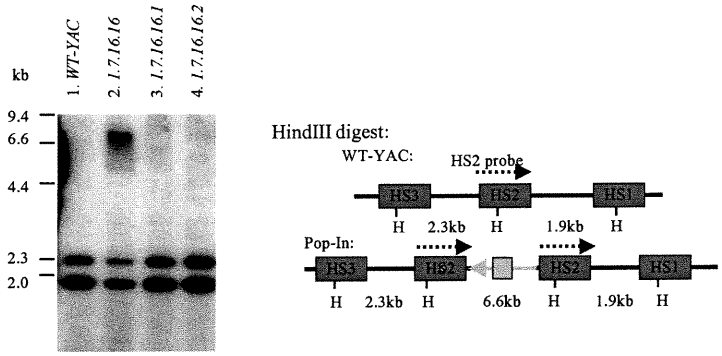


Fig 3.9: Southern blot on pop-out clones using a HS2 probe. Yeast genomic DNA was digested with HindIII to yield the expected bands at 1.9 and 2.2kb in the wild-type YAC (lane 1) with an extra band at 6.6kb in the pop-in clone (lane 2) corresponding to the YIp insert. The pop-out clones (lanes 3 and 4) retain the wild type YAC bands without the 6.6kb YIp derived band, indicating that the correct excision event occurred. Virtually all pop-out clones had undergone normal excision events as determined by this method.

Because the HS2 YIp construct has roughly equal amounts of flanking sequence on either side of the HPFH polymorphism, half of the pop-out clones obtained should theoretically retain the polymorphism and the other half should retain the wild type sequence. On this basis, 10 HS2 pop-out clones, which appeared intact based on the Southern blot and PFGE analyses described above, were sequenced in the polymorphic region. Unfortunately, the sequencing results revealed that the HS2 pop-out clones had all retained the wild type sequence. This indicated that homologous recombination events which yielded pop-out clones with the HPFH polymorphic sequence were either non-existant or relatively infrequent. In order to eliminate the possibility that the reversion to the wild type sequence in the pop-outs was due to a defect in the pop-ins or the HS2 YIp clones used in this experiment, other pop-ins generated from independent HS2 YIp clones were used to produce more pop-out clones. To analyse pop-out clones more expeditiously, two screening approaches were devised to distinguish the wild type from the polymorphic sequence. The first was by SSCP of a PCR fragment encompassing the polymorphic region in HS2. The second was by restriction digest of

the same PCR fragment which yielded fragment sizes of 36bp in the case of the wild type HS2 and 34bp for the HPFH polymorphic HS2. The SSCP and restriction fragment size assays allowed for the relatively rapid screening of over 150 HS2 pop-out clones derived from 5 independent pop-in clones generated with 3 independent YIps (similar to fig. 3.11). All of the 150 HS2 pop-out clones analysed by these methods, all were found to have the wild type sequence and had therefore not retained the HS2 HPFH polymorphism.

These results prompted the exploration of other means to integrate the HS2 YIp. Perhaps, slightly different sequence alignments in the HS2 region could increase the probability of obtaining pop-out clones with the HS2 HPFH polymorphic sequence. One of the options was to integrate the HS2 YIp using a restriction endonuclease which introduced a dsDNA break 3' to the polymorphic sequence as opposed to SmaI used in the previous experiment which cuts 5' to the polymorphism. The differences in sequence alignment are shown in figure 3.12. The BglII linearised HS2 YIps, 1.7.13 and 1.7.16 were integrated into the β-globin YAC to generate 8 pop-in clones each, which were verified by Southern blot. In addition, 18 clones were found to have integrated more than 1 copy of the BglII digested YIp in the HS2 region (fig. 3.10). The insertion of more than 1 YIp copy was probably due to an accidental extended incubation of competent yeast with linearised YIp during the transformation process.

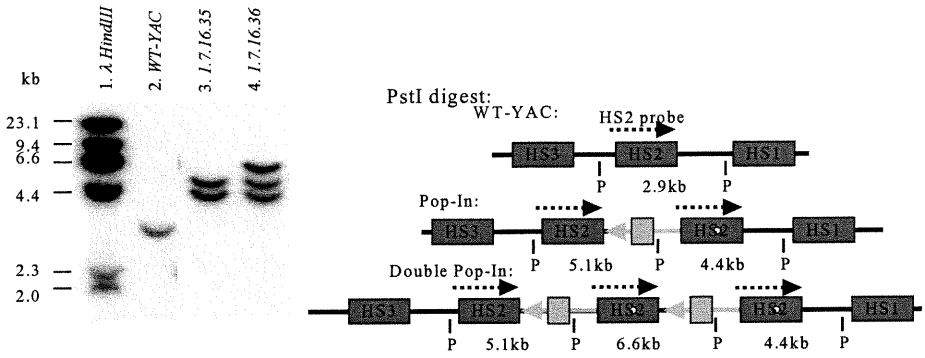


Fig 3.10: Southern blot on HS2 pop-in clones using genomic yeast DNA with an HS2 probe. Pop-in clones were obtained by transforming BglII digested HS2 YIp 1.7.16 into AB1380. PstI digestion yields bands at 2.9kb in the wild type YAC while pop-in clones with 1 YIp copy (1.7.16.35) display bands at 4.4 and 5.1kb. Pop-in clones containing more than 1 YIp copy show an extra band at 6.6kb corresponding to the HS2 YIp compared to the 1 copy pop-ins (1.7.16.36).

The HS2 pop-in clones with single YIp inserts obtained from BglII digested YIp were selected on 5-FOA to obtain pop-out clones. Since the vast majority of pop-out clones previously obtained with the SmaI digested YIp were intact as determined by Southern blotting and PFGE, these clones were first screened for the restriction fragment length of the HS2 polymorphic region. Unfortunately, similarly to the 150 clones generated with SmaI linearised HS2 YIp, all 29 pop-out clones derived from BglII linearised YIp had the wild type fragment length and therefore did not contain the

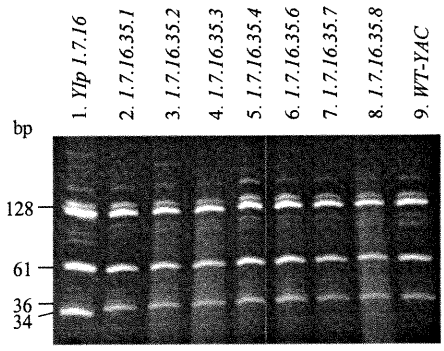


Fig 3.11: Restriction digest of the PCR amplified HS2 polymorphic region run on 20% polyacrylamide gel. Digestion by MfeI and SnaBI gives a 36bp fragment in the case of the wild type YAC (lane 9) and a 34bp fragment in the HPFH HS2 Yip (lane 1). All the pop-out clones (lanes 2-8) have retained the wild type sequence.

HPFH HS2 polymorphism (fig.3.11). This result indicates that the HS2 polymorphism cannot be introduced into the β-globin YAC by changing the location of YIp linearisation (fig. 3.12). It is paradoxical that the location of recombination for the pop-out is invariably in the same region as the homologous recombination event which generated the pop-in. In theory these two recombination events should be completely independent of each other.

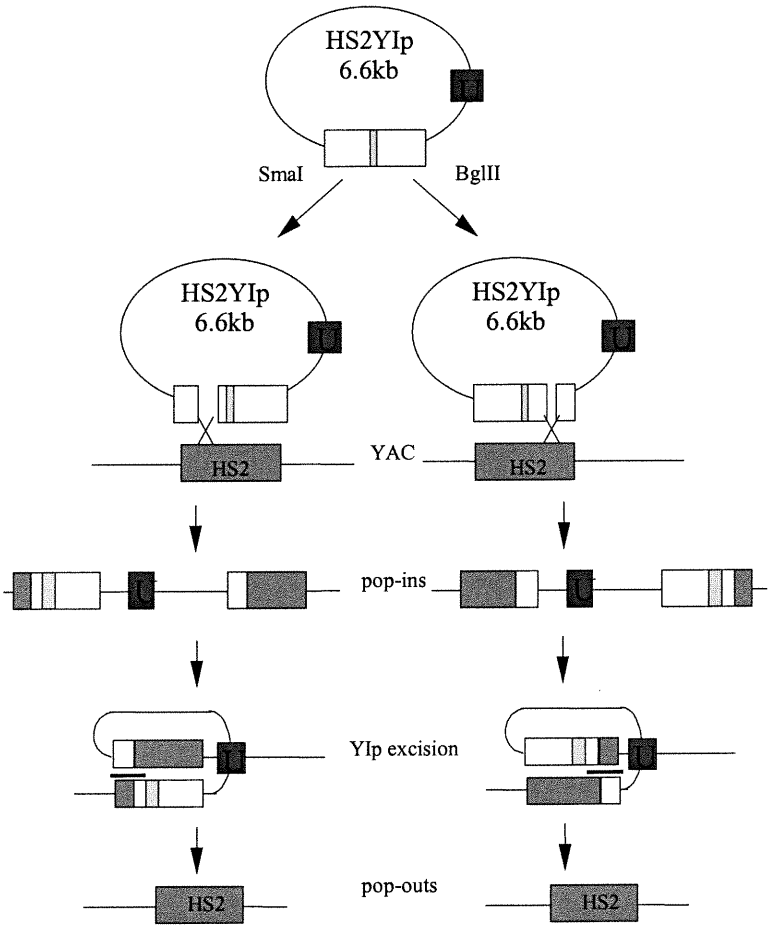


Fig 3.12: Homologous recombination in yeast. With the HS2 construct, pop-in with SmaI digested YIp yields pop-out clones which retain wild-type sequence. Therefore the second homologous recombination event occurs uniquely 5' to the polymorphic sequence as indicated by the solid bar. Conversely, when pop-ins are generated with BglII digested YIp, the pop-out clones undergo homologous recombination 3' of the polymorphic sequence, again retaining exclusively wild type sequence.

3.2 Generating the $^{G}\gamma$IVSII HPFH polymorphic YAC.

The $^{G}\gamma$ HPFH polymorphism was also introduced using the pop-in and pop-out system. To this end, a $^{G}\gamma$ polymorphic YIp was generated with a blunted StuI to SphI (genbank # 34094 to 36686) fragment of a polymorphic $^{G}\gamma$ sequence. This $^{G}\gamma$ segment was cloned from a sickle cell patient and was found to contain the HPFH polymorphism by sequencing. Of note, the polymorphism bordered by approximately 1.4kb of 5' sequences and 1.1kb of sequences in 3'. Therefore this fragment contains sufficient flanking sequences for efficient integration into yeast chromosomes. YIps with inserts in both orientations were generated and verified by restriction digest (fig. 3.13). Pop-in clones were generated by the same method as mentioned for the HS2 HPFH polymorphism. For the transformation, YIps with both insert orientations (clones 2.19 and 2.22) were linearised with EcoNI which cuts in the 5' of the $^{G}\gamma$ insert. After screening by Southern blot with a $^{G}\gamma$ probe, 6 clones had appropriate integration events. Of these 6 clones, 2.19.2, 2.19.3, 2.19.6, 2.19.14, and 2.19.18 were from a YIp with an inverted $^{G}\gamma$ insert while one clone, 2.22.9, was generated with a YIp containing a direct $^{G}\gamma$ insert.

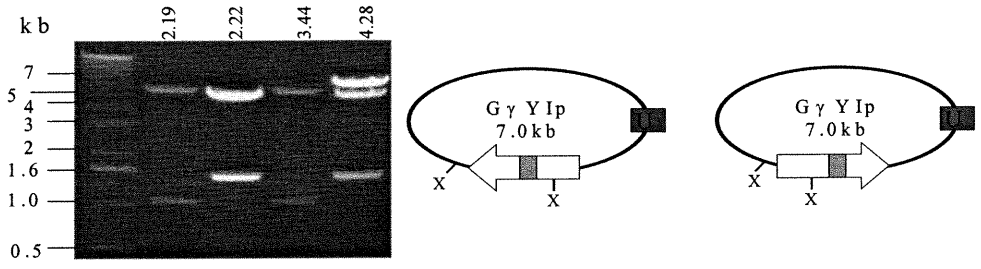


Fig 3.13: Diagnostic digest of the $^{G}\gamma$ YIp. XbaI digestion gives 1.1 and 5.9kb fragments when inserted in the reverse orientation and 1.5 and 5.4kb fragments when inserted directly.

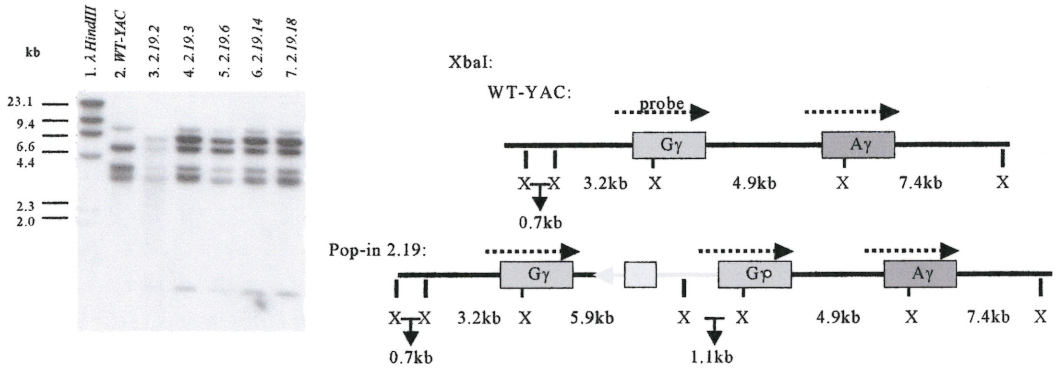


Fig 3.14: Southern blot on $^{G}\gamma$ pop-in clones from YIp 2.19 using a $^{G}\gamma$ probe. XbaI digestion yields extra bands at 1.1 and 5.9kb in the pop-in clones (lanes 3 to 7) as compared to the wild type YAC (lane 2). The numerous other bands are all accounted for by hybridization of the probe to $^{G}\gamma$ and homologous $^{A}\gamma$ sequences.

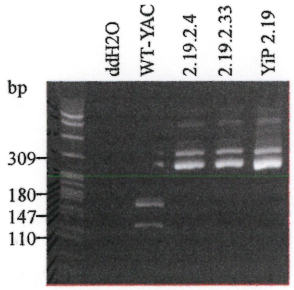


Fig 3.15: Screening of $^{G}\gamma$ pop-outs by BsaHI digest of the PCR amplified $^{G}\gamma$ polymorphic region. The wild type YAC has a BsaAI site in this region which generates 178 and 118bp fragments. The HPFH polymorphic YIp 2.19.2 does not contain this restriction site and remains undigested. The pop-outs 2.19.2.4 and 2.19.2.33 also remain undigested indicating that they retained the YIp sequences.

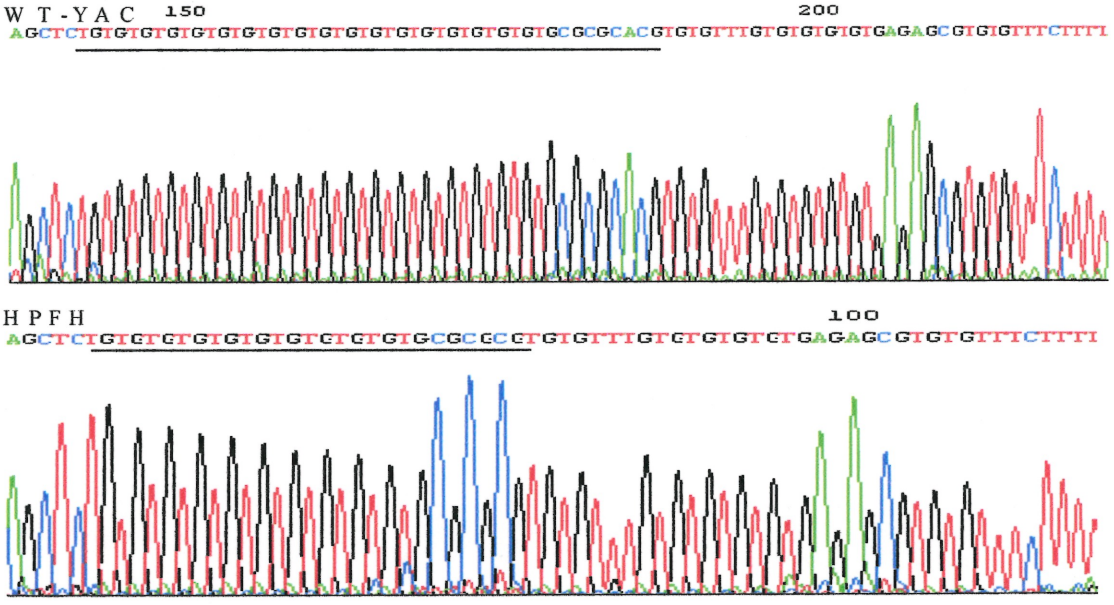


Fig 3.16: Sequencing of $^{G}\gamma$ polymorphic region with primer 00-53. Note that the wild type YAC (above) contains the sequence $(TG)_{19}(CG)_2CACG$ while the $^{G}\gamma$ pop-out clone 2.19.2.4 (below) contains the HPFH polymorphic sequence $(TG)_{11}(CG)_3$. The pop-out clone 2.19.2.33 was also found to have the correct HPFH sequence.

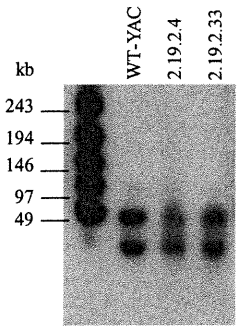


Fig 3.17: Southern blot on PFGE using left (LYA) and right (RYA) YAC arm probes. As expected, SfiI and SalI double digestion allows hybridization of a 30kb fragment with LYA probe and a 50kb fragment with RYA probe.

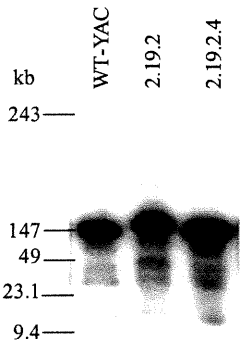


Fig 3.18: Southern blot on PFGE using $^{G}\gamma$ and HS1-4 probes. KpnI digest allows hybridization of a 40kb fragment with $^{G}\gamma$ and 30 and 13kb fragments with HS1-4.

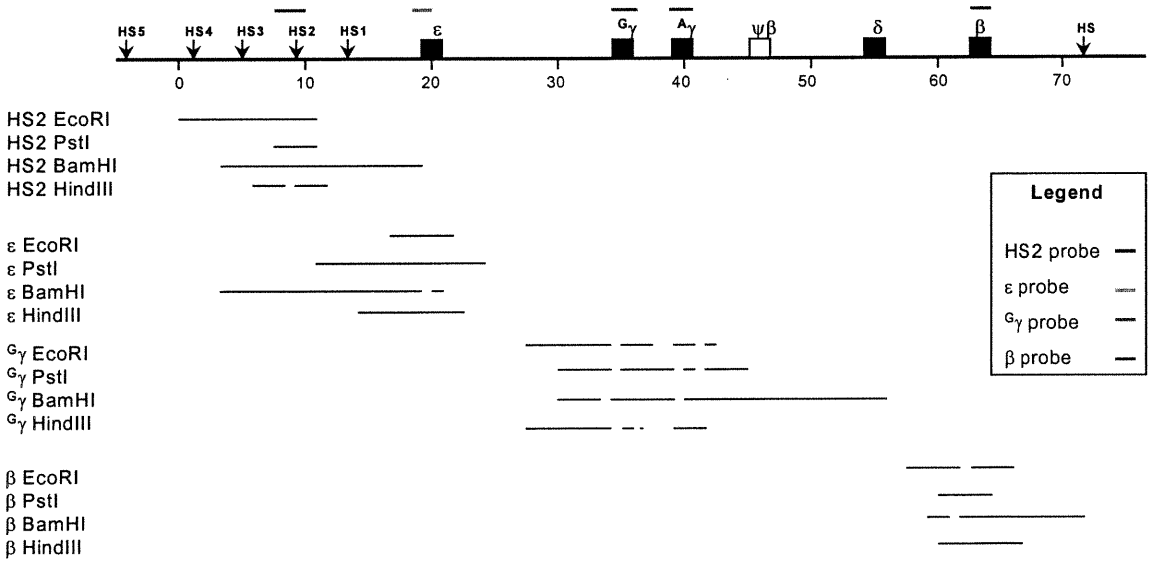


Fig 3.19: Schematic representation of digestions and probes used to assess the integrity of the pop-out YACs by Southern blotting. Note that almost the entire β-globin locus is covered by four digestions, BamHI, EcoRI, HindIII and PstI, when hybridised with probes in HS2, ε, $^{G}\gamma$ and β.

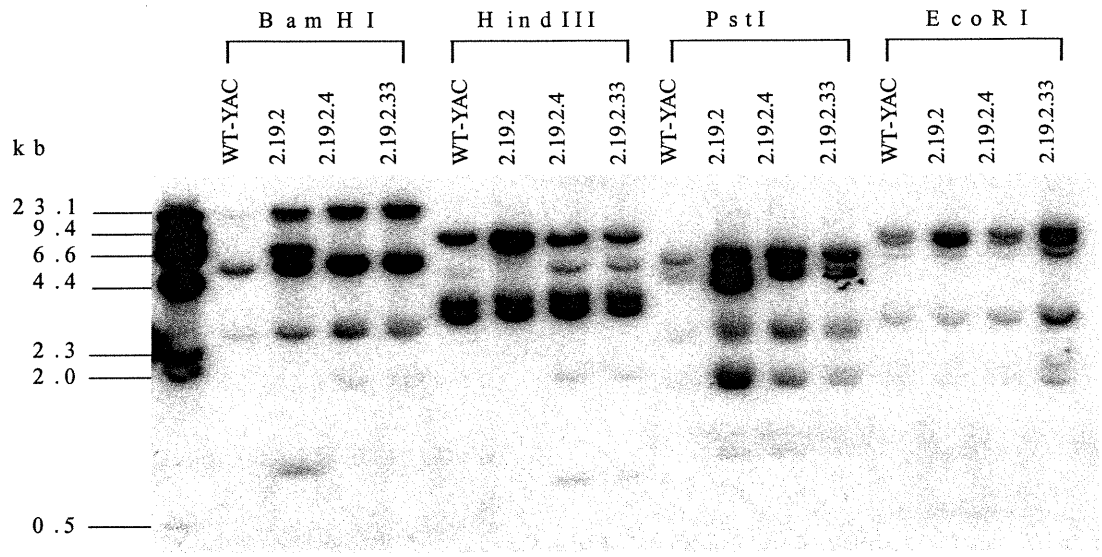


Fig 3.20: Southern blot on yeast genomic DNA using a $^{G}\gamma$ probe. $^{G}\gamma$ pop-out clones containing the HPFH polymorphic sequence yielded appropriate bands when digested: BamHI yields bands at 2.6, 4.9 and 15.3kb for the wild type YAC and pop-out clones (2.19.2.4 and 2.19.2.33) while the pop-in clone (2.19.2) has extra bands at 0.9 and 6.1kb corresponding to the YIp insert. HindIII yields bands at 3.3 and 6.8kb for the wild type YAC and pop-out clones while the pop-in has an extra band at 6.1kb corresponding to the YIp insert. PstI yields bands at 2.7, 4 and 4.9kb for the wild type YAC and pop-out clones while the pop-in clone has extra bands at 1.7 and 3.7kb corresponding to the YIp insert. EcoRI yields bands at 1.6, 2.6 and 7kb for the wild type YAC, pop-in and pop-out clones (the YIp insert gives a band at 7kb). Extra bands are accounted for by hybridization of the left YAC arm and YIp vector sequences with vector sequences remaining in the probe preparation.

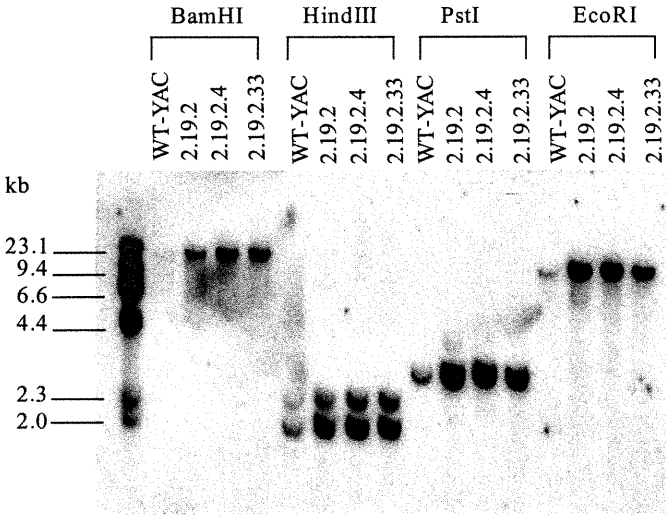


Fig 3.21: Southern blot on yeast genomic DNA using an HS2 probe. $^{G}\gamma$ pop-out clones containing the HPFH polymorphic sequence yielded appropriate bands when digested: BamHI yields bands at 25kb for the wild type YAC, pop-in clone (2.19.2) and pop-out clones (2.19.2.4 and 2.19.2.33). HindIII yields bands at 1.9 and 2.2kb for the wild type YAC, pop-in and pop-out clones. PstI yields bands at 2.9kb for the wild type YAC, pop-in and pop-out clones. EcoRI yields bands at 10.5kb for the wild type YAC, pop-in and pop-out clones.

Pop-in $^{G}\gamma$ clones 2.19.2 and 2.22.9, generated by YIps with opposite orientations of $^{G}\gamma$ insert, were chosen for selection on 5-FOA and the excision of the YIp sequence to produce pop-out clones. Since, the $^{G}\gamma$ YIp construct contains roughly equal amounts of flanking sequence on either side of the polymorphism and therefore the pop-outs should retain the wild type or HPFH polymorphic sequences in equal amounts. Fortunately, the $^{G}\gamma$ polymorphism abolishes a BsaAI restriction site found in the wild type sequence. Of the 8 pop-out clones analysed, 3 would not digest with this enzyme (fig. 3.15). Of these 3 clones, 2 (2.19.2.4 and 2.19.2.33) were found to have the appropriate sequence (fig. 3.16). To verify the integrity of the $^{G}\gamma$ pop-out YACs, which had undergone 2 homologous recombination events, PFGE was performed on SfiI/SalI and KpnI digests of these $^{G}\gamma$ polymorphic YACs. These gels were then Southern blotted for the LCR, the globin genes and the left and right YAC arms as shown in figs. 3.17 and 3.18. These results indicated that neither 2.19.2.4 or 2.19.2.33 had undergone large rearrangements, insertions or deletions. Therefore the integrity of the β-globin locus was examined with more scrutiny for smaller rearrangement events which could not be detected by PFGE. Southern blotting was performed by hybridizing 4 restriction digests with 4 probes from within the β-globin locus as illustrated in fig. 3.19. This enzyme and probe combination covers the entire 70kb region containing the LCR and the β-like genes. Both of the $^{G}\gamma$ pop-out clones with the HPFH sequence had the same Southern blotting profile as the wild type YAC when analysed by this method (figs. 3.20 and 3.21). In summary, the results indicate that these $^{G}\gamma$ HPFH polymorphic YACs are intact and have not undergone any rearrangements, insertions or deletions.

Chapter 4: Discussion.

4.1 Introduction of the HPFH HS2 polymorphism into the β-globin YAC.

4.1.1 Integration of the HS2 YIp by homologous recombination.

Of the four polymorphisms found in the HPFH individual, two were selected to assess their contribution to increased HPFH production in the 150kb YAC system. The HS2 polymorphism was successfully cloned into a YIp with approximately 1.2kb of flanking sequence 5' of the polymorphic region and 1.1kb of flanking sequence in the 3' direction. These large flanking sequences are required for efficient homologous recombination into the β-globin YAC. The first round of homologous recombination is termed the pop-in. For this process, several independent YIps were linearised and transformed into β-globin YAC containing yeast. This favors homologous recombination with the YIp and the β-globin YAC HS2 segment rather than the YIp and the disrupted *ura3* gene found in the AB1380 yeast strain. Subsequent plating on media without uracil, selects for clones which have integrated the YIp into the β-globin YAC. In the case of HS2, 24 clones with appropriate integration events were obtained as determined by Southern blotting, indicating that homologous recombination occurs normally in the β-YAC containing AB1380 yeast strain.

4.1.2 The excision of the HS2 YIp by homologous recombination.

The second round of homologous recombination is termed the pop-out. Yeast with these recombination events will survive selection against the *URA3* gene contained in the integrated YIp. The homologous segments of HS2 should in theory line up and undergo homologous recombination such that the YIp vector is excised

with an approximately 50% chance of retaining the wild type sequence and therefore a 50% chance of retaining the HPFH polymorphic sequence. This 50% chance is based on the roughly equal amount of flanking sequence on either side of the polymorphic sequence. However, this was not the case; of the 160 clones screened by sequencing and SSCP not one was found to have retained the HPFH polymorphic sequence. The ability to generate pop-in clones indicates that the yeast strain used is capable of the homologous recombination process. In addition, several independent YIp clones and pop-in clones were used to generate pop-outs, suggesting that the problem is not due to the YIp or incorrectly popped-in YIps. In an attempt to reverse the alignment of the sequences for the pop-out step, the YIp insert was popped-in using a restriction enzyme which cut on the opposite side (3') of the polymorphic region. This yielded 16 pop-ins which contained one YIp insert while 18 other clones were found to have integrated more than one copy of the Yip, as was determined by Southern blot. When 29 pop-out clones obtained from the pop-ins with a single integrated YIp were screened for the polymorphism by restriction fragment size, no clones had retained the HPFH sequence. These pop-in clones were obtained from two different Yips, indicating that the problem is one that pertains to the homologous recombination process and is not due to an artefact caused by a defective YIp or defective pop-in clones.

Since the yeast strain is capable of homologous recombination, it would suggest that this HPFH polymorphic region is intrinsically refractory to this process. This may be due to the sequence contained in the polymorphic region which is $(AT)_9N_{12}(AT)_{11}$ which should be able to form a stable hairpin structure. Perhaps these hairpins could impair access of proteins involved in the homologous recombination process to the

HS2 region. It is also possible that these structures could prevent homologous recombination from proceeding in a normal fashion by interfering with branch migration and preferentially selecting certain ways of resolving the Holliday junction. One method to counter these problems would be to force homologous recombination on the desired side of the construct. A YIp containing a very small 5' flanking region and a large 3' flanking region could be produced. Theoretically if this construct were introduced by a break in the short 5' moiety, then subsequent pop-outs would have no choice but to undergo recombination in the 3' flanking region and, in the process, retain the HPFH polymorphism. Alternatively, several pop-in clones have already been generated with 2 or more YIp inserts. Since the selection to generate pop-out clones is against functional *URA3* genes, all YIp copies would be excised as occurs in normal pop-in clones if placed on 5-FOA media. Perhaps, the positioning of sequences would differ sufficiently during the homologous recombination process to allow for the retention of the HPFH polymorphic sequence in the YAC. Indeed, another student in the Trudel laboratory has recently generated pop-out clones with the HS2 HPFH polymorphic sequence from one of these pop-in clones with 2 inserted YIps. This supports the hypothesis that the secondary structure of DNA can affect the outcome of homologous recombination events.

4.2 Introduction of the HPFH $^{G}\gamma$ polymorphism into the β-globin YAC.

4.2.1 Integration of the $^{G}\gamma$ YIp by homologous recombination.

For the introduction of the $^{G}\gamma$ polymorphism, the same technique of homologous recombination in yeast was used. This time the YIps contained the HPFH

polymorphism in $^{G}\gamma$ IVSII as well as 1.4kb of 5' flanking sequence and 1.1kb of 3' flanking sequence. These YIps were generated with the $^{G}\gamma$ insert in both directions with respect to the vector and both types of YIp were used to generate pop-in clones with the appropriate insertion events as determined by Southern blotting. The *URA3* gene was selected against in these clones to generate pop-out clones. Of the first 8 clones which were screened by restriction digest of a PCR generated DNA fragment encompassing the polymorphic region, 3 were positive. Out of these 3 clones, 2 were deemed to have retained the β-globin YAC in its native conformation, without other rearrangements as determined by Southern blots covering the approximately 70kb region containing the LCR and the genes. The regions outside of these areas were assessed for integrity by Southern blotting on PFGE and probing for the LCR, the β-like genes and the YAC arms. Therefore, 2 β-globin YAC clones have been generated with the $^{G}\gamma$ HPFH sequence and can be used for the generation of transgenic mice to assess the role of this sequence in fetal gene expression and regulation. It is of outmost importance to note that the $^{G}\gamma$ manipulations in yeast were done in the same strain and under the same conditions as were used for HS2. This is a further indication that the problem with the previous construct is sequence related and not due to technical problems. Due to the problems encountered with the HS2 polymorphic YAC, the $^{G}\gamma$ YIp should now be inserted into the HS2 polymorphic YAC in order to generate a compound HS2 and $^{G}\gamma$ polymorphic β-globin YAC.

4.3 The role of the HS2 and $^{G}\gamma$ polymorphisms in globin regulation.

Since transgenic mice carrying YACs with HPFH mutations can express higher levels of HbF, it can be inferred that some of the factors which regulate the fetal to adult globin switch are the same in the mouse and in the human. On this basis and provided that the HPFH polyporphisms have a role in hemoglobin switching, mice carrying the HS2 HPFH polymorphic YAC, the $^{G}\gamma$ polymorphic YAC or the compound polymorphic YAC could have increased levels of HbF in adulthood (54). However, in this respect the homozygosity of HPFH individual is worrisome. It supposes a certain level of cosanguinity and therefore this individual may be homozygous for other mutations and polymorphisms in the β-globin locus and elsewhere in the genome. There is a distinct possibility that the polymorphisms in HS2, $^{G}\gamma$, $^{A}\gamma$ and 5' β found in the HPFH individual and his family, are simply associated with HPFH and other sequences elsewhere in the β-globin locus are truly responsible for the vast amount of HbF produced in these adult individuals.

4.3.1 Importance of HS2 in fetal to adult globin gene switching.

Several reasons justified the study of the HS2 and $^{G}\gamma$ polymorphisms in the YAC system. The 3' HS2 was further studied in the 150kb YAC because it is the only polymorphic region in the LCR associated with increased HbF production in thalassemics and SCD patients. Once the HS2 polymorphic YAC is used to generate transgenic mice, the patterns of β-like globin gene expression may elucidate the role of this polymorphic region in HS2 function: No increase in HbF production at any stage of development could indicate that either the 3' HS2 exerts its effects through factors

which are not present in the mouse or the HS2 could require other sequences to increase fetal gene expression. Alternatively, HS2 polymorphic YAC transgenic mice could demonstrate an increase in fetal gene expression exclusive to the fetal stage. This would imply that the 3' HS2 has a fetal stage specific function which is in sharp contrast to the HPFH individual who expresses the fetal genes at high levels in adulthood. This could be caused by factors which bind 3' HS2 that are expressed through the fetal and adult stages in the human but only in the fetal stages of the mouse. Conversely, an increase in HbF specific to the fetal stages could implicate the other polymorphisms of the HPFH individual in the expression of the fetal genes during adulthood. Finally, it is also possible that these HS2 polymorphic YAC transgenic mice would express the fetal genes at moderate to important levels during the adult stages. This would indicate that the 3' HS2 at the very least, contributes to the increased expression of HbF in the HPFH individual during adulthood and that the factors which mediate this increase are similar in the mouse and the human.

Furthermore, if the latter possibility occurs, it may be possible to determine whether this increase in HbF production is mediated by the HOXB2 binding sites in the polymorphic region. The *hoxb2* knock out mouse has already been created and the vast majority of homozygous pups die immediately after birth due to an improperly closed sternum and presumably an inability to breathe (4, 15). Therefore, it can be assumed that even though erythropoiesis was not examined in these mice, any defect could not have been sufficiently severe to cause the death of these *hoxb2* knock outs. Since the transcriptional switch from fetal to adult globin genes in the YAC transgenic mice occurs prior to day 15 of gestation, it would be interesting to observe whether the

switching from fetal to adult genes is affected in 150kb β-globin YAC transgenic mice with inactivated *hoxb2* genes. If *hoxb2* is important for the expression of the fetal genes then these mice would be expected to show a decline in fetal gene expression or a precocious switch from fetal to adult genes. A similar study comparing these mice and HS2 polymorphic YAC mice with inactivated *hoxb2* genes should yield a similar expression of the human β-like globin genes. This result would be anticipated because the difference in HOXB2 binding sites between the wild type and HS2 polymorphic YACs should be irrelevant on a *hoxb2* null background, and therefore the expression and developmental of the human β-like genes should be the same in both cases. However, if these mice do have different patterns of β-like gene expression it would suggestive of factors other than *hoxb2* acting on the HS2 polymorphic sequence.

4.3.2 Possible roles for γIVSII in increased fetal gene expression.

The $^{G}\gamma$ and $^{A}\gamma$ polymorphisms seem to be related since they occur in the second introns of two extremely homologous genes. Furthermore, these polymorphisms both involve the TG repeats in microsatelite sequences indicating that these polymorphisms could increase fetal gene expression through a common mechanism. It is also possible that the $^{G}\gamma$ polymorphism which was selcted for study in the YAC system, acts exclusively upon the $^{G}\gamma$ gene or on both fetal genes: The $^{G}\gamma$IVSII sequence may locally recruit transcription factors and have a specific activity on its gene or conversely, the $^{G}\gamma$IVSII sequence may recruit factors which increase the accesibility of the DNA to transcription factors over a much wider region including the neighbouring $^{A}\gamma$ gene. Therefore, the observed levels and types of HbF present in transgenic mice containing

the $^{G}\gamma$ HPFH polymorphic YAC could prove to be quite interesting. In the first instance, there could be no increase in fetal hemoglobin production at any stage of development, suggesting that either this polymorphism requires other sequences to increase HbF levels or that the factors which bind this region to increase fetal hemoglobin expression are not present in the mouse. Secondly, there could be an increase in fetal hemoglobin expression due to increased $^{G}\gamma$ or increased $^{G}\gamma$ and $^{A}\gamma$ during the fetal stage only. As was noted for the HS2 polymorphism, this stage specific increase would not resemble the HPFH individual since the expression of the fetal genes would not persist into adulthood. In this case, the $^{G}\gamma$ sequence could not drive the expression of fetal genes in adulthood alone and other sequences in the β-globin locus would be partially or entirely responsible for the expression of HbF in adulthood in the HPFH individual. Alternatively, the lack of expression in adulthood could be explained by a dissimilar developmental regulation of the factors which bind the $^{G}\gamma$ IVSII to increase fetal gene expression, in the human and in the mouse. The latter explanation may still allow an assessment of effect of the $^{G}\gamma$ IVSII polymorphism on the expression of $^{G}\gamma$ and $^{A}\gamma$ individually. Thirdly, it is possible that the expression of HbF in the $^{G}\gamma$ IVSII polymorphic YAC transgenic mice will continue at modest to high levels into adulthood. This would define the $^{G}\gamma$ IVSII as an important regulatory element of the fetal genes *in vivo* and would account for the level of HbF expression in the HPFH individual. It would also imply that the factors which bind this polymorphic sequence are the same in both humans and mice.

4.3.3 Potential for interaction between HS2 and $^{G}\gamma$IVSII.

The linking and looping models for hemoglobin switching predict that the LCR can interact either indirectly or directly with gene proximal sequences. Thus it is theoretically possible that the HPFH polymorphic sequence in HS2 could interact with the $^{G}\gamma$ IVSII, allowing for a synergistic enhancement of fetal gene expression. This possibility renders it imperative to generate a compound HS2 and $^{G}\gamma$ IVSII HPFH polymorphic YAC to test this hypothesis. Furthermore, if a synergistic increase in HbF expression is noticed in compound polymorphic YAC transgenic mice, it would be interesting to determine whether this effect is HOXB2 dependent. This could elucidate the role of HOXB2 binding in HS2 enhancer activity *via* the homeobox dependent linking model.

Bibliography:

1. Acuto, S., Urzi, G., Schimmenti, S., Maggio, A., O'Neill, D., Bank, A. "An element upstream from the human δ-globin-encoding gene specifically enhances β-globin reporter gene expression in murine erythroleukemia cells." Gene. Vol. 168. pp. 237-241. 1996.

2. Ashe, H., Monks, J., Wijgerde, M., Fraser, P., Proudfoot, N.J. "Intergenic transcription and transinduction of the human β-globin locus." Genes and Development. Vol. 11. pp. 2494-2509. 1997.

3. Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, and J.G., Struhl, K. Current Protocols in Molecular Biology. J. Wiley Publishing, USA. 1st ed. 1988. (Chapters: 2.4, 6.1, 15.1)

4. Barrow, J.R., Capecchi, M.R. "Targetted disruption of the Hoxb-2 locus in mice interferes with expression of Hoxb-1 and Hoxb-4." Development. Vol. 122. pp. 3817-3828. 1996.

5. Bender, M.A., Bulger, M., Close, J., Groudine, M. "β-globin gene switching and DnaseI sensitivity of the endogenous β-globin locus in mice do not require the locus control region." Molecular Cell. Vol. 5. pp. 387-393. 2000.

6. Bender, M.A., Mehaffey, M.G., Telling, A., Hug, B., Ley, T., Groudine, M., Fiering, S. "Independent formation of DnaseI hypersensitive sites in the murine β-globin locus control region." Blood. Vol. 95, No. 11. pp. 3600-3604. 2000.

7. Boeke, J.D., Trueheart, J., Natsoulis, G., Fink, G.R. "5-fluoroorotic acid as a selective agent in yeast molecular genetics." Methods in Enzymology. Vol. 154. pp. 164-175. 1987.

8. Bulger, M., Groudine, M. "Looping versus linking: toward a model for long-distance gene activation." Genes and Development. Vol 13. pp. 2465-2477. 1999.

9. Bulger, M., von Doorninck, J.H., Saitoh, N., Telling, A., Farrell., C., Bender, M.A., Felsenfeld, G., Axel, R., Groudine, M. "Conservation of sequence and structure flanking the mouse and human β-globin loci: The β-globin genes are embedded within an array of odorant receptor genes." Proceedings of the National Acedemy of Sciences USA. Vol. 96. pp. 5129-5134. 1999.

10. Bungert, J., Dave, U., Lim, K-C., Lieuw, K.H., Shavit, J.A., Liu, Q., Engel, J.D. "Synergistic regulation of human β-globin gene switching by locus control region elements HS3 and HS4." Genes and Development. VCol. 9. pp. 3083-3096. 1995.

11. Bungert, J., Tanimoto, K., Patel, S., Liu, Q., Fear, M., Engel, J.D. "Hypersensitive site 2 specifies a unique function within the human β-globin locus control region to stimulate globin gene transcription." Molecular and Cellular Biology. Vol. 19, No. 4. pp. 3062-3072. 1999.

12. Case, S.S., Huber, P., Lloyd, J.A. "The γPE complex contains both SATB1 and HOXB2 and has positive and negative roles in human γ-globin gene regulation." DNA and Cell Biology. Vol. 18, No. 11. pp. 805-817. 1999.

13. Cavallesco, R., Tuan, D. "Modulatory domains of the HS2 enhancer differentially regulate enhancer activity in erythroid cells at different developmental stages." Blood Cells, Molecules, and Diseases. Vol. 23, No. 2. pp. 8-26. 1997

14. Collis, P., Antoniou, M., Grosveld, F. "Definition of the minimal requirements within the human β-globin gene and the dominant control region for high level expression." EMBO Journal. Vol. 9, No. 1. pp. 233-240. 1990.

15. Davenne, M., Maconochie, M.K., Neun, R., Pattyn, A., Chambon, P., Krumlauf, R., Rijli, F.M. "Hoxa2 and Hoxb2 control dorsoventral patterns of neuronal development in the rostral hindbrain." Neuron. Vol. 22. pp. 677-691. 1999.

16. Donovan-Peluso, M., Acuto, S., Swanson, M., Dobkin, C., Bank, A. "Expression of human γ-globin genes in human erythroleukemia (K562) cells." Journal of Biological Chemistry. Vol. 262, No. 35. pp. 17051-17057. 1987.

17. Ebb, D., Tang, D.C., Drew, L., Chin, K., Berg, P.E., Rodgers, G.P. "Identification of upstream regulatory elements that repress expression of adult β-like globin genes in a primitive erythroid environment." Blood Cells, Molecules, and Diseases. Vol. 24, No. 18. pp. 356-369. 1998.

18. Elion, J., Berg, P., Lapoumeroulie, C., Trabuchet, G., Mittelman, M., Krishnamoorthy, R., Schecter, A.N., Labie, D. "DNA sequence variation in a negative control region 5' to the β-globin gene correlates with the phenotypic expression of the β^s mutation." Blood. Vol. 79, No. 3. pp. 787-792. 1992.

19. Farrell, C.M., Grinberg, A., Huang, S.P., Chen, D., Pichel, J.G., Westphal, H., Felsenfeld, G. "A large upstream region is not necessary for gene expression or hypersensitive site formation at the mouse β-globin locus." Proceedings of the National Acedemy of Sciences USA. Vol. 97, No. 26. pp. 14554-14559. 2000.

20. Fiering, S., Epner, E., Robinson, K., Zhuang, Y., Telling, A., Hu, M., Martin, D.I.K., Enver, T., Ley, T.J., Groudine, M. "Targetted deletion of 5' HS2 of the murine β-globin LCR reveals that it is not essential for proper regulation of the β-globin locus." Genes and Development. Vol. 9. pp. 2203-2213. 1995.

21. Forrester, W.C., Epner, E., Driscoll, M.C., Enver, T., Brice, M., Papayannopoulou, T., Groudine, M. "A deletion of the human β-globin locus activation region causes a major alteration in chromatin structure and replication across the entire β-globin locus." Genes and Development. Vol. 4. pp. 1637-1649. 1990.

22. Forrester, W.C., Novak, U., Gelinas, R., Groudine, M. "Molecular analysis of the human β-globin locus activation region." Proceedings of the National Academy of Sciences USA. Vol. 86. pp. 5439-5443. 1989.

23. Francastel, C., Walters, M.C., Groudine, M., Martin, D.I.K. "A functional enhancer suppresses silencing of a transgene and prevents its localisation close to centromeric heterochromatin." Cell. Vol. 99. pp. 259-269. 1999.

24. Fraser, P., Hurst, J., Collis, P., Grosveld, F. "DNaseI hypersensitive sites 1,2 and 3 of the human β-globin dominant control region direct position-independent expression." Nucleic Acids Research. Vol. 18, No. 12. pp. 3503-3508. 1990.

25. Gaensler, K.M.L., Burmeister, M., Brownstein, B.H., Taillon-Miller, P., Myers, R.M. "Physical mapping of yest artificial chromosomes containing sequences from the human β-globin gene region." Genomics. Vol. 10. pp. 976-984. 1991.

26. Gaensler, K.M.L., Kitamura, M. Kan, Y.W. "Germ-line transmission and developmental regulation of a 150kb yeast artificial chromosome containing the β-globin locus in transgenic mice." Proceedings of the National Acedemy of Sciences USA. Vol. 90. pp. 11381-11385. 1993.

27. Giampaolo, A., Sterpetti, P., Bulgarini, D., Samoggia, P., Pelosi, E., Vatieri, M., Peschle, C. "Key functional role and lineage specific expression of selected HOXB genes in purified hematoietic progenitor differentiation." Blood. Vol. 84, No. 11. pp. 3637-3647. 1994.

28. Gilman, J.G., Huisman, T.H.J. "DNA sequence variation associated with elevated fetal $^{G}\gamma$-globin production." Blood. Vol. 66, No. 4. pp. 783-787. 1985.

29. Gilman, J.G., Huisman, T.H.J. "Two independent genetic factors in the β-globin gene cluster are associated with high $^{G}\gamma$-levels in the HbF of SS patients." Blood. Vol. 64, No. 2. pp. 452-457. 1984.

30. Gribnau, J., Diderich, K., Pruzina, S., Calzolari, R., Fraser, P. "Intergenic transcription and developmental remodeling of chromatin subdomains in the human β-globin locus." Molecular Cell. Vol. 5. pp. 377-386. 2000.

31. Gumucio, D.L., Heilstedt-Williamson, H., Gray, T.A., Tarle, S.A., Shelton, D.A., Tagle, D.A., Slightom, J.L., Goodman, M., Collins, F.S. "Phylogenetic footprinting

reveals a nuclear protein which binds to silencer sequences in the human γ and ε globin genes." Molecular and Cellular Biology. Vol. 12, No. 11. pp. 4919-4929. 1992.

32. Gumucio, D.L., Shelton, D.A., Bailey, W.J., Slightom, J.L., Goodman, M. "Phylogenetic footprinting reveals unexpected complexity in trans factor binding upstream from the ε-globin gene." Proceedings of the National Academy of Sciences USA. Vol. 90. pp. 6018-6022. 1993.

33. Hardison, R., Slightom, J.L., Gumucio, D.L., Goodman, M., Stojanovic, N., Miller, W. "Locus control regions of mammalian β-globin gene clusters: combining phylogenetic analyses and experimental results to gain functional insights." Gene. Vol. 205. pp. 73-94. 1997.

34. Hug, B.A., Wesselschmidt, R.L., Fiering, S., Bender, M.A., Epner, E., Groudine, M, Ley, T. "Analysis of mice containing a targetted deletion of β-globin locus control region 5' hypersensitive site 3." Molecular and Cellular Biology. Vol. 16, No. 6. pp. 2906-2912. 1996.

35. Jackson, C.E., O'Neill, D., Bank, A. "Nuclear factor binding in human β-globin IVS2." Journal of Biological Chemistry. Vol. 270, No. 47. pp. 28448-28456. 1995.

36. Kong, S., Bohl, D., Li, C., Tuan, D. "Transcription of the HS2 enhancer toward a cis-linked gene is independent of the orientation, position and distance of the enhancer relative to the gene." Molecular and Cellular Biology. Vol. 17, No. 7. pp. 3955-3965. 1997.

37. Lanclos, K.D., Oner, C., Dimovski, A.J., Gu, Y-C., Huisman, T.H.J. "Sequence variations in the 5' flanking and IVS-II regions of the $^{G}\gamma$-and $^{A}\gamma$-globin genes of β^{s} chromosomes with five different haplotypes." Blood. Vol. 77, No. 11. pp. 2488-2496. 1991.

38. Langdon, S.D., Kaufman, R.E. "Gamma-globin gene promoter elements required for interaction with globin enhancers." Blood. Vol. 91, No. 1. pp. 309-318. 1998.

39. Lapoumeroulie, C., Castiglia, L., Ruberto, C., Fichera, M., Amata, S., Labie, D., Ragusa, A. "Genetic variations in fetal globin gene microsatellites and their functional relevance." Human Genetics. Vol. 104. pp. 307-314. 1999.

40. Lavelle, D., Ducksworth, J., Eves, E., Gomes, G., Keller, M, Heller, P., DeSimone, J. "A homeodomain protein binds to γ-globin gene regulatory sequences.' Proceedings of the National Acedemy of Sciences USA. Vol. 88. pp. 7318-7322. 1991.

41. Lee, J-S., Lee, C-H., Chung, J.H. "The β-globin promoter is important for recruitment of erythroid Kruppel-like factor to the locus control region in erythroid

cells." Proceedings of the National Acedemy of Sciences USA. Vol. 96. pp. 10051-10055. 1999.

42. Lee, C-H., Murphy, M.R., Lee, J-S., Chung, J.H. "Targetting a SWI/SNF-related chromatin remodeling complex to the β-globin promoter in erythroid cells." Proceedings of the National Acedemy of Sciences USA. Vol. 96, No. 22. pp. 12311-12315. 1999.

43. Leonova, J.Y., Kazanetz, E.G., Smetanina, N.S., Adekile, A.D., Efremov, G.D., Huisman, T.H.J. "Variability in the fetal hemoglobin level of the normal adult." American Journal of Hematology. Vol. 53. pp. 59-65. 1996.

44. Li, J., Noguchi, C.T., Miller, W., Hardison, R., Schechter, A.N. "Multiple regulatory elements in the 5'-flanking sequence of the human ε-globin gene." Journal of Biological Chemistry. Vol. 273, No. 17. pp. 10202-10209. 1998.

45. Liu, Q., Bungert, J., Engel, J.D. "Mutation of gene proximal regulatory elements disrupts human ε-, γ- and β-globin expression in yeast artificial chromosome transgenic mice." Proceedings of the National Acedemy of Sciences USA. Vol. 94. pp. 169-174. 1997.

46. Liu, Q., Tanimoto, K., Bungert, J., Engel, J.D. "The ^{A}γ-globin 3' element provides no unique function(s) for human β-globin locus gene regulation." Proceedings of the National Academy of Sciences USA. Vol. 95. pp. 9944-9949. 1998.

47. Lloyd, J.A., Case, S.S., Ponce, E., Lingrel, J.B. "Positive transcriptional regulation of the human γ-globin gene: γPE is a novel nuclear factor with multiple binding sites near the gene." Journal of Biological Chemistry. Vol. 269, No. 30. pp. 19385-19393. 1994.

48. Magli, M.C., Barba, P., Celetti, A., DeVita, G., Cillo, C., Boncinelli, E. "Coordinate regulation of HOX genes in human hematopoietic cells." Proceedings of the National Acedemy of Sciences USA. Vol. 88. pp. 6348-6352. 1991.

49. Magram, J., Chada, K., Costantini, F. "Developmental regulation of a cloned adult β-globin gene in transgenic mice." Nature. Vol. 315. pp. 338-340. 1985.

50. Matthews, C.H.E., Detmer, K., Boncinellli, E., Lawrence, H.J., Largman, C. "Erythroid restricted expression of homeobox genes of the human Hox-2 locus." Blood. Vol. 78, No. 9. pp. 2248-2252. 1991.

51. Merghoub, T., Perichon, B., Maier-Redelsperger, M., Dibenedetto, S.P., Samperi, P., Ducrocq, R., Feingold, N., Elion, J., Schiliro, G., Labie, D., Krishnamoorthy, R. "Dissection of the association status of two polymorphisms in the β-globin gene cluster

with variations in F-cell number in non-anamic individuals." American Journal of Hematology. Vol. 56. pp. 239-243. 1997.

52. Milot, E., Strouboulis, J., Trimborn, T., Wijgerde, M., de Boer, E., Langeveld, A., Tan-Un, K., Vergeer, W., Yannoutsos, N., Grosveld, F., Fraseer, P. "Heterochromatin effects on the frequency and duration of LCR-mediated gene transcription." Cell. Vol. 87. pp. 105-114. 1996.

53. Morley, B.J., Abbott, C.A., Wood, W.G. "Regulation of human fetal and adult globin genes in mouse erythroleukemia cells." Blood. Vol. 78, No. 5. pp. 1355-1363. 1991.

54. Navas, P.A., Peterson, K.R., Li, Q., McArthur, M., Stamatoyannopoulos, G. "The 5'HS4 core element of the human β-globin locus control region is required for high level globin gene expression in definitive but not in primitive erythropoiesis." Journal of Molecular Biology. Vol. 312. pp. 17-26. 2001.

55. Navas, P.A., Peterson, K.R., Li, Q., Skarpidi, E., Rohde, A., Shaw, S.E., Clegg, C.H., Asano, H., Stamatoyannopoulos, G. "Developmental specificity of the interaction between the locus control region and embryonic or fetal globin genes in transgenic mice with an HS3 core deletion." Molecular and Cellular Biology. Vol. 18, No. 7. pp. 4188-4196. 1998.

56. O'Neill, D., Yang, J., Erdjument-Bromage, H., Bornschlegel, K., Tempst, P., Bank, A. "Tissue specific and developmental stage specific DNA binding by a mammalian SWI/SNF complex associated witth human fetal to adult globin gene switching." Proceedings of the National Acedemy of Sciences USA. Vol. 96. pp. 349-354. 1999.

57. Perkins, A.C., Gaensler, K.M.L., Orkin, S.H. "Silencing of human fetal globin expression is impaired in the absence of the adult β-globin gene activator protein EKLF." Proceedings of the National Academy of Sciences USA. Vol. 93. pp. 12267-12271. 1996.

58. Peterson, K.R., Clegg, C.H., Huxley, C., Josephson, B.M., Haugen, H.S., Furukawa, T., Stamatoyannopoulos, G. "Transgenic mice containing a 248kb yeast artificial chromosome carrying the β-globin locus display proper developmental control of human globin genes." Proceedings of the National Academy of Sciences USA. Vol. 90. pp. 7593-7597. 1993.

59. Peterson, K.R., Clegg, C.H., Navas, P.A., Norton, E.J., Kimbrough, T.G., Stamatoyannopoulos, G. "Effect of deletion of 5'HS3 or 5'HS2 of the human β-globin locus control region on the developmental regulation of globin gene expression in β-globin locus yeast artificial chromosome transgenic mice." Proceedings of the National Academy of Sciences USA. Vol. 93. pp. 6605-6609. 1996.

60. Peterson, K.R., Li, Q., Clegg, C.H., Furukawa, T., Navas, P., Norton, E.J., Kimbrough, T.G., Stamatoyannopoulos, G. "Use of yeast artificial chromosomes (YACs) in studies of mammalian development: Production of β-globin locus YAC mice carrying human developmental mutants." Proceedings of the National Acedemy of Sciences USA. Vol. 92. pp. 5655-5659. 1995.

61. Plant, K.E., Routledge, S.J.E., Proodfoot, N.J. "Intergenic transcritpion in the human β-globin gene cluster." Molecular and Cellular Biology. Vol. 21, No. 19. pp. 6507-6514. 2001.

62. Porcu, S., Kitamura, M., Witkowska, E., Zhang, Z., Mutero, A., Lin, C., Chang, J., Gaensler, K.M.L. "The human β-globin locus introduced by YAC transfer exhibits a specific and reproducible pattern of developmental regulation in transgenic mice." Blood. Vol. 90, No. 11. pp. 4602-4609. 1997.

63. Pruzina, S., Hanscombe, O., Whyatt, D., Grosveld, F., Philipsen, S. "Hypersensitive site 4 of the human β-globin locus control region." Nucleic Acids Research. Vol. 19, No. 7. pp. 1413-1419. 1991.

64. A. Ragusa, personal communication.

65. Samakoglu, S., Philipsen, S., Grosveld, F., Luleci, G., Bagci, H. "Nucleotide changes in the γ-globin promoter and the (AT)xNy(AT)z polymorphic sequence of the βLCRHS-2 region associated with altered levels of HbF." European Journal of Human Genetics. Vol. 7. pp. 345-356. 1999.

66. Sambrook, Fritsch, and Maniatis. Molecular Cloning: a laboratory manual. Cold Spring Harbour Laboratory Press, USA. 2nd ed. 1989. (Chapters: 1.77-78, 1.25-28, 1.42-46, 1.80, 6, 9.52-5510.7-12, E.3, E.10-14)

67. Sargent, T.G., Lloyd, J.A. "The human γ-globin TATA and CACCC elements have key distinct roles in suppressing β-globin gene expression in embryonic/fetal development." The Journal of Biological Chemistry. Vol. 276, No. 45. pp. 41817-41824. 2001.

68. Schubeler, D., Groudine, M., Bender, M.A. "The murine β-globin locus control region regulates the rate of transcription but not the hyperacetylation of histones at the active genes." Proceedings of the National Acedemy of Sciences USA. Vol. 98, No. 20. pp. 11432-11437. 2001.

69. Sengupta, P.K., Lavelle, D.E., DeSimone, J. "The 87kD $^{A}\gamma$-globin enhancer binding protein is a product of the HOXB2 (HOX2H) locus." Blood. Vol. 83, No. 5. pp. 1420-1427. 1994.

70.A. Sikorski, R.S., Hieter, P. "Manipulation of DNA in *Saccharomyces cerevisiae*." Genetics. Vol. 122. pp. 19-27. 1989.

71. Stamatoyannopoulos, G., Josephson, B., Zhang, J.W., Li, Q. "Developmental control of the human gamma-globin genes in transgenic mice." Molecular and Cellular Biology. Vol. 13, No. 12. pp. 7636-7644. 1993.

72. Stratagene web site: www.stratagene.com

73. Strouboulis, J., Dillon, N., Grosveld, F. "Developmental regulation of a complete 70-kb human β-globin locus in transgenic mice." Genes and Development. Vol. 6. pp. 1857-1864. 1992.

74. Tanimoto, K., Liu, Q., Bungert, J., Engel, J.D. "Effects of altered gene order or orientation of the locus control region on human β-globin gene expression in mice." Nature. Vol. 398. pp. 344-348. 1999.

75. Tanimoto, K., Liu, Q., Grosveld, F., Bungert, J., Engel, J.D. "Context dependent EKLF responsiveness defines the the developmental specificity of the human ε-globin gene in erythroid cells of YAC transgenic mice." Genes and Development. Vol. 14. pp. 2778-2794. 2000.

76. Torigoi, E., Bennani-Baiti, I.M., Rosen, C., Gonzalez, K., Morcillo, P., Ptashne, M., Dorsett, D. "Chip interacts with diverse homeodomain proteins and potentiates Bicoid activity in vivo." Proceedings of the National Acedemy of Sciences USA. Vol. 97, No. 6. pp. 2686-2691. 2000.

77. Tsang, A.P., Visvader, J.E., Turner, C.A., Fujiwara, Y., Yu, C., Weiss, M.J., Crossley, M., Orkin, S.H. "FOG, a multiple zinc finger protein, acts as a cofactor for transcription factor GATA-1 in erythroid and megakaryocytic differentiation." Cell. Vol. 90. pp. 109-119. 1997.

78. Tuan, D., Abeliovitch, A., Lee-Oldham, M., Lee, D. "Identification of regulatory elements of human β-like globin genes." Progress in Clinical and Biological Research. Vol. 235. pp. 211-220. 1987.

79. Tuan, D., Kong, S., Hu, K. "Transcription of the hypersensitive site HS2 enhancer in erythroid cells." Proceedings of the National Academy of Sciences USA. Vol. 89. pp. 11219-11223. 1992.

80. Tuan, D., London, I.M. "Mapping of DNase I-hypersensitive sites in the upstream DNA of the human embryonic ε-globin gene in K562 leukemia cells." Proceedings of the National Acedemy of Sciences USA. Vol. 81. pp. 2718-2722. 1984.

81. Tuan, D., Solomon, W., Li, Q., London, I.M. "The "β-like-globin" gene domain in human erythroid cells." Proceedings of the National Acedemy of Sciences USA. Vol. 82. pp. 6384-6388. 1985.

82. Vieille-Grosjean, I., Huber, P. "Transcription factor GATA-1 regulates human HOXB2 gene expression in erythroid cells." Journal of Biological Chemistry. Vol. 270, No. 9. pp. 4544-4550. 1995.

83. Whitelaw, E., Tsai, S-F., Hogben, P., Orkin, S.H. "Regulated expression of globin chains and the erythroid transcription factor GATA-1 during erythropoiesis in the developing mouse." Molecular and Cellular Biology. Vol. 10, No. 12. pp. 6596-6606. 1990.

84. Yoo, J., Herman, L.E., Li, C., Krantz, S.B., Tuan, D. "Dynamic changes in the locus control region of erythroid progenitor cells demonstrated by polymerase chain reaction." Blood. Vol. 87, No. 6. pp. 2558-2567. 1996.