the virus-coded nuclear antigen EBNA, do not appear to do so.

This kind of result, reminiscent of those obtained with rodent cells transfected by fragments of herpes simplex virus DNA[10], raises more problems than it solves. The important issue remains the relevance of this phenomenon to the part which EBV might play in the pathogenesis of nasopharyngeal carcinoma and, as Griffin and Karran intimate, this should become clearer as their work and that of others extends to include the transfection of human epithelial cells. If not the virus, then at least its students, are losing their B-lymphotropic tag. □

Alan Rickinson is at the Cancer Research Campaign Laboratories, The Medical School, University of Birmingham, Birmingham B15 2TJ.

Molecular biology

New twist to DNA methylation

from Edward E. Max

On average, the dinucleotide cytidine-guanosine (CpG) is present in vertebrate DNA at a frequency only one-quarter of that predicted by random distribution. This striking 'CpG suppression' was puzzling at first but has since been partially explained[1]. A recent analysis[2] of published vertebrate DNA sequences in our laboratory has added a new puzzle — why do some genes contain localized segments that are CpG-rich regional lapses in the usual CpG suppression?

The partial explanation of CpG suppression involves DNA methylation. Most CpG dinucleotides in vertebrate DNA are methylated as mCpG, in which the mC is 5-methyl cytidine. Because 5-methyl cytidine has a propensity to deaminate to thymidine, CpG dinucleotides in vertebrate DNA tend, over evolutionary time, to be replaced by TpG, the dinucleotide that results from the deamination of mCpG and the complementary CpA (on the other strand). The correlation between methylation, CpG suppression and TpG + CpA elevation in some species was originally pointed out by Bird[1]. While the deamination model explains a possible mechanism for CpG suppression in terms of CpG methylation, it leaves open the question of the function of DNA methylation in vertebrates. Many experiments suggest that this function is somehow related to the control of gene expression[3].

Our first observation of localized CpG-rich regions within genes came from an analysis of the class I genes of the major histocompatibility complex (MHC). These genes encode highly polymorphic antigens that are major determinants of graft rejection. CpG suppression is absent from all available class I gene nucleotide sequences upstream of the border between exon 3 and intron 3; but normal CpG suppression is found downstream of this border. This is not simply a result of asymmetric frequencies of C or G in these regions; in the *H-2K*^{d} gene, for example, the ratio of CpGs observed to that expected on the basis of C and G content is about 0.06 downstream of the border between exon 3 and intron 3, but upstream it is 0.96 — that is, essentially no CpG suppression.

Our analysis revealed similar asymmetries in five mouse, one human and two rabbit class I genes. We also found a CpG-rich region surrounding exon 2 of the genes encoding the β-chain of class II MHC antigens, which are found on B lymphocytes and macrophages and regulate cellular interactions in the immune system. Of the many other genes examined by us and by others[4], only a few contain CpG-rich regions: the chicken α2-collagen[5], hamster adenine phosphoribosyl transferase (Lowy, I. and Axel, R., personal communication) and mouse dihydrofolate reductase[5,6] genes are three examples.

How do the CpG clusters escape the usual CpG suppression of vertebrate DNA? If CpG suppression requires mCpG as an intermediate, then a simple model to explain CpG clusters would be that these regions of DNA are maintained substantially free of mCpG, that is, are undermethylated, in germ-line DNA. This model predicts that the CpG-rich regions will not show the elevated TpG + CpA levels that result from mCpG deamination, which is precisely what we find. (An alternative model, in which unknown mechanisms generate excess CpG with normal methylation and mCpG deamination, predicts elevated TpG + CpA frequencies.) In the *H-2K*^{d} gene, for example, the ratio of observed TpG + CpA frequencies to those expected is approximately 1.0 in the CpG-rich 5′ region, and about 1.4 in the CpG-suppressed 3′ region. Thus, the TpG + CpA frequencies in the CpG-rich regions are consistent with our model of regional germ-line undermethylation.

An experimental test of the proposed model would be to examine directly the methylation state of the CpG-rich regions in germ-line DNA, such as that isolated from sperm. As discussed previously in these columns[7], methylation of the small fraction of CpG dinucleotides that fall within the recognition sequences of certain restriction enzymes (for example, *Hpa*II, *Msp*I) can be assessed using Southern blot techniques. The multiplicity of sequences cross-hybridizing to MHC probes complicates such an analysis for MHC genes, but the relevant data already exist for three non-MHC genes that contain CpG-rich regions. The 5′ regions corresponding to CpG-rich segments of the mouse dihydrofolate reductase gene[8], the hamster adenine phosphoribosyl transferase gene[8] and the chicken α2-collagen gene[9] are, indeed, undermethylated in sperm DNA, whereas the 3′ regions are methylated; similar patterns were found for these genes in all other tissues examined. Thus, available data seem consistent with the hypothesis that at least some of the CpG-rich regions are associated with DNA segments that remain undermethylated in the germ-line genome.

Should our model relating CpG-richness and germ-line undermethylation prove correct, it would still represent only the first step towards an explanation of these regions. One would also want to know what features of these regions spare them from the usual methylation of their CpG residues, and what functional significance of the CpG-richness of these regions accounts for their evolutionary preservation. Regarding function, we might first ask whether it is the CpG-richness itself that has critical functional significance, with the undermethylated state perhaps serving only to protect the CpGs from rapid mutation via the mCpG deamination pathway; or conversely, is the undermethylation the critical feature of these regions — perhaps mediating active expression of these genes in sperm cells — with CpG preservation an inconsequential byproduct? Furthermore, is the significance of the CpG-richness the same in MHC and non-MHC genes? It is striking that for both class I and class II MHC genes, the protein domains that exhibit the greatest polymorphisms are the ones encoded in CpG-rich regions; is this more than coincidence?

Given the questions about DNA methylation that remain unanswered in other more thoroughly studied systems, it is evident that additional information about CpG-rich regions and germ-line undermethylation in many genes will be necessary before the full significance of these observations will be understood. Meanwhile, it is worth noting that interesting features can still be gleaned from pure 'sequence gazing' (a pejorative term in some quarters), even from gene sequences as intensively studied as those of the MHC. □

1. Bird, A.P. *Nucleic Acids Res.* **8**, 1499 (1980).
2. Tykocinski, M.L. & Max, E.E. *Nucleic Acids Res.* **12**, 4385 (1984).
3. Bird, A.P. *Nature* **307**, 503 (1984).
4. Smith, T.F., Waterman, M.S. & Sadler, J.R. *Nucleic Acids Res.* **11**, 2205 (1983).
5. Vogeli, G. *et al. Proc. natn. Acad. Sci. U.S.A.* **78**, 5334 (1981).
6. Crouse, G.F., Simonsen, C.C., McEwan, R.N. & Schimke, R.T. *J. biol. Chem.* **257**, 7887 (1982).
7. Felsenfeld, G. & McGhee, J. *Nature* **296**, 602 (1982).
8. Stein, R., Sciaky-Gallili, N., Razin, A. & Cedar, H. *Proc. natn. Acad. Sci. U.S.A.* **80**, 2422 (1983).
9. McKeon, C., Ohkubo, H., Pastan, I. & de Crombrugghe, B. *Cell* **29**, 203 (1982).

Edward E. Max is at the National Institute of Allergy and Infectious Diseases, NIH, Bethesda, Maryland 20205.