Chromatin remodelling and the interaction between enhancers and promoters in the β-globin locus

Ann Dean

Date received (in revised form): 20th October 2003

Abstract
The β-globin genes are among the most studied families of tissue-specific and developmentally regulated genes. Even before genome-wide sequencing was a reality, the sequences of extensive regions of the human, rabbit and mouse β-globin loci were available through the individual efforts of several research groups. Many of the genomic approaches now being pursued in order to understand the control of gene expression were applied early to this gene family with success. Higher levels of regulation attributable to networks of transcription factor interaction, chromatin remodelling and enhancer–promoter communication remain as challenges in the post-genomic era.

INTRODUCTION
Transcription activation by distant enhancers is a characteristic of many developmentally regulated gene families of which the β-globin genes are a prominent example. There are five human β-like-globin genes clustered on chromosome 11, which are differentially expressed during development in the order of their 5′ to 3′ arrangement: ε-, γy-, δγ-, δ- and β-globin (Figure 1). The ε-globin gene is expressed in the early stages of embryonic development, up to about week 10 of gestation. The two γ-globin genes are then expressed during the foetal period of development. Finally, the δ- and β-globin genes become predominant after birth and into adulthood.1

The overall structure of this locus is closely mirrored in the mouse β-globin cluster, where the εγ- and βh1-globin genes are expressed during early development and the adult β-globin genes predominate after birth. Comparison of the structure of the human globin locus with globin loci from other eutherian mammals led to the proposal that the locus arose by gene duplication during evolution.2,3 An alternative grouping of the globin genes is exemplified by the chicken globin locus. Here, the two embryonic genes, σ- and ε-globin, are at the ends of the locus and the two adult β-globin genes are in the middle. The two types of clusters were long thought to be orthologous; however, the discovery of both types of loci in a marsupial argues that the eutherian and bird clusters may have arisen from different ancestors.4,5

The approximately 100 kilobase (kb) human globin locus exhibits, in erythroid cells, a so-called ‘general sensitivity’ to DNase I digestion.6 In addition, local alterations in the promoter chromatin structure of highly expressed genes are observed contemporaneously as each gene is activated at the appropriate developmental stage. The high-level transcriptional activation of the globin genes is dependent on the β-globin locus control region (LCR).7,8 The LCR comprises four conserved regions of several hundred nucleotides each (HS1–HS4) that form DNase I hypersensitive sites (HSs) in erythroid cell chromatin. These are located between 6 and 60 kb upstream of the genes they regulate.9 One
of the most conceptually difficult problems in globin gene regulation has been to establish how the LCR acts at linearly distant promoters. Recent data discussed in this paper illuminate this issue, but much remains unclear.

DECONSTRUCTION OF GLOBIN GENE REGULATION

Intense interest in globin gene regulation has been fuelled by the desire to learn enough about the control of the expression of these genes to be able to diagnose precisely and modulate therapeutically human clinical β-globin gene disorders such as β-thalassaemia and sickle cell disease. Hence, beginning in the late 1970s and into the early 1980s, the globin genes were among the first human genes to be molecularly cloned and sequenced. Quite in advance of current genome-wide sequencing efforts, therefore, extensive regions of the human, mouse, rabbit and chicken β-globin loci were sequenced and available for analysis.10

Analyses revealed that the coding sequences and exon/intron arrangement of the β-globin genes, both within and among species, were highly conserved. Some of the earliest clues to regulators of the globin genes were obtained by sequence alignment, and by intra- and interspecies comparative analysis of non-coding regions of mammalian β-globins. ATA (TATA box) and CCAAT motifs were observed to be common to all the globin promoters. Soon after, scanning mutagenesis and functional testing of the rabbit β-globin promoter in vitro revealed the regulatory importance of these two elements — as well as the CACCC element — to transcription of the globin genes.11 These important sites of interaction, while relevant to gene expression in general, did not illuminate erythroid-specific gene regulation.

Subsequently, erythroid-specific factors were cloned from chicken and mammalian DNAs based either on their binding to functionally important regulatory elements (GATA-112 and NF-E213) or on a directed search for CACCC-binding factors (EKLF14). Comparative sequence analysis showed that all globin promoters and/or LCR HSs shared motifs for these factors. Biochemical studies revealed in vitro interactions among these factors, as well as between these factors and the adaptor molecule TAFII130 and the histone acetyl transferase (HAT) CBP/p300 (see Table 1). These observations led to the proposal that homo- and heteromeric interactions among these transcription factors could function in integrating enhancer–promoter communication and transcription activation.1,15,16

At the same time, efforts to uncover novel and important conserved elements in globin gene regulation were pursued...
through phylogenetic footprinting. This approach depends on detecting motifs shared among orthologous globin genes, sequences for many of which became available in the 1990s. Human–mouse sequence alignment and analysis by percent identity plot (PIP) indicated additional regions of identity outside the LCR, HS cores, and a conserved E box in the HS2 core, which binds upstream transcription factor (USF) and Tal1 and plays a role in enhancement. More recently, it was discovered that 5’ HS4 in the chicken locus, a boundary element insulating the locus from adjacent heterochromatin, is a site of interaction of the insulator factor CTCF. Although CTCF sites have also been localised 3’ of the chicken locus and both 5’ and 3’ of the mouse and human loci, these regions vary in insulator strength in enhancer blocking assays, and their function in vivo is unclear. Many of the phylogenetic footprints in globin loci remain to be functionally characterised, partially due to the general difficulty in designing appropriate tests.

Some progress has been made in identifying a second level of protein–protein interactions or networks of globin gene regulators. For example, FOG 1 (friend of GATA-1), a functionally important partner of GATA-1, was discovered in a yeast two-hybrid screen. Analysis by immunoprecipitation revealed that GATA-1, Tal1, LMO2, E47 and Ldb1 exist in a large, erythroid-specific complex. How this complex functions remains unclear. Other researchers have shown that EKLF provides specificity to globin promoter remodelling in vitro, in that it is required to recruit SWI/SNF remodelling complex components to the human β-globin promoter. A current major challenge in understanding how the human β-globin locus is regulated is both to deepen and broaden our knowledge of the factors at work integrating the locus with the complex processes of chromatin remodelling and transcription activation.

### CHROMATIN REMODELLING AND TRANSCRIPTION ACTIVATION

Studies in recent years have made it clear that chromatin structure plays a critical role in gene expression, adding an additional dimension to sequence-based genomics. In vivo, genes are complexed with histone proteins to form chromatin. The nucleosome, the basic unit of chromatin, is, in general, inhibitory to transcription. Transcriptional activation involves relief of this repression in two ways. ATP-dependent nucleosome remodelling complexes of the SWI/SNF type use energy from ATP hydrolysis to alter nucleosome structure and/or position along the chromatin fibre, while other complexes covalently modify the N-terminal tails of histones within nucleosomes. Several types of histone modifications have been described, including acetylation by SAGA-/GCN5-like HAT complexes, methylation, phosphorylation and ubiquitylation.

Evidence suggests that the actions of the two types of chromatin remodelling complexes are interdependent. Interestingly, the order in which these

#### Table 1: Erythroid transcription factors

<table>
<thead>
<tr>
<th>PROTEIN</th>
<th>MOTIF</th>
<th>CLASS</th>
<th>INTERACTS WITH</th>
</tr>
</thead>
<tbody>
<tr>
<td>EKLF</td>
<td>CACCC</td>
<td>Kruppel-like</td>
<td>GATA-1, CBP/p300,</td>
</tr>
<tr>
<td>GATA-1</td>
<td>(A/T)GATA(A/G)</td>
<td>GATA-like</td>
<td>TAF130, GATA-1,</td>
</tr>
<tr>
<td>NF-E2</td>
<td>TGCTGA/(C/G)TCA</td>
<td>C4 zinc fingers</td>
<td>Sp1, GATA-1,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>basic leucine zipper</td>
<td>FOG, EKLF, CBP/p300</td>
</tr>
</tbody>
</table>

---

A cadre of erythroid and non-erythroid transcription factors participate in globin gene regulation.
complexes act at promoters is not invariant. Recent studies of several yeast and mammalian genes revealed that, in some cases, promoter hyperacetylation preceded SWI/SNF remodelling, while at other promoters the reverse is true. These studies suggest the possibility that different promoters have evolved independent mechanisms with which to overcome barriers to transcription. In a mini-chromosome model system in vivo, human HS2-mediated ß-globin promoter remodelling is not dependent on prior promoter hyperacetylation, suggesting that remodelling may occur first at this promoter. In the yeast recruitment model, sequence-specific DNA-binding factors that associate with promoters or nearby upstream activating sequences (UASs) recruit the protein complexes that remodel the adjacent promoter chromatin (Figure 2). Yeast UASs are highly distance dependent, whereas vertebrate LCR/ enhancers are typically separated by tens of kilobases from the genes they activate. Nevertheless, LCRs act in a similar way to relieve nucleosome repression at their distant targets, resulting in DNase I HS site formation and histone hyperacetylation at promoter sequences that correspond with transcription activation. At present, the ways in which the recruitment model can be extended to gene activation from a distance remain unclear, but understanding how enhancers and promoters communicate is clearly essential to such an undertaking.

**ENHANCER–PROMOTER COMMUNICATION**

A number of models of enhancer-/LCR-dependent gene activation are available (Figure 3). One possibility is that activators and remodelers recruited to the enhancer migrate by tracking along the chromatin between the enhancer and promoter. Alternatively, a distant enhancer and promoter might contact each other directly or indirectly, with separation of the intervening DNA, as described in looping and linking models. Facilitated tracking is an intermediate view in which the activators track along the DNA from enhancer to promoter without losing contact with the enhancer, eventually forming a loop. It has been difficult to test these models directly, although recent technical advances have provided the first definitive results invoking one or the other model. Interestingly, one unifying model may not suffice (see below).

The models form a useful framework within which to investigate

---

**Figure 2:** The yeast recruitment model of gene activation is depicted. SWI/SNF and HAT complexes are recruited to the gene promoter via activators bound to the nearby upstream activating sequences (UASs). These complexes alter the promoter to permit binding of the transcriptional machinery and the commencement of transcription.

**Figure 3:** Looping and tracking models of enhancer–promoter communication are illustrated. Y indicates a target gene; X indicates activator complexes targeted to the locus control region (LCR).
enhancer—promoter communication. Recent results in systems where genes are activated by relatively simple distant enhancers have advanced the mechanistic understanding of how enhancers mediate chromatin alterations and transcription activation at distant genes. These advances have depended heavily on chromatin immunoprecipitation (ChIP) assays and polymerase chain reaction amplification to detect in vivo bound activators. For example, the prostate-specific antigen (PSA) gene and its enhancer are separated by 4.2 kb. ChIP experiments showed PSA activators, RNA polymerase II (pol II) and histone acetylation at both the enhancer and promoter of the gene, but not at the intervening sequences. In the TCRβ locus, the HAT CBP and pol II were similarly found to be associated with the promoter and distant enhancer. These results support a direct interaction between enhancer and promoter mediated by activator molecules that bind at both sites.

In contrast to these results, others have found that pol II is both recruited to the PSA gene enhancer independent of recruitment to the promoter and is detected (along with histone acetylation) at significant levels at sequences between enhancer and promoter, a result which is consistent with a tracking model. Finally, investigators have demonstrated the association of activators — HATs and the SWI/SNF component Brg1 — between an enhancer and the HNF-4α promoter, which occurs without loss of contact activator with the enhancer. This last result provides an example of a facilitated tracking mechanism.

The β-globin LCR illustrates a more complex enhancer organisation than the above examples as it contains at least four separate regions which bind clusters of erythroid transcription factors. A great deal of experimentation, from transient transfections to transgenic studies, has been devoted to understanding the independent or cooperative nature of the LCR component HSs in the globin locus. The prevailing model proposes that the globin LCR HS sites interact with one another to form a holocomplex, which acts in turn on each of the globin genes at the appropriate developmental stage. A considerable technical advance, called 3C (‘capturing chromosome conformation’), has been used to show direct contact between individual HSs in the endogenous mouse globin LCR. Other complementary data obtained with a model globin chromatin system indicate that HS2 and HS3 interact structurally.

Although the mechanism underlying the establishment and maintenance of contacts between the LCR HSs themselves remains unclear, these results provide strong support for the holocomplex concept. Nevertheless, a complete understanding of how the LCR HSs, probably mutually, actually affect each of the distant genes is still someway off.

Recent data from studies of the endogenous mouse globin locus using a combination of fluorescence in situ hybridisation and ChIP contribute significantly to this debate. These data show that LCR HSs and an active globin promoter are held in close physical proximity when transcription is active. Moreover, 3C experiments reveal that globin genes switch this close interaction with the LCR in a way that correlates with the switch in their transcriptional activity. The basis for these associations remains to be explored. Earlier studies in vivo on a model chromatinised globin locus, in which factor motifs could be individually ablated, had indicated that the erythroid factors, interacting through shared motifs in the LCR HSs and the globin promoters, could mediate physical contact between these regulatory elements. Subsequent results using this model system suggest that direct contact between the HS2 enhancer and the human ε-globin promoter is at least partially responsible for the mutual association of factors with these motifs. By contrast, other data accumulated in studies of the endogenous mouse globin locus argue for remote delivery of
transcription activators, including RNA polII, from the LCR to the β-globin promoter, or LCR-independent factor and polII binding to the promoter.

Further work will be required to understand the extent to which enhancer–promoter contact is critical to promoter remodelling and transcription activation dependent on the β-globin LCR. Moreover, the chromosomal and nuclear milieu is likely to provide additional levels of regulatory control to globin loci. Sites of attachment to the nuclear matrix have been described for the human β-globin locus, which could serve to maintain the locus, or parts of it, independent of other chromatin.

Others have described a role for the LCR in maintaining the globin locus away from centromeric heterochromatin. The integration of these observations into an overall picture of globin gene regulation awaits future developments.

MAKING NEW CONNECTIONS IN THE GLOBIN LOCUS

There have been extensive studies aimed at linking chromatin alterations, transcription activation and the β-globin LCR, and interesting counterpoints have arisen from the study of the bird and mammalian loci. The chicken locus has been the most extensively characterised. In this locus, the extent of general DNase I sensitivity and domain–wide hyperacetylation of histones correspond precisely, and the ends of the domain are marked by chromatin insulators.

Moreover, in 10-day-old red cells, both the actively transcribed adult β-globin genes, as well as the inactive γ- and ε-globin genes, exist within the hyperacetylated and DNase I-sensitive domain.

By contrast, the organisation of the mouse locus is considerably less tidy. The locus exhibits general DNase I sensitivity; however, regions of putative boundary activity do not correspond to the limits of locus–wide general sensitivity to DNase I, and domains of both hyperacetylation and hypoacetylation exist as sub-domains within the DNase I-sensitive locus.

Inactive genes, as exemplified by the ε-globin gene during adult erythropoiesis, exist in a hypoacetylated domain. In addition to domains of modification, peaks of histone acetylation characterise the mouse LCR HSs and the active β-globin gene promoters, although such peaks were not notable in the chicken locus.

Surprisingly, a series of studies in which the LCR was deleted from the endogenous mouse chromosome revealed that, although transcription of the β-globin gene was reduced to 1–4 per cent of the endogenous level, general DNase I sensitivity of the locus and β-globin promoter remodelling and hyperacetylation were unaffected.

Studies of the human locus provide an alternative picture to the mouse locus. These studies have relied on diverse experimental systems. One series of experiments utilised a hybrid murine erythroleukaemia cell line (MEL) carrying human chromosome 11, where the β-globin locus is situated. The adult β-globin gene is appropriately expressed from the human chromosome in this adult MEL environment; however, by contrast with the mouse LCR deletion, deletion of the LCR from the human chromosome resulted in complete loss of transcription and promoter remodelling, consistent with the idea that remodelling complexes and acetylases are recruited to the LCR and then modify promoters. Taken together, these observations suggest the possibility that: (1) the mouse and human loci are regulated differently in at least one major respect, in that sequences outside the LCR contribute to promoter remodelling and hyperacetylation in the mouse but not the human locus; or (2) evolution has altered human transcription activators or co-activators to the extent that the mouse homologues no longer function as perfect substitutes when the human locus is studied in a murine cellular environment.
As with the mouse LCR deletion, general DNase I sensitivity of the human locus was unaffected by the LCR deletion, although it has long been known that deletion of an additional 27 kb of upstream sequence, as occurs on a severe \(\beta\)-thalassaemic chromosome, results in loss of general sensitivity in the locus. Thus, it seems likely that sequences — possibly as yet unidentified clusters of erythroid factor motifs outside the confines of the LCR as traditionally defined — contribute to locus-wide alterations in chromatin structure in the endogenous locus.

The HS2 enhancer component of the human LCR and the \(\alpha\)-globin gene have been studied in a model human globin gene locus on mini-chromosomes in human erythroid cells. In this system, the HS2 enhancer alone mediated widespread histone acetylation between the enhancer and the target globin gene, indicating that LCR components have the capacity to recruit histone acetylase activity. Furthermore, local modifications at the promoter of the linked gene (acetylation, DNase I HS formation) associated with transcription activation were completely dependent on HS2, consistent with the idea that remodelling complexes and acetylases are recruited to the LCR and then modify promoters.

Interestingly, in this model human globin gene locus, as well as in the endogenous human globin locus in K562 cells, the HS2 enhancer and the promoter of the human \(\epsilon\)-globin gene were not seen as peaks in histone modification of the locus. The core of HS2 was essentially unmodified by histone acetylation and the gene promoter was no more highly enriched in acetylated histones than were coding sequences or 5’ promoter flanking sequences. This high resolution analysis revealed a significant depletion of HS2 core sequences from the input DNA. Although DNase I HSs have been assumed to be nucleosome free, these experiments are the first to distinguish between eviction of a nucleosome at HS2 and the presence of a remodelled nucleosome at the active \(\epsilon\)-globin gene promoter.

**PERSPECTIVE**

The existence of an increasing number of complete genome sequences is likely to offer new insights into globin gene regulation through sequence alignment and interspecies comparison. For example, the discovery of an orphan globin gene in marsupials raises the possibility that globin genes unlinked to the major \(\alpha\)- and \(\beta\)-globin loci may be identified in humans and birds. Furthermore, it was noted that the homology between chicken and human globin LCR regulatory regions is not high, even though they share transcription activator motifs. Combining the knowledge of genome sequences from species intermediate between birds and humans and current methodological approaches may increase the likelihood of discovering more regulatory sequences and interacting partners.

As described above, the regions of the globin locus responsible for chromatin opening have not been identified and presumably lie outside the confines of the LCR as classically defined. Such a function is likely to involve multiple transcription activators and may be amenable to computational and experimental approaches in order to search for clusters of transcription factor sites. This approach may also identify previously unknown \(\alpha\)-regulatory sequences in the region at \(-62.5\) kb to the mouse globin structural genes and the corresponding human \(-111\) kb region, or elsewhere. Phylogenetic footprints could also be further analysed using these algorithms, while ChIP experiments offer experimental confirmation of actual occupancy of potential regulatory sites.

A theoretically important approach is location analysis by ChIP followed by microarray analysis (so-called ChIP-chip). This approach may also identify previously unknown \(\alpha\)-regulatory sequences in the region at \(-62.5\) kb to the mouse globin structural genes and the corresponding human \(-111\) kb region, or elsewhere. Phylogenetic footprints could also be further analysed using these algorithms, while ChIP experiments offer experimental confirmation of actual occupancy of potential regulatory sites.

A theoretically important approach is location analysis by ChIP followed by microarray analysis (so-called ChIP-chip). One caveat to the approach is the possibility that a factor will be found that is widely associated with sequences in a domain and not just at functional sites.
Certainly, all bound sites need not be functional. Homeo-proteins bind to long stretches of sequence between enhancers as well as to their known response elements.69 This type of analysis has been carried out using antibodies to GATA-1 for the ChIP and globin locus sequences for the chip.70 In contrast to an overabundance of interactions, Horak et al. identified one previously known site of GATA-1 interaction within HS2 and one other site in an intergenic region; this is almost certainly an underestimate of the GATA-1 sites of functional interaction. These experiments are technically challenging, but the approach has great promise.

**CHALLENGES IN THE POST-GENOMIC ERA**

Several aspects of chromatin remodelling and enhancer-dependent transcription activation challenge the limits of current investigative tools.71 First, these processes require large complexes. For example, the 2 megadalton SWI/SNF remodelling complex in yeast is composed of 11 subunits. The basal transcription machinery is likewise made up of multiple components. Secondly, although many subunits of these complexes may participate in activation of diverse genes, some components will be required to confer specificity of interaction between remodelling complexes and the transcription machinery. These gene-specific adaptor proteins are not likely to be DNA-binding factors, precluding their identification by this experimentally useful characteristic which has been the focus of most functional genomic studies.

A future challenge is to use interaction proteomics to unravel these protein–protein interactions. Recent advances in several proteomics technologies have been adapted to investigate protein–protein interactions on a genome-wide scale.72 Although primarily successful to date in yeast, and to some extent in *Caenorhabditis elegans*, it is not unreasonable to suppose that these approaches could be applied with success to particular cellular compartments, such as the erythroid cells of mammals. For example, in a recent study, two-dimensional gel electrophoresis and high throughput mass spectrometric analysis were used to identify multiprotein complexes during myeloid stem cell development.73

The power of combining genomic sequencing with a post-genomic functional assay is illustrated by work in which novel enhancers of the SCL gene were discovered or in which a genome-wide search for boundary elements in yeast was carried out.74,75 It is only through innovative approaches such as these that heteromeric interacting partners of transacting factors and the means by which to study how these factors create networks to link activation signals with chromatin remodellers and the transcription machinery can be found. Understanding the integrated process of transcription activation will require novel approaches, if the wealth of genomic data is to be mined to this end.

**Acknowledgment**

I would like to thank members of my laboratory for many helpful discussions on these issues, and Jane Little and Hui Zhao for critical comments on the manuscript.

**References**


Chromatin remodelling and the interactions in the β-globin locus


