Developmental regulation of eukaryotic gene loci

which *cis*-regulatory information is required?

It is becoming increasingly accepted that gene loci comprise an extensive *cis*-regulatory system that encodes different layers of regulatory information, all of which are necessary to achieve and maintain tissue-specific gene expression in ontogeny. To gain a detailed understanding of developmental processes, it is clearly necessary to unravel the molecular basis behind the different regulatory processes that control gene expression. This information is also of utmost importance for any practical application that uses gene transfer technology.

The *cis*-regulatory elements, which are required for the correct activation of eukaryotic genes during development, are often separated by long DNA segments that, at first sight, appear to serve no purpose. This led researchers to believe that they were dispensable for the control of gene expression and initiated a number of experimental attempts to recapitulate gene-expression patterns with expression cassettes that carried deletion constructs or artificial combinations of minimal enhancer and promoter sequences. However, these studies were often designed without taking all of the structural and functional considerations that might influence gene expression into account, and as a consequence often gave inexplicable results. Recent experiments that have examined the structural and functional properties of individual gene loci and gene clusters indicate that the control of gene expression is more complex and requires more sequence information than was previously thought. The regulation of the spatial and temporal pattern of expression observed with any given gene turns out to be the result of multiple regulatory inputs within a genetic locus. A number of recent findings shed light on the regulatory sequence information required for correct control of gene expression.

Constanze Bonifer
c.bonifer@leeds.ac.uk

Molecular Medicine Unit, University of Leeds, St James's University Hospital, Leeds UK LS9 7TF.

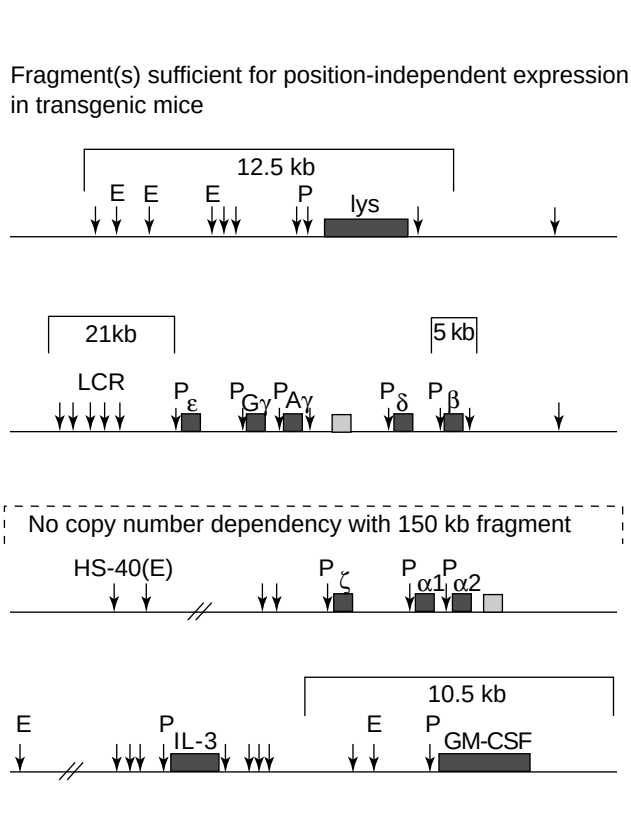
The *cis*-regulatory regions can be larger than anticipated

A gene locus encodes all necessary information for spatially and temporally correct gene expression in ontogeny, as well as the appropriate level of expression. However, at present our understanding of how this is achieved is incomplete. Much research in the past few years has been devoted to the characterization of sequence information necessary to replicate the expression pattern of endogenous loci at ectopic sites of integration. To examine all the regulatory properties of a gene locus, gene constructs must be assayed in transgenic animals. Tissue-specific and position-independent expression at the level of the endogenous gene was often (but not always, see below) observed with transgenes that harbour large genomic fragments[1]. This could even be observed when transgenes were inserted into heterochromatic regions[2]. The molecular basis for these results is that sequences that comprise the information for correct spatial and temporal regulation of a particular gene locus during development often extend over hundreds of kilobases of DNA. A few examples of complex genetic loci for which correct *in vivo* regulation have been demonstrated are depicted in Fig. 1.

In the case of the chicken lysozyme locus, any of the three enhancer elements can be deleted in the context of the whole locus without losing tissue-specific expression. However, expression levels vary with the chromosomal insertion sites of the transgene[3] and furthermore, with the deletion of the most distal enhancer at –6.1 kb, the developmental regulation of the gene is also disturbed[4]. A similar phenomenon is observed when any one of the DNaseI-hypersensitive sites of the major regulatory element of the human β-globin locus, the control region (LCR), is removed[5,6]. Interestingly, this is observed only in the context of the complete locus. When the LCR is fused directly with the gene for β-globin (Fig. 1), the transgene is expressed in a position-independent fashion[7]; however, it is not correctly regulated in development[8]. This indicates that

 PII: S0168-9525(00)02029-1

FIGURE 1. *cis*-regulatory requirements for correct transgene expression



Gene locus	*Cis*-regulatory requirements for correct transgene expression (tissue specificity, correct developmental regulation, position independence)
Chicken lysozyme locus	All *cis*-regulatory elements plus flanking regions necessary.
Human β-globin cluster	All elements of LCR necessary. Cluster has to be in original conformation.
Human α-globin cluster	Larger fragments show correct developmental regulation, but not high expression level.
Human GM-CSF locus	Correct tissue specificity of expression. Developmental regulation not known.

trends in Genetics

The organization of coding and *cis*-regulatory regions for four gene loci are shown, each of them serving as a paradigm for different *cis*-regulatory requirements for correct transgene expression. The extensions of regulatory information necessary for correct transgene expression are shown, and the sequence arrays needed for position-independent expression only are indicated by brackets. The individual elements are not drawn to scale. Coding regions are depicted as red boxes (lighter in case of pseudogenes) with the name of the gene indicated above. DNaseI-hypersensitive chromatin sites, some of which mark the position of *cis*-regulatory elements like enhancers (E) and promoters (P), are indicated as vertical arrows.

all *cis*-regulatory elements of a gene locus have to be present on a transgene in order to recapitulate all aspects of developmental regulation. It also indicates that, in different gene loci, *cis*-regulatory elements cooperate differently.

What comprises a *cis*-regulatory element? A large number of experiments aimed at unravelling the molecular basis of tissue-specific gene expression have ascribed major control functions to only a limited collection of *cis*-regulatory sequence arrays. However, for those genes whose function and tissue-specific regulation are conserved in evolution, a surprising degree of conservation in intron–exon structure, in the position of *cis*-regulatory elements and in the transcription-factor binding-site composition within these elements is often observed. This suggests that *cis*-regulatory elements could be much larger than previously anticipated, and that they might contain previously undefined structural and functional information. Supporting evidence for this idea was forthcoming when large sequence arrays of known regulatory function were compared between species, which yielded a 'phylogenetic' footprint. For example, with the β-globin LCR, this type of analysis revealed a striking conservation of sequence motifs outside of the *cis*-regulatory elements previously mapped by classical structural and functional assays (reviewed in Ref. 9). Within a region of 1400 bp 5′ and 3′ around the core of one of the HS3 subelements, a large number of short conserved sequence motifs were discovered that contribute to regulation of the gene for globin. They bind transcription factors and their presence or absence makes a pronounced difference to reporter-gene activity in stable (but not transient) transfections assays[10,11]. This indicates a regulatory role that is apparent only in chromatin. A possible molecular explanation for this finding is that DNA is wrapped around nucleosomes, which brings factor binding sites together, thereby allowing the recruitment of proteins that flank the enhancer core into a large complex (as reviewed in Ref. 12). In support of this hypothesis, it was found that the deletion of enhancer core flanking sequences can affect the chromatin-remodelling activity of these elements[13,14]. These observations indicate that standard assays such as *in vitro* DNA binding and transfection studies, identify only clusters of high-affinity transcription factor binding sites, which might represent only a part of a much larger regulatory complex. Although such assays suggest a high degree of redundancy in transcription factor use, it is now apparent that all factors are required for the precise spatial and temporal control of gene expression in a chromatin context.

Differential gene regulation requires defined and correctly spaced sets of *cis*-elements

Experiments that attempt to achieve stable gene transfer using artificial or nonhomologous enhancer–promoter combinations have, very often unaccountably, failed. This is in part because of the absence of important *cis*-regulatory elements, as described above. However, an additional reason for such failure is probably that such constructs did not obey the specificity and spatial requirements for the formation of a robust, high-affinity transcription complex that is required for correct gene locus activation.

Important insights into the requirements for *cis*-element interactions in development were achieved when the spatial organization or the sequence composition of transgenes was altered relative to their natural counterparts. A change in gene order (or distance between *cis*-regulatory elements) or the reversal of the orientation of the LCR within the globin cluster disturbs the regulation of the globin genes[8,15,16]. Furthermore, substituting one component of the LCR with another sometimes disturbs the developmental regulation of the locus (e.g., see Refs 6, 17). Even in the presence of the LCR, the insertion of a *LacZ*-expression cassette interferes with regulation of the gene encoding β-globin in transgenic mice[18]. Insertion of a heterologous coding region or a heterologous promoter into genes in their natural chromosomal location can lead to their deregulation, as was observed after the insertion of a selectable marker expression cassette into the granzyme gene cluster[19]. A major reduction of expression of the globin genes was also observed when the same marker cassette was inserted into the globin LCR. Strikingly, this inhibition was alleviated after removal of the inserted cassette by CRE-mediated recombination[20]. Therefore, deregulation was caused by the disturbance of the *cis*-regulatory system of the respective gene(s) and not by the removal of an important *cis*-element. The most likely explanation for such a phenomenon is that these insertions interfere with the assembly of a specific complex that comprises a large number of interacting components. However, with the discovery of intergenic transcripts in a variety of gene loci, an additional, but not mutually exclusive, explanation for the interference of an ectopic promoter is possible. In the β-globin locus, intergenic transcripts have been identified that initiate at developmentally regulated promoters upstream of the LCR, and between the embryonic–foetal and adult globin genes. The presence of the latter is vital for transcription of the adult globin genes[21]. It was suggested that by recruiting chromatin-remodelling complexes to the basal transcription machinery, chromatin between and around the genes would be modified progressively, leading to an increase in general DNaseI sensitivity and factor accessibility of the promoters of the genes located downstream. If ectopic promoters had similar activities, this would interfere with the activity of the endogenous 'chromatin-opening' elements and suggests a molecular explanation for their long-range deregulatory effects. Even without knowing the exact mechanism(s) of deregulation, it is evident that the insertion of ectopic sequence information into a gene locus can lead to unpredictable deregulatory effects on transgene regulation.

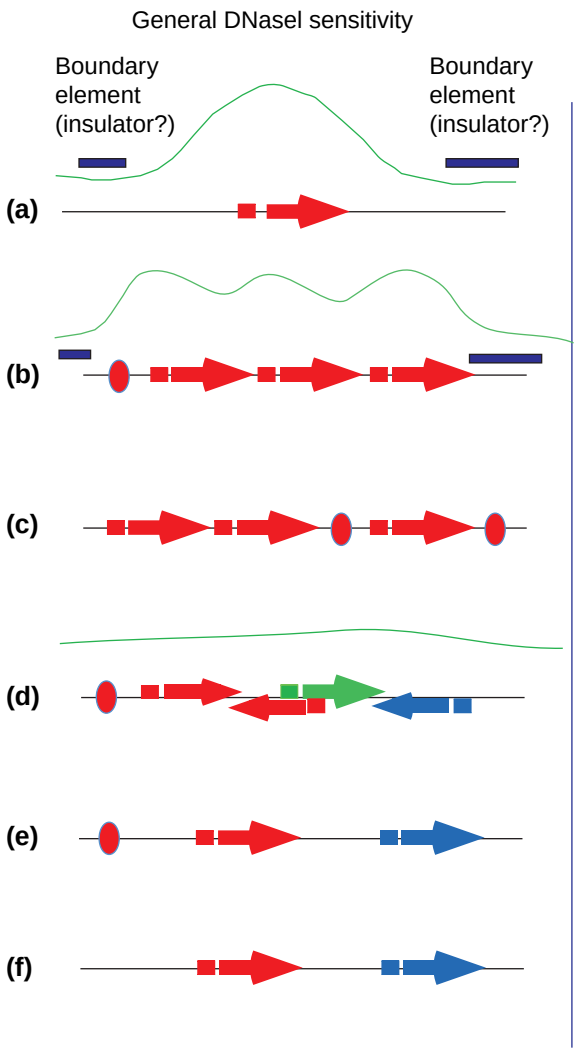
The regulatory organization of the eukaryotic genome

The intriguing finding that gene loci can be arranged in the order of their expression during ontogeny suggests that the spatial organization of genes might also be a parameter that influences their regulation. This clearly holds true for the mammalian β-globin and for the *Hox* gene clusters, within which the most 5′ located gene is expressed first. The order of genes, including their local *cis*-regulatory elements is crucial for correct developmental regulation within each cluster[8,22]. However, as the sequence information of particular genomes and information about the expression profile of individual genes increases, it becomes clear that the spatial arrangements of genes on their respective chromosomes do not always reflect the way in which they are regulated (Fig. 2). Coordinately regulated gene loci, such as the human apolipoprotein E (Apo E) locus can be clustered as well[23]. Two of the examples listed in Fig. 2 (the developmentally regulated β-globin cluster and the coordinately regulated Apo E cluster) contain shared LCR-like *cis*-regulatory elements (reviewed in Ref. 24). However, clusters of differentially regulated genes have also been found to share *cis*-regulatory elements[25], in contrast to others, in which individual genes are regulated autonomously[26]. In extreme cases, such as in the α-globin locus, differentially regulated genes can be organized in overlapping transcription units. Because all genes possess regulatory regions, the question arises of how the different regulatory units are functionally separated in order to prevent *cis*-regulatory elements capable of acting over long ranges from interfering with each other. Two main mechanisms to explain such a functional compartmentalization have been suggested: the organization of coordinately regulated or single genes in chromatin domains, as exemplified by the chicken lysozyme locus or the β-globin gene cluster; and promoter specificity, which allows only interaction between matching *cis*-regulatory elements. It has emerged that both models are correct.

Chromatin domains were originally defined by an increase in accessibility to DNaseI (Ref. 1). In several cases it was demonstrated that the boundary region between open and closed chromatin is marked by a region with insulator properties. These regions are proposed to protect transgenes from chromosomal position effects[27]. However, not all gene loci have a defined structurally organized chromatin domain. Support for promoter specificity is provided by evidence from experiments that study enhancer–promoter interactions at the neighbouring *decapentaplegic* (*dpp*), SLY1 *homologous* (*slh*) and *out at first* (*oaf*) loci in *Drosophila*, which demonstrated only tissue-specific interactions between matching *cis*-regulatory elements, irrespective of their location[28]. The human α-globin cluster is located in a region of constitutively open chromatin at the telomeric end of chromosome 16q (Ref. 29). Surprisingly, the most important regulatory region of the α-globin locus overlaps with the intron of a nearby gene. The α-globin genes are surrounded by differently regulated genes[30], indicating that α-globin gene promoters are unable to interact with the elements that regulate the activities of these genes.

Overlapping transcriptional units that are regulated by specific interactions between *cis*-regulatory elements raise the interesting issue of interdependencies of functional and structural features of individual regulatory units. In this respect it is noteworthy that the sequence organization of the genes within the β-globin locus and also of differently regulated flanking genes is highly conserved between mouse and man[31]. Interestingly, it has not yet been possible to define conditions in which the α-globin locus is

FIGURE 2. The regulatory organization of the eukaryotic genome



Gene locus arrangement and regulation	Examples
Individual gene locus	Chicken lysozyme locus Human apo B locus
Tissue specifically but developmentally regulated multigene cluster	Vertebrate β-globin cluster[1]
Coordinately regulated multigene cluster	Human ApoE cluster
Differentially regulated multigene cluster with overlapping transcription units	α-globin cluster
Differentially regulated multigene cluster with shared control sequences	Human growth hormone cluster
Differentially regulated multigene cluster without shared control sequences	IL-3, GM-CSF locus

trends in Genetics

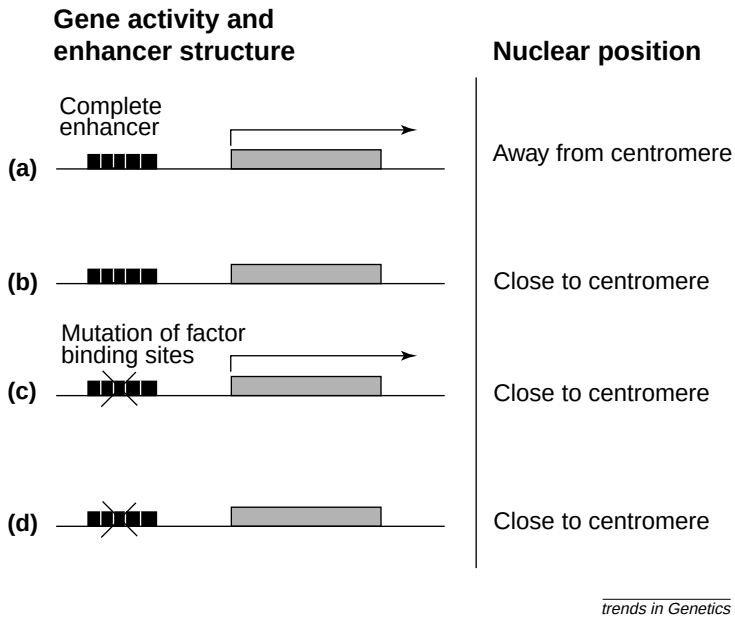
The left panel depicts six different potential gene arrangements and some examples of genes arranged in the indicated manner are given on the left. *Cis*-regulatory elements are indicated as rectangles or ovals, whereby matching elements are depicted in the same colours. The green line depicts the extensions of DNaseI-sensitive chromatin, if known. The boundary regions between DNaseI-sensitive and -insensitive chromatin, which might contain insulator sequences, are indicated as blue boxes. (a) A gene cluster with its *cis*-regulatory region organized in a DNaseI-sensitive chromatin domain. (b) A cluster of developmentally regulated genes expressed in one tissue type. These genes can share important *cis*-regulatory elements (oval), as exemplified by the β-globin locus control region (Ref. 1). For the human β-globin cluster, subdomains of DNaseI sensitivity have been identified[21]. (c) A multigene cluster harbouring coordinately regulated genes with a shared *cis*-regulatory element. (d) A multi-gene cluster comprising of differentially regulated genes that are organized in overlapping transcription units, some of which share a *cis*-regulatory element. (e) Differentially regulated genes with shared *cis*-elements. (f) Differentially regulated genes organized in separate transcription units with *cis*-regulatory elements that are specific for each gene.

expressed in a position-independent fashion, even though large fragments encompassing extensive flanking regions have been used to generate transgenic mice (reviewed in Ref. 29). These data indicate that the natural chromosomal environment could participate in regulation. This phenomenon has also been observed previously with genes that are normally localized in heterochromatic regions in *Drosophila* (Ref. 32). Here, regulation is disturbed when these genes are placed into euchromatic regions of the genome. Such 'next-neighbour' effects, which highlight a role of the particular chromosomal environment on gene expression, might provide part of the molecular explanation for some as yet unexplained discrepancies in the *cis*-regulatory requirements of large transgene constructs, compared with their function in their natural chromosomal location (reviewed in Ref. 33). For example, a deletion of the LCR in its natural genomic context reduces the expression level, whereas a deletion of the β-globin LCR in a transgenic locus leads to integration-site-dependent, highly variable expression levels[34,35].

Structural and spatial organization of gene loci within the nucleus

A wealth of experiments indicates that a developmental-specific chromatin architecture is an essential prerequisite for correct gene regulation. It exceeds the scope of this

FIGURE 3. Enhancers are involved in the regulation of intranuclear position



Intranuclear localization is dependent on the presence of an intact enhancer, but not on expression of a transfected gene. At one particular chromosomal integration site, reporter genes can exist in two states: active and inactive. The ratio of active and inactive genes at a given time after integration is a function of the chromosomal position and the binding-site composition of the enhancer. Enhancers are depicted as red boxes with binding sites indicated as black stripes. The reporter gene is indicated as a blue box. The expression status is shown by the presence or absence of an horizontal arrow. (a) Stably integrated reporter construct driven by an intact enhancer that is expressed. (b) Intact enhancer that has been silenced because of genomic position effects at its specific integration site. (c) Mutated enhancer that is integrated at chromosomal position where it is rapidly silenced, but is still expressed. (d) Mutated enhancer at the same chromosomal position as in (c), but silenced.

article to describe this in detail (see Refs 36, 37 for reviews). DNA, transcription factors, polymerases and associated proteins, as well as chromatin components, all contain distinct biochemical tags such as methylation, phosphorylation or acetylation, which define their current regulatory status[38]. Thus, the genome participates in development and establishes heritable patterns of gene expression by acquiring cell-specific epigenetic states. In turn, this suggests that correct development depends on a coordinated series of epigenetic reorganizations that must occur at the right time and in the correct order. It is most likely that the complexity of such developmentally controlled processes requires most, if not all, the sequence information of a gene locus. However, our current understanding of how structural information is encoded in the DNA sequence is in its infancy.

One unexpected activity of *cis*-regulatory elements that is worth highlighting in more detail is their possible role as regulators of intranuclear position. It is now clear that the nucleus is a compartmentalized structure with particular chromosomes occupying distinct territories[39]. A variety of findings has indicated that such a compartmentalization might also be of functional relevance. In yeast, telomeres are clustered near the nuclear envelope and the integration of genes into telomeric regions leads to gene silencing[40]. In mammalian cells, an analogy is the association of transcriptionally silent genes with centromeric heterochromatin[41]. Genes reposition within the nucleus during cell differentiation[42] and, in addition, specific chromosomes co-localize with large, specific transcription factor assemblies[43]. More direct evidence for a correlation between gene function and nuclear position has come from a series of elegant experiments described by Francastel *et al.*[44], which are summarized in Fig. 3. They demonstrated that the presence or absence of a functional enhancer (and not the expression status of the linked gene) determines the intranuclear localization of transgenes inserted at the same chromosomal position in relation to centromeric heterochromatin. However, this is true only for chromosomal integration sites that are nonpermissive for expression where integrated transgenes are progressively silenced. The finding that gene silencing and centromeric position is a function of the chromosomal integration site, and thus of the flanking chromatin structure, indicates a complex regulation of intranuclear positioning in natural gene loci.

Outlook

It is clear that eukaryotic genomes comprise highly organized structural and functional units that are shaped by evolution, participate in development by acquiring distinct epigenetic states and might occupy distinct positions in the nucleus. As a consequence, development-specific gene locus activation is a highly complex multistep process that requires the precisely timed and ordered interactions of the basal transcription machinery, tissue-specific activator and repressor proteins, and the different chromatin-modification complexes. The assembly of these components is dynamic and is influenced by the chromosomal environment, chromatin architecture and higher-order structures within the nucleus.

The findings described here have far-reaching practical implications. Without obeying the rules governing the control of gene expression in chromatin, the outcome of gene transfer will always have an element of uncertainty.

Acknowledgements

I apologize to all my colleagues whose work I have not been able to cite owing to space constraints. I thank P. Fraser for sharing unpublished results, my colleagues A. Markham and P. Robinson for critically reading the manuscript, and, in particular, R. Hardison and P. Cockerill for important suggestions. Work in the author's laboratory is supported by grants from the Wellcome Trust, the LRF, the BBSRC, the YCR and the Candlelighter's Trust.

References

1 Sippel, A.E. *et al.* (1993) Chromatin domains constitute regulatory units for the control of eukaryotic genes. *Cold Spring Harbor Symp. Quant. Biol.* 58, 37–44
2 Kioussis, D. *et al.* (1997) Locus control regions: overcoming heterochromatin-induced gene inactivation in mammals. *Curr. Opin. Genet. Dev.* 7, 614–619
3 Bonifer, C. *et al.* (1994) Dissection of the locus control function located on the chicken lysozyme gene domain in transgenic mice. *Nucleic Acids Res.* 22, 4202–4210
4 Jägle, U. *et al.* (1997) Role of positive and negative *cis*-regulatory elements regions in the regulation of transcriptional activation of the lysozyme locus in developing macrophages of transgenic mice. *J. Biol. Chem.* 272, 5871–5879
5 Milot, E. *et al.* (1996) Heterochromatin effects on the frequency and duration of LCR-mediated gene transcription. *Cell* 87, 105–114
6 Bungert, J. *et al.* (1995) Synergistic regulation of human β-globin gene switching by locus control region elements HS3 and HS4. *Genes Dev.* 9, 3083–3096
7 Grosveld, F. *et al.* (1987) Position-independent, high-level expression of the human β-globin gene in transgenic mice. *Cell* 51, 975–985
8 Hanscombe, O. *et al.* (1991) Importance of globin gene order for correct developmental expression. *Genes Dev.* 5, 1387–1394
9 Hardison, R.C. *et al.* (1997) Long human–mouse sequence alignments reveal novel regulatory elements: a reason to sequence the mouse genome. *Genome Res.* 7, 959–966

10 Jackson, J.D. *et al.* (1996) Role of DNA sequences outside the cores of DNase hypersensitive sites (HSs) in function of the β-globin locus control region. *J. Biol. Chem.* 271, 11871–11876
11 Shelton, D.A. *et al.* (1997) Phylogenetic footprinting of hypersensitive site 3 of the β-globin locus control region. *Blood* 89, 3457–3469
12 Wolffe, A.P. (1994) Nucleosome positioning and modification: chromatin structures that potentiate transcription. *Trends Biochem. Sci.* 19, 244–247
13 Aronow, B. *et al.* (1995) Dissecting a locus control region: facilitation of enhancer function by extended enhancer-flanking sequences. *Mol. Cell. Biol.* 15, 1123–1135
14 Huber, M.C. and Bonifer, C. (1998) The activation of the chicken lysozyme locus in development is a cooperative process. *Gene Ther. Mol. Biol.* 3, 1–13
15 Dillon, N. *et al.* (1997) The effect of distance on long-range chromatin interactions. *Mol. Cell.* 1, 131–139
16 Tanimoto, K. *et al.* (1999) Effects of altered gene order or orientation of the locus control region on human β-globin gene expression in mice. *Nature* 398, 344–348
17 Sabatino, D.E. *et al.* (1998) Substitution of the human β-spectrin promoter for the human ^{A}γ-globin promoter prevents silencing of a linked human β-globin gene in transgenic mice. *Mol. Cell. Biol.* 18, 6634–6640
18 Guy, L-G. *et al.* (1996) The β-globin locus control region enhances transcription but does not confer position-independent expression onto the *lacZ* gene in transgenic mice. *EMBO J.* 15, 3713–3721
19 Pham, C.T.N. *et al.* (1996) Long-range disruption of gene expression by a selectable marker cassette. *Proc. Natl. Acad. Sci. U. S. A.* 93, 13090–13095
20 Reik, A. *et al.* (1998) The locus control region is necessary for gene expression in the human beta-globin locus but not the maintenance of an open chromatin structure in erythroid cells. *Mol. Cell. Biol.* 18, 5992–6000
21 Gribnau, J. *et al.* (2000) Intergenic transcription and developmental remodelling of chromatin sub-domains in the human β-globin locus. *Mol. Cell.* 5, 377–386
22 Kmita, M. *et al.* (2000) Mechanisms of *Hox* gene colinearity: transposition of the anterior Hoxb1 gene into the posterior *HoxD* complex. *Genes Dev.* 14, 198–211
23 Allan, C.M. *et al.* (1997) Two hepatic enhancers, HCR.1 and HCR.2, coordinate the liver expression of the entire human apolipoprotein E/C-I/C-IV/C-II gene cluster. *J. Biol. Chem.* 272, 29113–29119
24 Li, Q. *et al.* (1999) Locus control regions – coming of age a decade plus. *Trends Genet.* 15, 403–408
25 Su, Y. *et al.* (2000) The human growth hormone gene cluster locus control region supports position-independent pituitary- and placenta-specific expression in the transgenic mouse. *J. Biol. Chem.* 275, 7902–7909
26 Cockerill, P.N. *et al.* (1999) The human granulocyte-macrophage colony stimulating factor gene is autonomously regulated *in vivo* by an inducible tissue-specific enhancer. *Proc. Natl. Acad. Sci. U. S. A.* 96, 15097–15102
27 Bell, A.C. and Felsenfeld, G. (1999) Stopped at the border: boundaries and insulators. *Curr. Opin. Genet. Dev.* 9, 191–198
28 Merli, C. *et al.* (1996) Promoter specificity mediates the independent regulation of neighboring genes. *Genes Dev.* 10, 1260–1270
29 Higgs, D.R. *et al.* (1998) Understanding α-globin expression: a step towards gene therapy. *Semin. Hematol.* 35, 93–104
30 Vyas, P. *et al.* (1992) *Cis*-acting sequences regulating expression of human α-globin cluster lie within constitutively open chromatin. *Cell* 69, 781–793
31 Bulger, M. *et al.* (1999) Conservation of sequence and structure flanking the mouse and human β-globin oci: the β-globin genes are embedded within an array od odorant receptor genes. *Proc. Natl. Acad. Sci. U. S. A.* 96, 5129–5134
32 Weiler, K.S. and Wakimoto, B.T. (1995) Heterochromatin and gene expression in *Drosophila*. *Annu. Rev. Genet.* 29, 577–605
33 Grosveld, F. (1999) Activation by locus control regions? *Curr. Opin. Genet. Dev.* 9, 152–157
34 Epner, E. *et al.* (1998) The β-globin LCR is not necessary for an open chromatin structure or developmentally regulated transcription of the native mouse β-globin locus. *Mol. Cell.* 2, 447–455
35 Bender, M.A. *et al.* (2000) β-globin gene switching and DNaseI sensitivity of the endogenous β-globin locus in mice do not require the locus control region. *Mol. Cell.* 5, 387–393
36 Wolffe, A.P. and Hayes, J.J. (1999) Chromatin disruption and modification. *Nucleic Acids Res.* 27, 711–720
37 Kornberg, R. and Lorch, Y. (1999) Twenty-five years of the nucleosome, fundamental particle of the eukaryotic chromosome. *Cell* 98, 285–294
38 Strahl, B.D. and Allis, C.D. (2000) The language of covalent histone modifications. *Nature* 403, 41–45
39 Zink, D. and Cremer, T. (1998) Cell nucleus: chromosome dynamics in nuclei of living cells. *Curr. Biol.* 8, 321–324
40 Andrulis, E.D. *et al.* (1998) Perinuclear localization of chromatin facilitates transcriptional silencing. *Nature* 394, 592–595
41 Brown, K.E. *et al.* (1997) Association of transcriptionally silent genes with Ikaros complexes at centromeric heterochromatin. *Cell* 91, 845–854
42 Brown, K.E. *et al.* (1999) Dynamic repositioning of genes in the nucleus of lymphocytes preparing for cell division. *Mol. Cell.* 2, 207–217
43 Pombo, A. *et al.* (1998) Regional and temporal specialization in the nucleus: a transcriptionally-active nuclear domain rich in PTF, Oct1 and PIKA antigens associates with specific chromosomes early in the cell cycle. *EMBO J.* 17, 1768–1778
44 Francastel, C. *et al.* (1999) A functional enhancer suppresses silencing of a transgene and prevents its localization close to centromeric heterochromatin. *Cell* 99, 259–269

Organellar genes

why do they end up in the nucleus?

Many mitochondrial and plastid proteins are derived from their bacterial endosymbiotic ancestors, but their genes now reside on nuclear chromosomes instead of remaining within the organelle. To become an active nuclear gene and return to the organelle as a functional protein, an organellar gene must first be assimilated into the nuclear genome. The gene must then be transcribed and acquire a transit sequence for targeting the protein back to the organelle. On reaching the organelle, the protein must be properly folded and modified, and in many cases assembled in an orderly manner into a larger protein complex. Finally, the nuclear copy must be properly regulated to achieve a fitness level comparable with the organellar gene. Given the complexity in establishing a nuclear copy, why do organellar genes end up in the nucleus? Recent data suggest that these genes are worse off than their nuclear and free-living counterparts because of a reduction in the efficiency of natural selection, but do these population–genetic processes drive the movement of genes to the nucleus? We are now at a stage where we can begin to discriminate between competing hypotheses using a combination of experimental, natural population, bioinformatic and theoretical approaches.

Jeffrey L. Blanchard
jlb@ncgr.org

Michael Lynch*
mlynch@oregon.uoregon.edu

National Center for Genome Resources, 2935 Rodeo Park Drive East, Santa Fe, NM 87505, USA.
*Biology Department, University of Oregon, Eugene, OR 97403, USA.

Endosymbiosis, the persistence of one organism within another, can lead to the melding of two organisms into one. At least two common eukaryotic organelles, the mitochondrion and the plastid, are derived from bacteria whose fate became linked with unicellular eukaryotes some two billion years ago. One of the most obvious features of organelle evolution has been the reduction in genome size that occurred following endosymbiosis. Some organellar genes became expendable in the internal environment of the host, and nuclear genes have replaced the function of some organellar genes, but much of the reduction occurred through the transfer of organellar genes to the nucleus[1,2].

Many genes might have been transferred to the nucleus early on, but we now realize that there are different rates of gene transfer and loss in eukaryotic lineages[1,3]. The most extreme cases lie within the cradle of eukaryotic diversity, the protists. The mitochondrial genome of the retortomonad protist, *Reclinamonas americanus*, codes for 97 genes[4] and the plastid genome of the red alga,

0168-9525/00/$ – see front matter PII: S0168-9525(00)02053-9