Acad. Sci. USA 88, 204–208
10 Nathan, C.F. and Hibbs, J.B. (1991) *Curr. Opin. Immunol.* 3, 65–70
11 Govoni, G. *et al.* (1995) *Genomics* 27, 9–19
12 Cellier, M. *et al.* (1994) *J. Exp. Med.* 180, 1741–1752
13 Gruenheid, S., Cellier, M., Vidal, S. and Gros, P. (1995) *Genomics* 25, 514–525
14 Cellier, M. *et al.* (1995) *Proc. Natl Acad. Sci. USA* 92, 10089–10093
15 Rodrigues, V., Cheah, P.Y., Ray, K. and Chia, W. (1995) *EMBO J.* 14, 3007–3020
16 Belouchi, A. *et al.* (1995) *Plant Mol. Biol.* 29, 1181–1196
17 West, A.H. *et al.* (1992) *J. Biol. Chem.* 267, 24625–24633
18 Reithmeier, R.A.F. (1995) *Curr. Opin. Struct. Biol.* 5, 491–500
19 Wood, M.W., Vandongen, H.M.A. and Vandongen, A.M.J. (1995) *Proc. Natl Acad. Sci. USA* 92, 4882–4886
20 Descoteaux, S., Yu, Y. and Samuelson, J. (1994) *Infect. Immunol.* 62, 3572–3575

The complexities of β globin gene regulation

WILLIAM G. WOOD
bwood%immads2.jr2@ox.ac.uk

MRC MOLECULAR HAEMATOLOGY UNIT, INSTITUTE OF MOLECULAR MEDICINE, UNIVERSITY OF OXFORD, JOHN RADCLIFFE HOSPITAL, HEADINGTON, OXFORD, UK OX3 9DU.

The human β globin gene cluster has been one of the favourite model systems for analysing the control of gene expression, particularly developmental regulation[1]. Not only is the system of innate biological interest, but an understanding of its regulation could be of enormous therapeutic benefit to sufferers of β thalassaemia and sickle cell anaemia. Long-term strategies for these disorders include replacement gene therapy and manipulation of the switch from foetal to adult haemoglobin production, but both approaches will require a fuller understanding of the regulatory sequences within the gene cluster and the *trans*-acting factors with which they interact.

The β globin LCR: separate elements or a holocomplex?

The embryonic (ε), foetal (Gγ and Aγ) and adult (δ and β) β-like globin genes are regulated by an upstream region known as the locus control region (LCR)[2]. This 'super enhancer' confers tissue-specific position-independent expression on attached globin genes in transgenic mice through its ability to provide an open chromatin domain wherever it integrates into the genome. A thorough understanding of how the LCR works will be a necessary prerequisite to effective gene therapy for the β globin gene disorders, but dissecting out the various functions of these complex regulatory regions has not proved easy. The β globin LCR spans about 20 kb and is marked by four DNase I hypersensitive sites (HS1 to HS4; Fig. 1), which mark the core regions of separate elements. Each core element contains binding sites for ubiquitous transcription factors as well as the erythroid-restricted factors, such as GATA1 and NF-E2 (Ref. 3).

It is believed that activation of the globin genes involves contact between the LCR and the promoters (or between the proteins bound to each), with looping out of the intervening chromatin. As yet there is no direct evidence for this. Furthermore, whether or not the LCR acts as a single holocomplex or its individual elements are free to interact with different promoters has been a matter of debate and speculation[4,5]. Recently, Wijgerde *et al.*[6] have used probes to visualize the short-lived primary transcripts in transgenic mouse erythroid cells at the time of human γ to β globin gene switching. Both genes are expressed within the same cell, but analysis of the nascent transcripts strongly suggests that only one gene at a time can be transcribed from each chromosome. The genes appear to be in competition for activation by the LCR and this interaction is a dynamic process with the potential for the LCR to switch from one promoter to another after one or more rounds of transcription. These results are clearly consistent with the LCR acting as a holocomplex, rather than individual sites acting on the γ and β promoters at the same time, but do not exclude the possibility that different HS elements preferentially interact with different promoters resulting in the LCR complex adopting alternative conformations.

Hypersensitive site 'knockouts'

The size of the intact LCR precludes its use in retroviruses for gene therapy and, therefore, an understanding of the roles of the individual HS elements will be a prerequisite for designing an effective 'mini-LCR'. Three of the individual elements (HS2, -3 and -4) can give high level (but not maximum) position-independent expression when linked separately to globin genes and additive effects are observed when different combinations of these sites are used[2,7–9]. One potential problem with such experiments though, is that small artificial constructs lack many of the known (let alone unknown) regulatory regions. Gene order and distance from the LCR are important for correct regulation and these factors are often altered in small constructs. Therefore, to determine the function of each element it is also important to determine the effects of their removal from the intact locus. Using knockout technology in ES cells, first, to replace the mouse HS2 with a *neo* gene and, second, to remove the *neo* gene using flp recombinase, Fiering *et al.*[10] generated lines of HS2-deleted mice. Homozygotes were viable and the only phenotype observed was a mild reduction in globin mRNA. Interestingly, in mice in which the *neo* gene remained, there was a marked reduction in expression of all the genes and homozygotes died *in utero*. Thus, when the *neo* gene is present within the LCR it disrupts regulation, possibly by out-competing the globin gene promoters, but simple loss of HS2 has only minor effects. Deletion of HS2 from a YAC containing the whole β globin gene cluster also resulted in a mild reduction in output of all the genes without

Copyright © 1996 Elsevier Science Ltd. All rights reserved. 0168-9525/96/$15.00
PII: S0168-9525(96)30043-7

affecting erythroid specificity or developmental regulation[11]. If the LCR does, indeed, act as a single complex it appears to be able to withstand the significant disruption that the loss of HS2 would seem likely to cause.

The effects of deletions of HS3 from the YAC in transgenic mice have produced somewhat conflicting results. Peterson *et al.*[11] found a decrease in ϵ gene expression with a concomitant increase in γ mRNA in embryonic blood and little or no effect on the γ and β genes in foetal or adult life. Bungert *et al.*[12], however, reported considerably reduced levels of expression (less than 10% of normal) of all the globin genes throughout development, albeit with some differences between the two lines studied. Bungert *et al.*[12] also showed a similar or greater downregulation of all the genes when HS4 was deleted from the YAC. Furthermore, expression levels were restored when HS4 was substituted by HS3, but not in the converse situation.

How is one to interpret the discrepancies in these two reports? HS2, -3 and -4 originally appeared to be similarly effective in transgenic mice when used alone or combined to make a 'mini' LCR. However, most mice contain multiple concatamerized copies of the transgenes and more recent data[13] suggests that the functions of HS2 and HS3 can be differentiated in single copy mice. High level expression was only observed with HS3β mice, suggesting that only this element includes a dominant chromatin opening function, a result consistent with the loss of HS3 being more detrimental than losing HS2. However, preliminary evidence suggests that HS4 also lacks this function[13], which does not explain why its deletion should be more similar to HS3 than HS2. Furthermore, if loss of HS3 does dramatically reduce β gene transcription, it is surprising that there are no natural mutations that mimic this deletion, with over 150 different β thalassaemia alleles described.

Could it be that subtle rearrangements that frequently accompany the transference of large fragments of DNA, such as YACs, have resulted in the differences between these two reports? Rearrangements were observed by Peterson *et al.*[11] (although not within the globin gene complex itself) and, while the mice generated by Bungert *et al.*[12] appeared to contain single intact copies of the YAC, it will require more detailed mapping to confirm that this is the case. An alternative explanation is that different amounts of DNA were deleted from the HS elements by the two groups. Peterson *et al.* deleted about 2 kb from HS2 or HS3, while Bungert *et al.* removed only the core region (about 250 bp) of either HS3 or HS4. Is it possible that sequences around the core can still interact with the other HS elements in a holocomplex and so loss of the core alone can be more disruptive than loss of the whole element? This brings us back to whether the LCR elements really do form a single complex and, if so, how much flexibility there is in its assembly. New techniques that allow direct examination of the LCR-promoter interaction might be necessary to answer these questions.

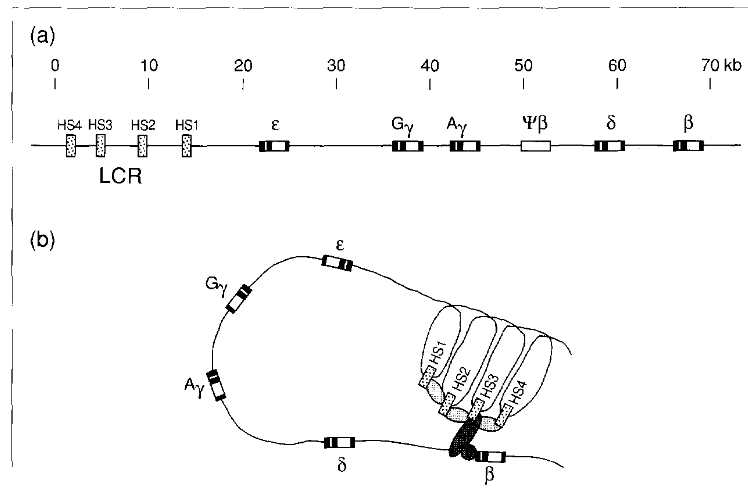

FIGURE 1. (a) The organization of the β globin cluster and its major regulatory region, the locus control region (LCR). (b) Representation of a possible interaction between the LCR elements to form a holocomplex activating the β gene promoter (modified from Ref. 6). The existence of such a complex has yet to be demonstrated (see text).

Developmental regulation: *trans*-acting factor dependent?

If the mechanism by which the *cis*-active elements regulate globin gene transcription require further clarification, what of the *trans*-acting factors that must mediate the various interactions? Erythroid-restricted and ubiquitous factors bind to the LCR elements and the globin gene promoters. It is generally assumed that among these (or interacting with them) are developmental stage-specific factors that mediate activation of the required gene at the appropriate time. Searches for such factors have so far had limited success. Two candidates have been partially characterized (one that binds to the β gene in adult chicken erythroblasts[14], the other to the human γ gene promoter[15]) and contain the common ubiquitous factor CP2, but the erythroid component has proved elusive so far[15]. Another factor, EKLF (erythroid *Krüppel*-like factor), binds specifically to the β globin gene CACCC box (Ref. 16), and when its gene is knocked out in mice, severe β thalassaemia results[17,18]. However, EKLF seems to be present in erythroid cells at all stages of development[19]. While it is clearly essential for adult β globin gene expression, it could be that structural modifications or quantitative changes are required for it to become active. It could also be that epigenetic changes affecting the chromatin conformation of the complex[20] prevent access of EKLF to the β gene early in development.

An alternative approach to identifying *trans*-acting factors involved in globin gene regulation is to identify genes that affect haemoglobin production, but are not linked to the β globin cluster. One such condition is heterocellular hereditary persistence of foetal haemoglobin (HPFH), which results in small increases in foetal haemoglobin (HbF) in adults, but interacts with the severe genetic anaemias to produce sufficiently high levels of HbF to ameliorate the disease. Identification of the gene has been difficult because of genetic

heterogeneity, phenocopies and accurate phenotype identification. However, by using a large inbred kindred, Craig *et al.*[21] have localized a gene involved in HPFH to 6q22.3–q24. The power of multiple regression analysis was used to sift out the confounding effects of other genetic and environmental influences on a quantitative continuum of HbF levels. Identification of the gene itself will be no trivial problem because, not only is its phenotype subtle, but it is unclear whether its action on the globin genes is direct or indirect, for instance, by altering the kinetics of erythropoiesis. Nevertheless, this example and the dissection of LCR action described above both suggest that a number of avenues are available that could ultimately lead to improvements in the treatment of thalassaemia and related disorders.

References

1. Stamatoyannopoulos, G. and Nienhuis, A.W. (1993) in *The Molecular Basis of Blood Diseases* (Stamatoyannopoulos, G., Nienhuis, A.W., Majerus, P. and Varmus, H., eds), pp. 107–154, W.B. Saunders
2. Grosveld, F., Blom van Assendelft, G., Greaves, D.R. and Kollias, G. (1987) *Cell* 51, 975–985
3. Grosveld, F., Dillon, N. and Higgs, D.R. (1993) in *Baillière's Clinical Haematology: The Haemoglobinopathies* (Higgs, D.R. and Weatherall, D.J., eds), pp. 31–56, Baillière Tindall
4. Fraser, P., Pruzina, S., Antoniou, M. and Grosveld, F. (1993) *Genes Dev.* 7, 106–113
5. Engel, J.D. (1993) *Trends Genet.* 9, 304–309
6. Wijgerde, M., Grosveld, F. and Fraser, P. (1995) *Nature* 377, 209–213
7. Fraser, P., Hurst, J., Collis, P. and Grosveld, F. (1990) *Nucleic Acids Res.* 18, 3503–3508
8. Collis, P., Antoniou, M. and Grosveld, F. (1990) *EMBO J.* 9, 233–240
9. Forrester, W.C., Novak, U., Gelinas, R. and Groudine, M. (1989) *Proc. Natl Acad. Sci. USA* 86, 5439–5443
10. Fiering, S. *et al.* (1995) *Genes Dev.* 9, 2203–2213
11. Peterson, K.R. *et al. Proc. Natl Acad. Sci. USA* (in press)
12. Bungert, J. *et al.* (1995) *Genes Dev.* 9, 3083–3096
13. Ellis, J. *et al.* (1996) *EMBO J.* 15, 562–568
14. Choi, O-R.B. and Engel, J.D. (1988) *Cell* 55, 17–26
15. Jane, S.M., Nienhuis, A.W. and Cunningham, J.M. (1995) *EMBO J.* 14, 97–105
16. Miller, I.J. and Bieker, J.J. (1993) *Mol. Cell. Biol.* 13, 2776–2786
17. Perkins, A.C., Sharpe, A.H. and Orkin, S.H. (1995) *Nature* 375, 318–322
18. Nuez, B. *et al.* (1995) *Nature* 375, 316–318
19. Donze, D., Townes, T.M. and Bieker, J.J. (1995) *J. Biol. Chem.* 270, 1955–1959
20. Stanworth, S.J. *et al.* (1995) *Mol. Cell. Biol.* 15, 3969–3978
21. Craig, J.E. *et al.* (1996) *Nat. Genet.* 12, 58–64

Myosin diversity and disease

ALAN F. WRIGHT AND IAN JACKSON
alanw@hgu.mrc.ac.uk
ian@hgu.mrc.ac.uk

MRC HUMAN GENETICS UNIT, WESTERN GENERAL HOSPITAL, CREWE ROAD, EDINBURGH, UK EH4 2XU.

It is generally believed that molecular motors, such as dynein, kinesin and myosin, act by the application of force to relatively fixed structures, such as actin and microtubule filaments or membranes, to bring about movement of the cytoskeleton. The myosin superfamily is a diverse set of motor proteins grouped on the basis of the structural similarities of the myosin 'motor' or head domain[1,2]. The myosin head is an actin-associated motor consisting of a relatively conserved domain of about 80 kDa that can undergo actin-stimulated ATP hydrolysis. Specialized neck and tail domains can contain binding sites for calmodulin, ATP-independent actin-binding sites and membrane-binding sites suitable for anchoring the protein. The most striking property of some myosin tails, such as the 'conventional' (type II) myosin of smooth and striated muscle, is the ability to assemble into α-helical coiled-coils and to promote dimerization of the heavy chain heads with the formation of thick anti-parallel filaments. Myosin-generated movements along an actin 'tightrope' can be measured and although it is still not clear whether or not two heads are better than one, single heads powered by ATP hydrolysis can produce short 4 nm steps, which are consistent with conformational change models based on analysis of the myosin crystal structure[3,4]. Larger steps might require cooperative interactions or a particular head orientation.

At least 11 major classes of myosin can be defined, each of which has its origins as far back as early eukaryotic evolution[5,6]. A summary of the current classification is shown in Table 1. Their functions are clearly diverse and in many cases remain to be elucidated, but there is evidence of myosin involvement in vesicle trafficking, cytokinesis, budding, phagocytosis, contractile vacuoles and cellular movement as well as in muscle contraction[1,2]. Single-cell studies have shown the overlapping expression of at least 11 different 'unconventional' (non-type II) myosins in different cell types[7]. Some degree of functional redundancy is possible. Interactions with microtubule-based trafficking systems have also been proposed; Golgi-derived vesicles are transported by dynein along microtubules and then by myosin to the plasma membrane along an actin 'web'[8].

Does the analysis of myosin mutations throw any light on the roles of this diverse family? Inherited disorders of conventional myosin can cause a form of familial hypertrophic cardiomyopathy (FHC), an autosomal dominant disorder in which there is hypertrophy of the heart, disarray of the myocytes and a risk of sudden death[9]. Mutations in at least five different genes can cause FHC, four of which

Copyright © 1996 Elsevier Science Ltd. All rights reserved. 0168-9525/96/$15.00
PII: S0168-9525(96)30050-4