Rather like those of Albert Einstein, DNA's popular images are hardly representative. While it is fashionable in these post-genome days to show it as an endless string of A's, C's, G's and T's, this year's anniversary will surely be illustrated with two kinds of picture. One shows the famous double helix, delightfully suggesting the twin snakes of Wisdom and Knowledge intertwining around the caduceus, the staff of the medic's god Hermes. The other reveals the X-shaped symbol of inheritance, the chromosome.

But it is rare that DNA looks this good. For a couple of hours during the prophase and metaphase of the cell cycle, as the cell prepares for mitosis, the genome is compacted into its distinctive chromosomal fragments. The rest of the time you will search the eukaryotic cell in vain for those molecular tetrapods. What you find instead in the cell nucleus is, apparently, a tangled mess.

And don't think that this will, on closer inspection, turn out to be woven from a double helix as elegant as that in publicity shots. Rather, the threads are chromatin—a filamentary assembly of DNA and proteins—in which only very short stretches of the unadulterated helix are fleetingly revealed. While the chromosomes are often equated with DNA, there is actually around twice as much protein as DNA in chromatin, as well as typically around 10 percent by mass of RNA—mostly as nascent transcribed chains.

Zooming in on DNA

If we want to know how DNA really functions, it is not enough to zoom in to the molecular level with its beautifully simple staircase of base pairs. Textbooks will, understandably, show replication as the steady progress of DNA polymerase along a linear single strand laid out like a railway line, and RNA polymerase doing likewise in transcription. One has the impression of the genome as a book lying open, waiting to be read.

It is not so simple, however. The book is closed up, sealed, and packed away. Moreover, the story is not simply on the pages of the book; these operations on DNA involve information transmission over many length scales. Perhaps those who do not routinely have to delve into the intricacies of genome function have acquired such a simplistic picture of it all because these length scales were largely considered out of bounds for molecular science until relatively recently. We know about molecules; we know about cells and organelles; but it is the stuff in between that is messy and mysterious.

We speak of molecular biology and cell biology, but no one really talks of mesobiology. Yet that is the level of magnification at which much of the action takes place: the scale of perhaps a few to several hundred nanometres. How DNA is arranged on these scales seems to be central to the processes of replication and transcription that we have come to

think of in terms of neat base pairings, yet it is precisely here that our understanding remains the most hazy.

Partly that's because the mesoscale represents, quite literally, a difficult middle ground. It encompasses too many atoms to be able to apply straightforward molecular mechanics, with its bond bending and breaking; yet the graininess still matters, the continuum has not yet become a good approximation. As Bustamante *et al*. show elsewhere in this issue, looking at DNA on a scale where it flexes and twists like a soft rod reveals how the mechanical and the molecular interact.

Take the problem of supercoiling, for example. The closed loops of bacterial DNA can develop twists like those in a Möbius strip, which either 'overwind' or 'underwind' the helix. Generally there is some degree of underwinding—negative supercoiling—such that there is one negative supercoil for every 200 base pairs (bp) or so. This has an energy cost of around -9 kcal mol^{-1}, which manifests itself in physiological effects. In bacteria, too much supercoiling can inhibit growth, which is why enzymes (topoisomerases) exist to release it. On the other hand, negative supercoiling tends to unwind the double helix, which might be useful to initiate strand separation for replication.

Although the chromosomal DNA of eukaryotes has free ends, it too is prone to supercoiling, since it appears to be typically attached in large loops to a filamentous structure called the nuclear matrix that coats the inside of the nuclear membrane. The attachment may in fact be necessary for both replication and transcription to take place.

Packaging problem

Stretched into a linear double helix, the four billion or so base pairs of human DNA would measure 1.8 m. This strand, snipped into 46 chromosomes, has to be packed into a nucleus just 6 μm or so across. As a result, the DNA chains are far from the idealized picture of molecules floating in an infinite solvent. They have a density of around 100 mg ml^{-1}, comparable to that of a highly viscous polymer gel.

The packing ratio for the chains is therefore enormous. In the smallest human chromosome, a length of DNA 14 mm long is compressed into a chromosome about 2 μm long: a packing ratio of 7,000. The first stage in solving this packaging problem is to wind the DNA around protein disks to form a bead-like nucleosome. Each disk is an octamer of four types of histone protein; a fifth histone, called H1, seals the DNA to the disk at the point where the winding starts and ends. Each nucleosome, 6 nm high by 11 nm in diameter, binds around 200 bp of DNA in two coils, and there is very little 'free' DNA between adjacent nucleosomes: sometimes as little as 8 bp.

The string of nucleosomes forms a fibre about 10 nm thick, which is then packaged into a filament three times as wide. This 30-nm fibre is the basic element of chromatin—yet we still don't know its structure. It is widely held to be composed of nucleosomes arranged in a solenoid, but hard evidence for this is scanty. How many celebrations of the double

helix will admit that, fifty years on, we don't really know what DNA at large in the cell looks like?

The 30-nm fibre is further folded and condensed to give a packing ratio of around 1,000 in interphase chromosomes, and around ten times that in the X-shaped mitotic chromosomes. How this happens is even more of a mystery. For mitotic chromosomes it was thought until only recently that there might be a contiguous protein scaffold holding the whole affair together; but now it seems that the structural integrity must come from chromatin crosslinking.[1] All the histones seem to have higher-order structural functions. Multi-subunit protein complexes in yeast called SWI/SNF and RSC (both of which seem to have human homologues) are chromatin-remodelling machines which use ATP hydrolysis to distort histone-DNA contacts or to transfer histones between DNA molecules, exposing the DNA to attack by nucleases. Quite how they work, or what they do to the nucleosomes, remains hazy.[2] According to one recent study,[3] DNA engaged by such complexes 'behaves as if it were free and bound at the same time.' Or in other words, as if 'free' and 'bound' were notions too simplistic to have much meaning here. What is clear is that these chromatin-shaping machines play an important part in transcription: cells lacking RSC are no longer viable. One way or another, there is a lot a mesoscale activity involved in this fundamental cell process.

There are in fact two types of chromatin in the nucleus of an interphase eukaryotic cell. Euchromatin is the most abundant: it is relatively dispersed and gel-like. Heterochromatin is much denser, comparable to the density of mitotic chromosomes, and is confined to a few small patches. The invitation is to regard euchromatin as 'active' DNA, unpacked enough to let the transcription apparatus get to work on it, while heterochromatin is compressed, like a big data file, until needed. But like just about any other generalization about DNA's structure and behaviour, this one quickly breaks down. Clearly only a small fraction of a cell's euchromatin is made up of transcribable DNA in the first place (so why not pack the rest away?); and even chromosomes containing a large amount of heterochromatin can be transcriptionally active. Some researchers think that 'euchromatin' is actually just a blanket term for many things we don't understand: further hierarchies of DNA organization yet to be revealed.

Structured chaos

Certainly, there seems to be more to the nucleus than a disorderly mass of DNA. It is a constantly changing structure, but not randomly: there is method in there somewhere. Specific chromosomes occupy discrete nuclear positions during interphase, and these positions can change in a deterministic way in response to changes in the cell's physiological state.

And the euchromatin itself has an internal logic, albeit one only partly decoded. It has been proposed that DNA has sequences called scaffold/matrix-attached regions (S/MARs), recurring typically every 10-100 kbp, that bind to the nuclear matrix to divide up the chromosome into loops.[2] Yet the existence of not only S/MARs but even the nuclear scaffold itself has been questioned. There is no sign of the scaffold during

mitosis, and the material it is thought to be composed of may be nothing more than a mess of denatured proteins.

Be that as it may, the organization of the loops seems to be important for compaction of DNA and for the regulation of transcription, and each loop may act as an independent unit of gene activity. In other words, there is at least one level of superstructural organization in the chromosomes that makes its influence felt at the level of molecular information transfer. Topoisomerase II is one of several proteins that bind specifically to the putative S/MARs, suggesting that these points are important for controlling supercoiling in the strands.

With all this high-level structure, transcription of DNA is not so much a matter of slotting the parts in place as tugging on the rope. DNA is highly curved around the nucleosomes, the inward-facing groove compressed and the outer one widened. (This bending is sequence-dependent: A•T-rich regions have the minor groove facing inwards, C•G regions place it on the outside.) RNA polymerase, at 13 by 14 nm, is about the same size as the nucleosome, yet it binds to a region of DNA around 50 bp long: about a quarter of the entire histone-bound length. So clearly some DNA must leave the surface of the histone core for transcription to proceed. But this core needn't be displaced completely. The histone disk actually has a considerable amount of mobility, sometimes described as a corkscrew motion through the DNA coil. The reality is undoubtedly more complex, involving a kind of diffusion of localized defects in the DNA-histone contact.

This mesoscale mechanical behaviour may be governed by chemical switches. Actively transcribing chromatin contains histones modifed by acetyl groups, which seems to perturb and loosen the chromatin structure. Histones also become phosphorylated at some stages in the cell cycle, which again affects their packing. The DNA itself is also chemically labelled for transcription. Between 2 and 7 percent of the cytosines in animal nuclear DNA aren't as one sees in classic illustrations of paired bases, but have methyl groups attached. (This proportion can be as great as 30 percent in some plants.) Typically methylation occurs in paired ^{Me}CG doublets on the twin strands. Active genes, meanwhile, are generally *under*-methylated: the removal of some methyls is a chemical signal that the gene is ready for transcription.

If all of this destroys the pretty illusion created by the iconic model of Watson and Crick, it surely also opens up a much richer panorama. The fundamental mechanism of information transfer in nucleic acids—complementary base pairing—is so elegant that it risks blinding us to the awesome sophistication of the process. These molecules do not simply wander up to one another and start talking. They must first be designated for that task, and must then file applications at various higher levels before permission is granted. For those who would like to control these processes, and those who seek to mimic them in artificial systems, the message is that the mesoscale, far from being a regime where order and simplicity descend into unpredictable chaos, has its own structures, logic, rules and regulatory mechanisms. This is the next frontier at which will unfold the continuing story of how DNA works.

1. Poirier, M. G. & Marko, J. F. *Proc. Natl Acad. Sci. USA* **99**, 15393-15397 (2002).
2. Van Driel, R. & Otte, A. P. (eds) *Nuclear Organization, Chromatin Structure, and Gene Expression* (Oxford University Press, 1997).
3. Asturias, F. J., Chung, W.-H., Kornberg, R. D. & Lorch, Y. *Proc. Natl Acad. Sci. USA* **99**, 13477-13480 (2002).