Editorial

Genetic Variation in the MHC II Promoter: Lessons for Regulation and for Comparative Genomics

N. A. MITCHISON

Windeyer Institute of University College London, London, U.K.

Abstract. Sequence data have been accumulating that reveal variation in gene promoters of the immune system, notably in MHC class II, cytokines and chemokines. The variation is non-random: it occurs most often in proximity to and within certain regulatory elements such as CRE and NFY (in MHC class II these are respectively the X2 and Y boxes). These are elements that are widely used elsewhere in the genome, and appear to act as rheostats (modulators of expression) in contrast to the type of on-off switch operated by the RFX element that is unique to a single family of promoters such as MHC class II. It is proposed that a complex mouse phenotype described in Prague and elsewhere may reflect this pattern of variation in/around CRE. Such rheostats are expected to operate in other promoters. Their identification will be facilitated by short-range comparisons (e.g. human – chimp), and indeed this is a motive for extending comparative genomics.

The major histocompatibility complex (MHC) class II gene promoter provides the best understood example of *cis*-regulation. Having just determined the promoter sequence of three class II genes in each of thirteen strains of wild mouse in parallel, my intention here is to summarize the outcome against the background of their known function. The main interest lies in the light that these sequences cast on the new subject of comparative genomics. Following its triumph in man, the world-wide genome effort is now turning to mouse and chimp. Both species are logical choices, one because so much is known about its biochemistry and genetics, and the other because it is man's closest relative. Mice have the additional advantage that the laboratory mouse has a wide range of variant strains, as well as many related wild species. Exploring this genetic variation provides a rehearsal of what is likely to emerge from the main man – chimp – mouse effort. It confirms the view that close-up comparisons – notably between man and chimp – are likely to tell us most about *cis*-regulatory sequences. It is polymorphism in these sequences that accounts for most of genetic variation in man (Carroll, 2000; Mitchison, 2000).

The working of the MHC class II promoter is summarized in Fig. 1. To initiate transcription, the eight transcription factor proteins shown in the figures have to assemble together into a complex that interacts with the transcription bubble. The bubble is itself composed of the assembly of proteins around polymerase II that enables the enzyme to transcribe DNA into RNA. The three regulatory factor X (RFX) proteins assemble together so as to bind to the conserved X1 DNA box, and also to the large CIITA protein that connects this assembly to the transcription bubble (Emery et al., 1996). Two other assemblies are also involved (Zhu et al., 2000). One is composed of CREB (the protein that binds to the CRE sequence, so called because it functions as a cyclic-AMP response element) bound to CBP (CREB-binding protein) that connects this assembly to the transcription bubble (De Cesare et al., 1999; Shaywitz and Greenberg, 1999; Haus-Seuffert and Meisterernst, 2000). The CRE sequence is an octomer of cannonical sequence tgacgtca that varies slightly between different MHC II genes, the other is composed of the three nuclear factor Y (NFY) proteins shown, which together bind to the ccaat pentamer in the H2Eb promoter (Linhoff et al., 1997; Caretti et al., 2000). The X1, X2 and Y boxes were originally defined by their conservation in vertebrate evolution (Benoist and Mathis, 1990), and were thought to correspond to the binding sites of these three assemblies. Recent gel-shift analyses have broadly confirmed this view, although the relationship of the biochemical data to the previously defined Y box is unclear. Although the composition of each of the three assemblies is known, their arrangement and role in the final assembly at the bubble is unclear. The folded-up configuration shown in Fig. 1 is therefore speculative.

The variation indicated by arrows in Fig. 1 is taken from the unpublished wild-mouse parallel sequences mentioned above. The study was prompted by previous work on laboratory mice that had identified sites in and

Received May 15, 2001.

This work was supported by the Leverhulme and Wellcome Trusts.

Corresponding author: N. Avrion Mitchison, Windeyer Institute, University College London, 46 Cleveland Street, W1T 4JF, U.K.

Abbreviations: CRE – cyclic-AMP response element, CREB – protein binding to CRE, MHC – major histocompatibility complex, NFY – nuclear factor Y, RFX – regulatory factor X.

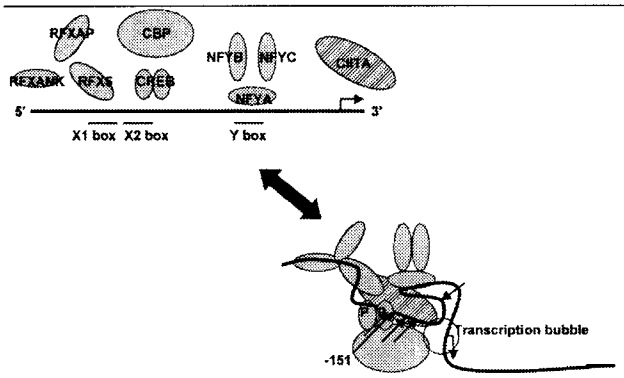

Fig. 1. The MHC class II promoter, inactive (upper left) and activated (lower right). The transcription factor proteins are shown in diagramatic shape (although the sizes are approximately correct), as are the DNA loops and the packing shown in the activated form. P indicates the key phosphorylation sites in CREB. Arrows indicate four sites of most conspicuous polymorphism and evolutionary variation, with the polymorphism at position –151 in $H2Ab^b$ shown.

around the X2 box as prone to vary (Cowell et al., 1998). Once it was discovered that the X2-binding protein was in fact CREB (Moreno et al., 1999), further testing of what we began to call the *circum/intra*-CRE hypothesis was called for. The new study confirmed and extended the previous findings. The four sites of variation found in wild mice and shown in the figure do indeed show the expected distribution in and around the CRE octomer. Furthermore, variation previously reported in the human DRβ promoter conforms to this pattern of CRE-related variation (Cowell et al., 1998). A polymorphism near CRE has also been reported in the human IL-6 promoter, which affects not only the level of expression, but also susceptibility to systemic-onset juvenile idiopathic arthritis (Fishman et al., 1998). It makes sense to find some of the variation around rather than within the regulatory element itself, as that should alter the level of transcription without switching it off entirely. Unlike the RFX-binding sequences, both CRE and the NFY-binding sequence are widely used elsewhere in the genome – CRE notably in the neuropeptide genes that are important in memory and other brain functions. I hypothesize that the *cis*-regulatory elements that are unique to one locus tend to serve as on-off binary switches (although switching by the RFX proteins seems more leaky than had been supposed (Williams et al., 1998)). The more widespread elements serve as rheostats that vary the level of expression under the influence of neighbouring sequences (I thank Adam Lacy-Hulbert for suggesting this nomenclature). The weakness of this hypothesis is that the effect of neighbouring variation is poorly understood. Is it possible that this editorial might encourage others to investigate this possibility?

Selection or hitchhiking?

Polymorphism in a transcriptional unit may reflect natural selection, or alternatively may have accumulated as more-or-less non-functional junk. It is an attractive hypothesis that heterozygosity at a *cis*-regulatory element may confer selective advantage by providing advantageous flexibility of expression (Villard et al., 1996; Mitchison, 2000). For genes with highly polymorphic coding sequences, such as the classical MHC genes, the *cis*-regulatory polymorphism might result from hitchhiking, i.e. be derived from linkage disequilibrium with the coding variants. This possibility receives support from the finding of high variation that extends far (~20 kb) upstream of an HLA DQ gene (Horton et al., 1998; Beck and Trowsdale, 2000)). That is much further than the extent of the classical transcription unit, although effects as far as 4 kb upstream have been identified in the mouse MHC and enhancers can operate over long distances (van Ewijk et al., 1988; Carson and Wiles, 1993). The large number of uniformly distributed single nucleotide changes found upstream of HLA DQ certainly argues against natural selection. The balance of evidence is summarized in Table 1. The positive controls entry refers to the relatively low level of polymorphism within the *cis*-regulatory unit of MHC class I genes, particularly around their CRE sequence (the alpha-site) found in our

Table 1. Hitchhiking or natural selection?

Type of evidence	Data supporting hitchhiking	Data supporting natural selection
Positive controls (level of upstream variation in other loci with equally variable coding sequences)	√	
Negative controls (level of upstream variation in other loci lacking variable coding sequences)	√	√
Meaningful distribution of upstream variation	√	√
Upstream variation affects transcription	√	
Upstream variation affects expression	√	
Upstream variation affects disease susceptibility	√	
Upstream variation investigated by gene knocking		Not done

unpublished data. The meaningful distribution entry refers to the unpublished data shown here in Fig. 1 and previously reported (Hesse et al., 1996; Janitz et al., 1997; Baumgart et al., 1998). Effects on transcription, expression and disease susceptibility are known in detail for only a tiny sample of the variation, namely the single nucleotide substitution between $H2Ab^b$ and other H2Ab alleles, located in CRE, at position –151 as shown in Fig. 1. Clearly, the final answer is not yet in; hence the need for the taxing experiment mentioned in the bottom line of the table.

The case for natural selection by no means implies that diversity of MHC II *cis*-regulatory sequences has evolved independently of the diversity of coding sequences. Rather, one imagine that a sequence that favoured, say, Th2 differentiation would associate with MHC molecules able to present worm epitopes, while one that favoured Th1 differentiation would associate with molecules able to do likewise for internal virus epitopes. With this sort of scenario neither promoter nor coding sequence could be said to occupy the driving seat. Rather, the two would have co-evolved together.

Shrinking biology into molecular genetics: the Prague connection

Over the years my group and the group in Prague led by Říhová have both been investigating a particular MHC-associated difference between C57BL and other mouse strains. Between us we have published some twenty papers on the serology, and on the antigen-presentation involved with its effect on T cells (Říhová, 1995; Mitchison, 2000), and the lead paper appeared even earlier (Silver et al., 1972). To me it now seems likely that all these biological effects stem from the single nucleotide substitution at position –151 mentioned above. There is nothing unusual about small changes in DNA making sense of a wide swathe of clinical and biological observations, and of course this does not mean that detailed characterization of the phenotype is not worthwhile. Still, it's remarkable how long it took for the penny to drop.

The role of comparative genomics

These data from parallel sequencing strengthen the case for further genome sequencing. The argument is summarized in Fig. 2. Future comparison between the mouse and human genomes are expected to prove informative about structural genes and the location of *cis*-regulatory elements, but less so about variation around these elements. If the argument advanced here holds good, the real pay-off will come in the form *circum*-element variation data. The name of the game is to spot the rheostats as defined by their pattern of variation. The mouse – human comparison cannot be expected to help much, because of the accumulation of junk. That is where the human – chimp comparison comes in, with its lower expected level of junk. Well, we shall have to wait and see, but perhaps not for very long. Spotting rheostats is important for understanding the networks that control gene expression. It is also potentially valuable in its applications, since each rheostat flags a potential target for pharmaceutical intervention. Rheostats are where natural selection has chosen to intervene, and where we would do well to follow.

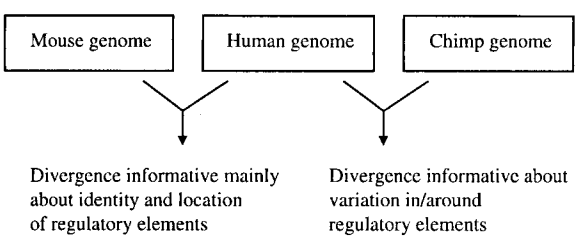

Fig. 2. A view of the future of comparative genomics

Acknowledgement

I thank Jurgen Roes for expert help and unswerving encouragement.

References

Baumgart, M., Moos, V., Schuhbauer, D., Muller, B. (1998) Differential expression of major histocompatibility complex class II genes on murine macrophages associated with T cell cytokine profile and protective/suppressive effects. *Proc. Natl. Acad. Sci. USA* **95,** 6936-6940.

Beck, S., Trowsdale, J. (2000) The human major histocompatibility complex: lessons from the DNA sequence. *Annu. Rev. Genomics Hum. Gene* **1,** 117-1137.

Benoist, C., Mathis, D. (1990) Regulation of major histocompatibility complex class-II genes: X, Y and other letters of the alphabet. *Annu. Rev. Immunol.* **8,** 681-715.

Caretti, G., Cocchiarella, F., Sidoli, C., Villard, J., Peretti, M., Reith, W., Mantovani, R. (2000) Dissection of functional NF-Y-RFX cooperative interactions on the MHC class II Ea promoter. *J. Mol. Biol.* **302,** 539-552.

Carroll, S. B. (2000) Endless forms: the evolution of gene regulation and morphological diversity. *Cell* **101,** 577-580.

Carson, S., Wiles, M. V. (1993) Far upstream regions of class II MHC Ea are necessary for position-independent, copy-dependent expression of Ea transgene. *Nucleic Acids Res.* **21,** 2065-2072.

Cowell, L. G., Kepler, T. B., Janitz, M., Lauster, R., Mitchison, N. A. (1998) The distribution of variation in regulatory gene segments, as present in MHC class II promoters. *Genome Res.* **8,** 124-134.

De Cesare, D., Fimia, G. M., Sassone-Corsi, P. (1999) Signaling routes to CREM and CREB: plasticity in transcriptional activation. *Trends Biochem. Sci.* **24,** 281-285.

Emery, P., Durand, B., Mach, B., Reith, W. (1996) RFX proteins, a novel family of DNA binding proteins conserved in the eukaryotic kingdom. *Nucleic Acids Res.* **24,** 803-807.

Fishman, D., Faulds, G., Jeffery, R., Mohamed-Ali, V., Yudkin, J. S., Humphries, S., Woo, P. (1998) The effect of novel polymorphisms in the interleukin-6 (IL-6) gene on IL-6 transcription and plasma IL-6 levels, and an association with systemic-onset juvenile chronic arthritis. *J. Clin. Invest.* **102,** 1369-1376.

Haus-Seuffert, P., Meisterernst, M. (2000) Mechanisms of transcriptional activation of cAMP-responsive element-binding protein CREB. *Mol. Cell. Biochem.* **212,** 5-9.

Hesse, M., Bayrak, S., Mitchison, A. (1996) Protective major histocompatibility complex genes and the role of interleukin-4 in collagen-induced arthritis. *Eur. J. Immunol.* **26,** 3234-3237.

Horton, R., Niblett, D., Milne, S., Palmer, S., Tubby, B., Trowsdale, J., Beck, S. (1998) Large-scale sequence comparisons reveal unusually high levels of variation in the HLA-DQB1 locus in the class II region of the human MHC. *J. Mol. Biol.* **282,** 71-97.

Janitz, M., Mitchison, A., Reiners-Schramm, L., Lauster, R. (1997) Polymorphic MHC class II promoters exhibit distinct expression pattern in various antigen-presenting cell lines. *Tissue Antigens* **49,** 99-106.

Linhoff, M. W., Wright, K. L., Ting, J. P. (1997) CCAAT-binding factor NF-Y and RFX are required for in vivo assembly of a nucleoprotein complex that spans 250 base pairs: the invariant chain promoter as a model. *Mol. Cell. Biol.* **17,** 4589-4596.

Mitchison, N. A. (2000) Polymorphism in regulatory gene sequences. *Genome Biol.* **2,** 1-6.

Mitchison, N. A., Muller, B., Segal, R. M. (2000) Natural variation in immune responsiveness, with special reference to immunodeficiency and promoter polymorphism in class II MHC genes. *Hum. Immunol.* **61,** 177-181.

Moreno, C. S., Beresford, G. W., Louis-Plence, P., Morris, A. C., Boss, J. M. (1999) CREB regulates MHC class II expression in a CIITA-dependent manner. *Immunity* **10,** 143-151.

Říhová, B. (1995) Presentation of antigen and MHC-restricted unresponsiveness. *Folia Biol. (Praha)* **41,** 140-162.

Shaywitz, A. J., Greenberg, M. E. (1999) CREB: a stimulus-induced transcription factor activated by a diverse array of extracellular signals. *Annu. Rev. Biochem.* **68,** 821-861.

Silver, D. M., McKenzie, I. F., Winn, H. J. (1972) Variations in the responses of C57BL-10J and A-J mice to sheep red blood cells. *J. Exp. Med.* **136,** 1063-1071.

van Ewijk, W., Ron, Y., Monaco, J., Kappler, J., Marrack, P., Le Meur, M., Gerlinger, P., Durand, B., Benoist, C., Mathis, D. (1988) Compartmentalization of MHC class II gene expression in transgenic mice. *Cell* **53,** 357-370.

Villard, E., Tiret, L., Visvikis, S., Rakotovao, R., Cambien, F., Soubrier, F. (1996) Identification of new polymorphisms of the angiotensin I-converting enzyme (ACE) gene, and study of their relationship to plasma ACE levels by two-QTL segregation-linkage analysis. *Am. J. Hum. Genet.* **58,** 1268-1278.

Williams, G. S., Malin, M., Vremec, D., Chang, C. H., Boyd, R., Benoist, C., Mathis, D. (1998) Mice lacking the transcription factor CIITA – a second look. *Int. Immunol.* **10,** 1957-1967.

Zhu, X. S., Linhoff, M. W., Li, G., Chin, K. C., Maity, S. N., Ting, J. P. (2000) Transcriptional scaffold: CIITA interacts with NF-Y, RFX, and CREB to cause stereospecific regulation of the class II major histocompatibility complex promoter. *Mol. Cell. Biol.* **20,** 6051-6061.