DEVELOPMENTAL GENETICS

Living with bad architecture

Robin Lovell-Badge

PAPERS in *Nature Genetics*[1] and *Cell*[2], and another on page 771 of this issue[3], provide a common link between two very different forms of aberrant growth — benign tumours in humans and the pygmy phenotype in mice. It emerges that both conditions stem from mutations in a gene encoding a so-called architectural transcription factor known as HMGI-C.

In the cell, DNA is packaged into chromatin by histone and non-histone proteins Chromatin structure varies during the cell cycle and cell differentiation, and with gene expression, and these changes are under the control of architectural transcription factors[4]. Particular attention has centred on the HMG family of proteins[5], which get their name (high-mobility group; see box) from their small size in gel electrophoresis compared with other non-histone proteins associated with chromatin. They lack activation domains, have little if any sequence specificity in binding, and generally differ from classical transcription factors; instead they are thought to be involved in organizing chromatin at a local level to provide the correct architecture for other transcription factors and the basal transcription machinery to operate[6].

Human defects

HMG proteins are in fact very diverse, and HMGI-C forms a subgroup of its own with one other member, HMGI-Y. Both bind DNA sequences rich in adenine and thymine (AT) by domains termed AT hooks, and are fairly ubiquitous during embryonic development but largely absent in adult tissues. Given what has seemed to

be the somewhat general function of HMGI-C, it is surprising that three new studies[1-3] should come up with the protein as the cause of two highly specific but very different growth defects.

Schoenmakers *et al.*[1] and Ashar *et al.*[2] were searching for a gene that could be affected by translocation breakpoints mapping to human chromosome region 12q15. These are often associated with a variety of benign tumours, lipomas in particular. Schoenmakers *et al.* used positional cloning to search within a previously defined 1.7-megabase 'multiple aberration region' (MAR), which was thought to contain the tumour-associated gene, and found *HMGI-C*. Ashar *et al.* had already considered *HMGI-C* as a reasonable candidate for the tumour-associated gene: they knew that it mapped to the MAR; it is required for retrovirally induced neoplastic transformation[7]; and transcription factors have been identified at breakpoints in a variety of tumours[8].

Both groups determined the gene's organization. The first three exons each encode an AT-hook-motif DNA-binding domain, and the last two encode an acidic region. Importantly, there is a large intron of over 25 kilobases in length between exons 3 and 4 within which most translocation breakpoints had occurred.

More cases need to be analysed, but the common theme emerging is that the cause of the lipomas and other benign tumours is the generation of fusion proteins in which the AT-hook DNA-binding domains of HMGI-C are linked to a regulatory domain from another gene in the translocation partner chromosome. In one case

the latter was a highly acidic, serine-and-threonine-rich domain from a new gene, which could be turning HMGI-C into a transcriptional activator. In a second lipoma studied by both groups, the RNA was found to stem from the first three exons of HMGI-C and two tandemly arrayed LIM domains; these are involved in protein–protein interactions, and could recruit other transcriptional activators to the DNA sites bound by the AT hooks of HMGI-C. The benign nature of the tumours needs an explanation, but HMGI-C is expressed at very low levels. Also, why do they tend to be lipomas? Perhaps HMGI-C is particularly important in adipose tissue, which brings us to the second part of the story.

Mouse defects

Zhou, Chada and co-workers[3] were interested in the cause of the *pygmy* mutation in mice. Homozygotes for *pygmy* have adult body weights that are 40 per cent, and heterozygotes 80 per cent, of those of their wild-type littermates. Chada and colleagues had previously identified a new allele at *pygmy* by transgenic insertion[9]. In the latest study[3], a flanking marker from the insertion site was cloned and used to initiate chromosome walks. The authors identified a common region of fifty-six kilobases that was deleted in the mutant alleles, and exon amplification was used to identify a gene within the deletion — the gene turned out to be *Hmgi-c*. Because other genes could have been affected by the deletions, the authors used gene targeting by homologous recombination to create a much more specific null mutation in the gene, thus confirming that the pygmy phenotype results from lack of HMGI-C.

How is this phenotype explained? Analysis of *Hmgi-c* expression showed

---

The HMG family of proteins

THE HMG proteins are a collection of small acidic proteins that were first described many years ago with the advent of gel electrophoresis techniques for separating macromolecules. Their name, which stands for high-mobility group (and not Her Majesty's Government) proteins, simply reflects their small size (all have relative molecular masses under 30,000) compared with other non-histone chromatin-associated proteins — a terminology that conceals the diverse structure and function of these proteins.

One class is represented by proteins such as HMG-1, which bind bent DNA structures irrespective of sequence[15]. This binding occurs through several domains, first identified by comparison of HMG-1 with UBF, a protein involved in regulating the expression of genes encod-

ing ribosomal RNA[16]. These domains were termed HMG boxes, which unfortunately is rather confusing as not all HMG proteins have them and because similar domains are found in many other proteins that fall outside the original definition. Many of these, including factors important for decisions of cell fate in development, such as LEF-1, TCF-1, SRY and the related SOX proteins, have single HMG-box domains, with the additional property of sequence-specific binding to linear DNA.

On binding, these proteins cause DNA to bend through large angles, which may facilitate the action of transcription factors bound nearby[6,17] For this reason they can be considered as architectural proteins, although some also have strong activation domains and may contribute

directly to the activation of target genes[18].

A second class of HMG proteins is represented by HMG-14 and HMG-17 (refs 4–6). Like HMG-1, these do not appear to activate or repress transcription of other genes directly, but they have different DNA-binding domains (random-coil secondary structure) and show affinity for DNA within nucleosomes of active chromatin. HMGI-C and HMGI-Y represent a third class and use domains termed AT hooks to bind A+T-rich DNA sequences associated with many genes[4-6]. Again, they are thought not to play a direct role in transcriptional activation, but participate in organizing complexes of other transcription factors, which in turn recruit the basal transcriptional machinery.

R. L.-B.