Cell, Vol. 100, 499–502, March 3, 2000, Copyright ©2000 by Cell Press

Looping, Linking, and Chromatin Activity: New Insights into β-*globin* Locus Regulation

Minireview

James Douglas Engel* and Keiji Tanimoto
Department of Biochemistry, Molecular Biology, and Cell Biology
Northwestern University
Evanston, Illinois 60208-3500

Analysis of globin gene transcription has provided us with one of the most thoroughly studied examples of tissue- and temporally specific transcriptional control. Many of the concepts and principles of eukaryotic gene regulation are a direct consequence of the analysis of globin loci. Yet, even after decades of molecular analysis and despite the availability of dozens of naturally occurring mutations correlated with specific disease phenotypes, the mechanisms underlying the simple yet exquisitely reproducible pattern of sequential embryonic, fetal, and adult β-type globin gene transcription remain sharply controversial and surprisingly elusive. The basic facts underlying major controversies in this field are largely undisputed, so we refer interested readers to two reviews for a meaningful introduction (Bell and Felsenfeld, 1999; Bulger and Groudine, 1999). Two new papers examining different aspects of chromatin structure in the β-*globin* locus revealed, on the one hand, a surprising lack of influence of the locus control region (LCR) on overall chromatin sensitivity, and on the other, a subtlety in the structure of active chromatin that correlates with nonglobin transcription within the locus.

Many of these controversies focus on the in vivo function(s) of the LCR, a group of five DNase I hypersensitive sites (HS) located from 6 to about 25 kbp 5′ to the embryonic ϵ-*globin* gene (Tuan et al., 1985; Forrester et al., 1986; Figure 1). As but one example of temporal complications that can arise, even our definition of what constitutes the LCR has changed only recently, since it was discovered that additional HSs are located even further 5′ to the five original HSs (Figure 1). Individual globin genes, or even the whole cluster of genes in their natural configuration, are characteristically expressed at the proper developmental time in transgenic animals but at greatly reduced abundance in comparison to endogenous globin genes. It was discovered that when *cis*-linked to globin genes, the LCR conferred roughly physiological expression levels in transgenic animals (Grosveld et al., 1987). These seminal observations suggested that earlier globin transgenes failed to express at high levels simply because they lacked the LCR.

Experimental Systems

We are at a stage in the evolution of this field when much of our current knowledge and direction is based on observations made in transgenic mice bearing what many investigators believe represents the entire human β-*globin* gene locus. More recently, studies of germline mutants in the endogenous murine locus have provided additional insight either in support of or in contrast to transgenic investigations. Thus, one level of potential problems arises because many of us evaluate the data from these two sources as equivalent, even though they represent a comparison between (randomly integrated human and murine) transgenes and the natural (murine) locus. Both mice and humans have similar gene organizations relative to the LCR (Figure 1) and, thus, if pressed, most investigators would probably agree that the loci are likely to be regulated by qualitatively similar mechanisms, since highly sequence-conserved LCRs regulate the β-*globin* genes in every organism analyzed to date (Hardison et al., 1998).

Although there are no specific examples, a second potential pitfall could arise as a consequence of using different (cosmid, BAC, or YAC) base DNAs as starting materials in the transgenic studies. Small constructs incorporating different parts of the locus are the simplest transgenes to construct and analyze, but more recent work has focused on mutations in large linked cosmids and even more recently on very large YACs bearing the locus (Figure 1). At present, we have no way of knowing whether all these fragments of DNA (or indeed, if any of them) confer to a transgene an environment equivalent to that encountered by the endogenous genes.

A third complication is that various laboratories have (quite naturally) examined different mutations that lead to some of the controversy between independent studies. For example, it was shown using a single YAC that a 2.3 kb deletion encompassing LCR HS3 has far *less* deleterious consequences on globin gene expression than an analogous deletion of 234 bp defined as the "core" of HS3 (Bungert et al., 1995; Peterson et al., 1996; Navas et al., 1998). With this revelation, there should be increased appreciation that any such comparisons between "similar" mutations is like comparing apples to oranges.

Questions

In spite of experimental and conceptual complications each approach presents, the central questions in β-*globin* locus gene regulation are fundamental. First, what regulatory properties does the LCR actually confer to β-locus gene regulation, and do these properties functionally effect transcription of the genes? Second, is differential replication timing through the locus in nonerythroid versus erythroid cells a determinant of gene expression, and if so, how? (Neither of the studies reviewed here address this specific issue, so it will not be further explored). Finally, is the differential chromatin structure detected in erythroid versus nonerythroid cells a cause or a consequence of globin gene activation, and if causal, how does it contribute to differential gene regulation? It is this final issue that was the focus of two recent publications in *Molecular Cell* (Bender et al., 2000; Gribnau et al., 2000).

One of the original properties attributed to the LCR was a chromatin opening function, a hypothesis consistent with transgenic data as well as with studies of a natural Hispanic LCR deletion ($\gamma\delta\beta^0$-thalassemia; Figure 1). The latter work demonstrated that the globin genes were not transcribed and that the chromatin surrounding them was inactive in the Hispanic deletion locus in both

*E-mail: d-engel@nwu.edu.

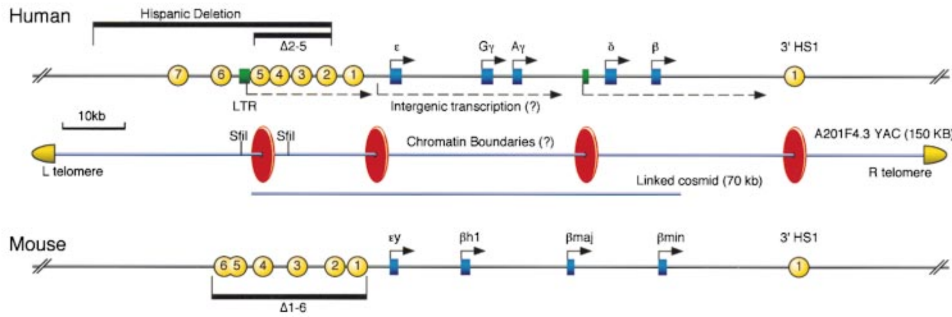

Figure 1. Proportional Representation of the Human (Upper) and Mouse (Lower) β-*globin* Loci

Globin genes are represented by blue boxes, putative intergenic transcription start sites by green boxes, and DNase I hypersensitive sites by numbered yellow spheres. Two lines beneath the human sequences represent the extent of the human genomic locus borne on either a 150 kb YAC or the 70 kb cosmid clones used for transgenic experiments (text). Hypothetical boundary elements (red disks) and endogenous SfiI restriction enzyme sites are shown on the YAC. Brackets indicate the extent of several (natural or experimental) deletions within the loci.

nonerythroid and erythroid cells, while transfer of the wild-type locus from nonerythroid to an erythroid environment was accompanied by both gene activation and increased DNase I sensitivity (Forrester et al., 1990; Figure 1). Consequently, it had become virtually axiomatic that the transcriptional stimulatory component of the LCR might be difficult to disentangle from its chromatin opening activity. The Bender et al. (2000) and Gribnau et al. (2000) reports highlighted here address qualities of differential chromatin structure in the locus that demand reevaluation of these perceptions.

Observations

Bender et al. (2000) presented in vivo evidence that despite severely diminished transcriptional activity in the locus after deletion of the LCR (as anticipated from earlier studies, including their own; Epner et al., 1998), the mutation had a negligible effect on the general DNase I sensitivity throughout the murine β-*globin* locus. This conclusion was congruent with an earlier study in which homologous recombination-mediated deletion of HS sites 2–5 of the human LCR abolished transcription of the genes without altering general chromatin sensitivity (Reik et al., 1998; ΔHS2-5, Figure 1), in stark contrast to studies on the Hispanic deletion (above). The important conclusion from these studies is that a second domain of chromatin opening activity must lie outside the LCR. This second domain must be in addition to that established by the LCR itself (since the linked cosmids used in many transgenic experiments can open chromatin, but only minimally encompass HS1-5), and thus deductively a second chromatin opening domain must lie within the 5′ most part of the Hispanic deletion (Figure 1).

Extra- and intergenic transcripts within the β-*globin* locus have been documented for some time, but a report 3 years ago revealed new insights into how pol II–dependent unidirectional intergenic transcripts might be generated (Ashe et al., 1997). Now, Gribnau et al. (2000) describe analysis of the chromatin structure of the transgenic human locus, as well as further properties of the globin intergenic transcripts. Importantly, they found that the intergenic transcripts surrounding the ε- and γ-*globin* genes were significantly more abundant in primitive erythroid cells (which transcribe only ε- and γ-*globin*) than those from the β-*globin* gene region, while the opposite was true in definitive erythroid cells. Furthermore, intergenic transcription correlated with modest changes in the regional DNase I chromatin sensitivity in the locus: an εγ "domain" was reportedly 2-fold more sensitive than a δβ domain in primitive erythroid cells, while the reverse was again true in definitive cells. Finally, they reported that an element lying between the two domains contained a discrete transcription initiation site 5′ to the δ-*globin* gene. Deletion of that element from a transgenic YAC had been shown to diminish β-*globin* gene transcription in definitive erythroid cells, and Gribnau et al. (2000) found that this mutation similarly inhibited δβ-domain intergenic transcription. The important conclusion from this report is that chromatin sensitivity of large regions flanking actively transcribed globin genes may be greater than that of domains that are (momentarily) transcriptionally inert and that this DNase I sensitivity correlates with differences in the abundance of the domain-restricted intergenic transcripts.

Future Questions

The Bender and Gribnau papers are of pivotal interest because they contribute new insights into one of the most basic precepts underlying globin transcriptional activity. First, Bender et al. (2000) demonstrate that at least one chromatin opening activity is separable from the transcriptional activation function of the β-*globin* LCR. Several questions emerge from this provocative discovery. Is the LCR alone capable of the same function (thus identifying this as a redundant property within the locus), or do HS1-5 cooperatively open chromatin through *cis* linkage with the genes? Is the new LCR-independent chromatin opening activity found in the region described by the difference in the ΔHS2-5 deletion and the $\gamma\delta\beta^0$-thalassemia Hispanic deletion (Figure 1)? What are the molecular identities of the chromatin modulatory activities that generate independent chromatin opening?

Gribnau et al. (2000) found that the locus appears to be subdivided into functionally discrete subdomains that reflect the developmental stage at which the genes within those domains are transcriptionally active, revealing a host of new possible levels of regulation. Do the intergenic transcripts arise as a cause or a consequence of chromatin opening or gene transcription? Can other (than the δβ domain) nongenic promoters be identified

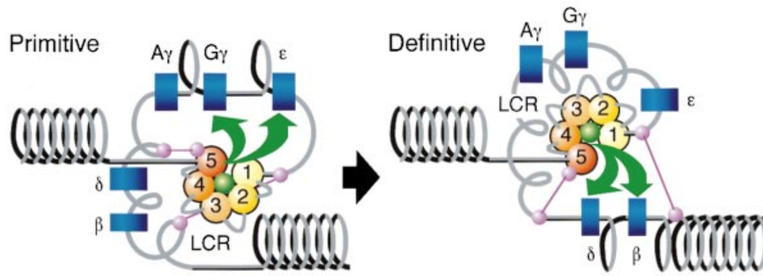

Figure 2. Stage-Specific Activation within Subdomains of the β-*globin* Locus by an LCR Holocomplex

Chromatinized DNA status (either relatively open or closed, which is meant to reflect general sensitivity to DNase I) is depicted as spring shapes with the more condensed chromatin represented as tighter coiling. Individual LCR HS sites and genes are shown as spheres (shaded from yellow to red) and cylinders (blue), respectively. Purple lines show possible structural constraints generated by linkage of two putative boundary elements (purple spheres) and how this might contribute to the formation of chromosomal subdomains. The architectural constraints within the locus and/or the LCR holocomplex configuration may also influence which gene(s) can be activated by LCR at specific developmental stages (green arrows representing the domains that are rendered more accessible after differential orientation of the LCR in the primitive and definitive erythroid stages).

in the locus? Do these intergenic RNAs represent long, uninterrupted primary transcripts? Is a small deletion, encompassing only the definitive δβ domain promoter, equally disruptive to adult β-*globin* transcription as the 2.5 kbp deletion? What is the mechanistic link, if any, between intergenic and globin gene transcription?

Mechanistic Interpretations

Both reports highlighted here challenge our perception of how vertebrate β-*globin* loci are regulated, although neither observation is inconsistent with currently popular models for how the LCR stimulates globin gene transcription. Two current models have been abbreviated "looping" or "linking," and the unstated, underlying characteristics attributed to both are specific and diverse (if not conflicting). The looping model is predicated on the idea that the LCR acts as an integral unit (a "holocomplex"; Bungert et al., 1995; Wijgerde et al., 1995) to stimulate the transcription of individual globin genes by looping through nucleoplasm to activate (or recruit) the transcriptional apparatus assembled at globin gene promoters. Since the LCR holocomplex would be limiting (one activation center), only one gene at a time could be transcribed, and thus in an equivalent environment the five genes of the human locus would compete for the activity of the LCR. While there are caveats and corollaries accompanying these basic precepts (e.g., to promote differential stage-specific transcription), the basic looping model cannot be refuted by current clinical or laboratory observations.

Even while numerous experiments *are consistent with* looping activation of, and competition between, individual genes and the LCR, none of them prove fundamental aspects of the hypothesis. In the absence of definitive experimental validation, further consideration of the profusion of laboratory and clinical observations led to an alternative model that might have fewer constraints, could explain the clinical and laboratory observations, and would be consistent with the properties of newly defined chromatin and/or transcription factor modulatory proteins. Thus evolved the notion of linking.

At its heart, linking is envisioned to be the product of sequential stage-specific binding of transcription factors and chromatin "facilitators" throughout the locus (in a topologically unconstrained form) to an array of chromatin elements that initially define the domain to be transcribed (Bulger and Groudine, 1999; Dorsett, 1999). Thus, the proteins bound to a to-be-transcribed locus are linked to one another by facilitating non-DNA-binding factors, and these form a continuous protein chain from (for example) the LCR to the adult β-*globin* gene located 50 kbp away. One can easily envision ways in which conceptual constraints imposed by this model (for example, intermediate masking points to preferentially transcribe β-*globin* and bypass γ-*globin* during definitive erythropoiesis) could be incorporated.

cis *Speculation*

Given our association with the looping and competition models (Choi and Engel, 1988), we can readily incorporate these new data into predictive corollaries to the competition hypothesis; however, we note that these predictions can be applied to support linking as well. If the human locus is defined by intergenic elements that promote or restrict the ability of the LCR to activate transcription of genes within subdomains (Figure 2), a primary layer of chromatin differentiation could serve to augment the positive or negative regulatory function of discrete transcriptional *cis* elements neighboring the genes. For example, we showed that deletion of the ε-*globin* gene silencer from a YAC has very modest effects on ε-*globin* gene expression in adult erythroid cells (Liu et al., 1997), while the effect of its deletion in transgenic mice bearing only fragments of the locus is quite dramatic (Li et al., 1998). We originally proposed that functional redundancy of *cis* elements could account for the lack of effects detected when the small ε-*globin* gene silencer was deleted from a large transgenic YAC. The report of Gribnau et al. (2000) support an additional alternative: that discrete *cis* element mutations are insufficient to confer full expressivity of local "silencer" or "enhancer" activities, and that these gene proximal elements can only act at the proper developmental stage when they do so within the context of regionally imposed (metastable) chromatin constraints. Thus, deletion of the ε-*globin* silencer may not confer a significant positive transcriptional effect on the ε-*globin* gene when examined in the context of the whole locus because the chromatin subdomain harboring ε-*globin* is suppressed in adult erythroid cells (i.e., the primary effect of chromatin suppression dominates a secondary effect of silencer deletion).

Clearly there are a host of challenges remaining in this field, not the least of which concerns our ability to clearly define the caveats and possible limitations in interpreting each new study within the context of a huge

body of existing literature. Lest we become complacent, it's important to remember that each new year for the past three decades has been accompanied by often startling revelations about how globin diseases arise, how the locus is regulated, how LCRs function, and, most recently, how chromatin structure impinges on differential modulation of the genes. The formidable tasks of devising clear distinctions among the possible models, as well as executing defining experiments to distinguish among them, is still a challenge we face daily, and often optimistically.

Selected Reading

Ashe, H.L., Monks, J., Wijgerde, M., Fraser, P., and Proudfoot, N.J. (1997). Genes Dev. *11*, 2494–2509.

Bell, A.C., and Felsenfeld, G. (1999). Curr. Opin. Genet. Dev. *9*, 191–198.

Bender, M.A., Bulger, M., Close, J., and Groudine, M. (2000). Mol. Cell *5*, 387–394.

Bulger, M., and Groudine, M. (1999). Genes Dev. *13*, 2465–2477.

Bungert, J., Dave, U., Lim, K.C., Lieuw, K.H., Shavit, J.A., Liu, Q., and Engel, J.D. (1995). Genes Dev. *9*, 3083–3096.

Choi, O., and Engel, J.D. (1988). Cell *55*, 17–26.

Dorsett, D. (1999). Curr. Opin. Genet. Dev. *9*, 505–514.

Epner, E., Reik, A., Cimbora, D., Telling, A., Bender, M.A., Fiering, S., Enver, T., Martin, D.I., Kennedy, M., Keller, G., et al. (1998). Mol. Cell *2*, 447–455.

Forrester, W.C., Thompson, C., Elder, J.T., and Groudine, M. (1986). Proc. Natl. Acad. Sci. USA *83*, 1359–1363.

Forrester, W.C., Epner, E., Driscoll, M.C., Enver, T., Brice, M., Papayannopoulou, T., and Groudine, M. (1990). Genes Dev. *4*, 1637–1649.

Gribnau, J., Diderich, K., Pruzina, S., Calzolari, R., and Fraser, P. (2000). Mol. Cell *5*, 377–386.

Grosveld, F., van Assendelft, G.B., Greaves, D.R., and Kollias, G. (1987). Cell *51*, 975–985.

Hardison, R., Riemer, C., Chui, D.H., Huisman, T.H., and Miller, W. (1998). Genomics *47*, 429–437.

Li, Q., Blau, C.A., Clegg, C.H., Rohde, A., and Stamatoyannopoulos, G. (1998). J. Biol. Chem. *273*, 17361–17367.

Liu, Q., Bungert, J., and Engel, J.D. (1997). Proc. Natl. Acad. Sci. USA *94*, 169–174.

Navas, P.A., Peterson, K.R., Li, Q., Skarpidi, E., Rohde, A., Shaw, S.E., Clegg, C.H., Asano, H., and Stamatoyannopoulos, G. (1998). Mol. Cell Biol. *18*, 4188–4196.

Peterson, K.R., Clegg, C.H., Navas, P.A., Norton, E.J., Kimbrough, T.G., and Stamatoyannopoulos, G. (1996). Proc. Natl. Acad. Sci. USA *93*, 6605–6609.

Reik, A., Telling, A., Zitnik, G., Cimbora, D., Epner, E., and Groudine, M. (1998). Mol. Cell Biol. *18*, 5992–6000.

Tuan, D., Solomon, W., Li, Q., and London, I.M. (1985). Proc. Natl. Acad. Sci. USA *82*, 6384–6388.

Wijgerde, M., Grosveld, F., and Fraser, P. (1995). Nature *377*, 209–213.