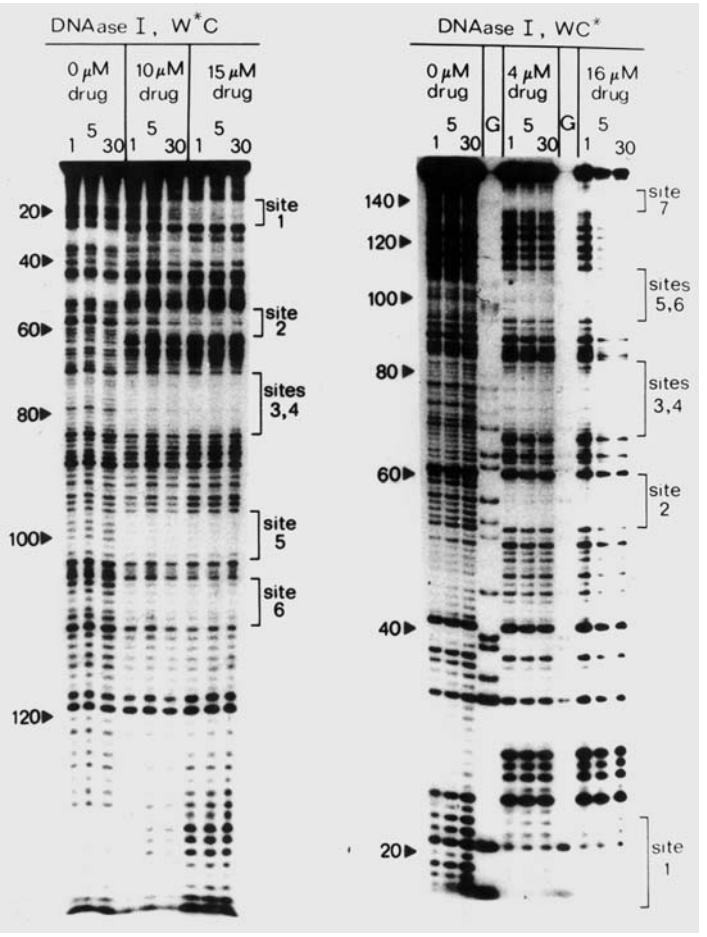

Figure 9.8 Use of a gel for 'footprinting' of the antibiotic echinomycin onto a DNA molecule of 200 base-pairs. The left- and right-hand sides of the figure show cutting by DNAase I on either of the two strands of the double helix, at different drug concentrations. Courtesy of Loretta Low and Michael Waring.

The effect of adding the antibiotic to DNA can be seen in the remaining six gel lanes on either side: different sets of lanes contain different concentrations of echinomycin. The lanes labeled 'G' are markers for guanine. One can easily identify the sites of binding of echinomycin to DNA, by looking for short regions within any lane of the gel where the bands are relatively faint, as compared with their intensities for free DNA. The antibiotic binds tightly to seven different locations along this 200-base-pair DNA, and blocks DNAase I cutting for 5 to 6 bonds at each of its binding sites. 'Site 1' is located near one end of the molecule, far from the radioactive end-label used on the left-hand side of the figure; so there it runs

slowly, near the top of the gel. But 'site 1' lies close to the radioactive end-label used on the right-hand side; so there it runs rapidly, near the bottom of the gel. The opposite holds true for 'site 7', which runs off the bottom of the gel on the left, but near the top of the gel on the right. If we were to compare the locations of these echinomycin binding sites with the base sequence of the DNA used in this experiment, we would find that each binding site is centered on a short sequence of the kind CG. After this experiment was published, it was found by X-ray diffraction methods that echinomycin binds at a step CG in its crystalline complex with DNA.

Near 'site 2' on the left-hand side, or 'site 1' on the right-hand side of Fig. 9.8, several bands adjacent to each echinomycin-binding-site actually become more intense in the presence of the antibiotic than in its absence. It turns out that these are AT-rich regions of high propeller twist, where DNAase I cuts only poorly in the free DNA. Once echinomycin binds next to such DNA, it flattens the propeller twist and lets DNAase I cut more rapidly. Other antibiotics such as ethidium bromide are also thought to flatten the propeller twist, so as to let DNAase I cut more rapidly.

Footprinting studies like the one shown can now be carried out even on single-copy DNA sequences within a whole chromosome, to see where different proteins bind along the DNA in living cells.

As a final example of useful enzymatic methods, let us look at the cutting activity of an enzyme called 'S1 nuclease'. This enzyme cuts DNA only where it can detect an unwound single-strand, as opposed to a double helix. Figure 9.9(a) shows two particular 20-nucleotide strands of DNA, called '20-Watson' and '20-Crick' respectively. When these two strands are mixed in equal amounts, they form a 20-base-pair double helix with Watson–Crick pairs as shown. But if the two strands are kept separate, each may fold back on itself to form a 'hairpin loop', with six base-pairs in the stem and eight unpaired bases in the loop. These hairpin loops are shown schematically in Fig. 9.9(b).

When the 20-base-pair double helix from part (a) is treated with S1 nuclease, and run through a denaturing gel, the cutting pattern shown on the right-hand side of Fig. 9.9(c) is obtained. In this experiment, the 3'-end of 20-Watson has a radioactive label, while the strand 20-Crick remains unlabeled and therefore 'invisible'. It seems that S1 nuclease can cut only at the very ends of a double helix, where the two single strands of DNA are not so firmly connected to one another. Yet when you treat the hairpin 20-Watson from Fig. 9.9(b) with S1 nuclease, you see the pattern shown on the left-hand side of Fig. 9.9(c). Once again the lowest part of the double helix is cut at its fraying end, but in addition you can see extensive

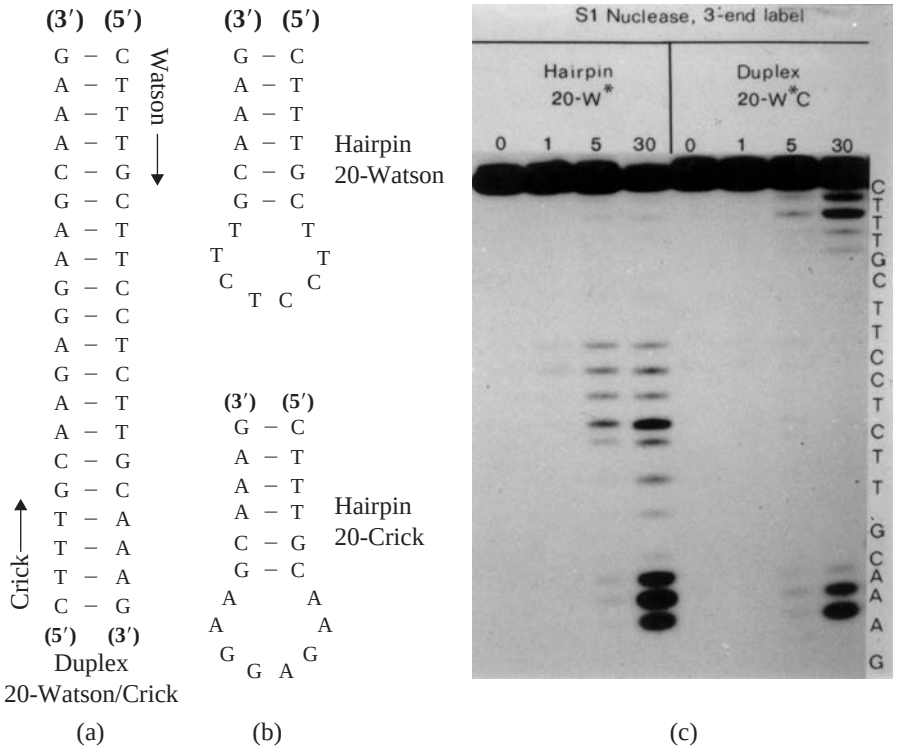

Figure 9.9 Investigation of hairpin-loops by means of cleavage with single-strand specific S1 nuclease, followed by gel electrophoresis of the cleavage products.

cutting of the strand within its loop of eight unpaired bases, halfway up the gel. This result shows conclusively that S1 nuclease recognizes the structure of the DNA rather than its sequence, for the same sequence TTCCTCTT is not cut when it is part of a double helix on the right.

In summary, one can cut the DNA by using a wide variety of enzymes and chemicals, in order to probe its structure. Some of these recognize the base sequence, while others recognize the double-helical structure, or lack thereof. In all such cases, the standard procedure is to separate the fragments of DNA so obtained by means of electrophoresis in gels, where each double helix or single strand runs according to its size.

Finally, although it is not widely known, it seems that the gels themselves can be used to find out something about DNA structure. It is not strictly true that molecules of DNA run in gels according to their *size*; rather, they run according to their *size*, *shape*, and *net electric charge*. For example, if a piece of negatively charged DNA is bound to a positively charged protein, it will have a reduced overall negative electric charge, and so it will run more slowly

through a gel than free DNA of the same length. That is easy enough to understand; but how can a molecule of DNA run in gels according to its shape? How can the gel fibers sense the shape of the DNA, in addition to its chain length or size? This result also follows in a trivial way, once you understand the theory of DNA motion through gels. We shall explain briefly how it works, without recourse to the difficult mathematics that are required to account precisely for the motion of DNA through gels.

Actually, there are two kinds of gel electrophoresis, that use *either constant* or *regularly alternating* electric fields. All of the examples we have given so far have been for the method of constant field. In this case, a sample of DNA is applied to one end of a gel near a negatively charged electrode (see Fig. 9.5), and this sample then migrates through the gel for a distance of 20 cm (or 50 cm in some cases) towards a positively charged electrode. The potential difference between the two electrodes is typically 100 to 1000 V, or enough to generate a current of 10 to 100 mA for a typical salt solution, at neutral pH. Most of the current between the two electrodes is carried by positively and negatively charged ions in the salt solution, such as sodium (Na^+) and chloride (Cl^-), which move through the gel in opposite directions owing to their different charges. The DNA, of course, is negatively charged, so it moves in the same direction as the chloride ions; that is, it moves towards the positively charged electrode. It is necessary to keep a solution of salt in the pores of the gel, because DNA in distilled water will not move at all; it remains bound to sodium ions that pull it in one direction, while it wants to go in the other. In the presence of excess salt, DNA can move towards the positively charged electrode by exchanging sodium ions continuously as it moves forward. (The most commonly-used salt solutions for gels actually include ions such as $Tris^+$ and $borate^-$ or $acetate^-$, which maintain a firmer control on the pH than do Na^+ and Cl^-.)

When this experiment is carried out in a vessel filled with simple salt solution, in the absence of any gel, all DNA molecules move toward the positively charged electrode with about the same velocity, regardless of their size or shape. Thus, a single nucleotide will move at the same speed as several thousand base-pairs of double helix. The explanation for this is that the larger DNA molecules contain more phosphates, and hence they have more electric charge to pull themselves forward in the electric field; but they also experience more viscous drag from their contacts with water molecules, because they are bigger. The two effects cancel, and so DNA moves through a simple salt solution at a speed independent of its size.

The purpose of a gel is to exert an even greater friction or drag on the larger DNA molecules, so as to cause them to move more slowly

than the smaller ones. Gels can be made from fibers of agarose or polyacrylamide, at concentrations of 0.5 to 20 gram of solid per 100 ml of fluid. Single nucleotides of DNA move through a gel at the same speed as they move free in solution, but large DNA molecules are retarded in proportion to their size by contacts with gel fibers. The gel method at constant field is capable of separating by size DNA molecules as long as 50 000 base-pairs, with a resolution of better than 1% of the size of the DNA. It is also capable of separating by size a wide variety of protein molecules, with a similar resolution.

There is thus no mystery as to how a gel works, in principle. But how exactly does a gel sense the size and shape of the DNA (or a protein), and retard its motion accordingly? To answer this question, we must first see what a gel looks like on a molecular scale. A typical agarose gel, as visualized by electron microscopy, is shown in Fig. 9.10. It contains many fibers of diameter about 100 Å that cross over one another like the strands of a grass mat. The most important property of an agarose gel is that its fibers are arranged *randomly in space*. Suppose that we could somehow make a gel from *regularly* arranged fibers or points, as shown schematically in Fig. 9.11(a). Then all of the passages through the gel would be of the same size, and hence a plot of particle velocity *versus* size would show a sharp cut-off where the particle (whether protein or DNA) becomes too large to go through any passage. But, as we have said, real gels are constructed from fibers that are arranged randomly in space, as shown in Fig. 9.11(b). Thus some passages through the gel are small, while others are large, and the speed of the particle will decrease gradually with increasing size. Scientists usually adjust

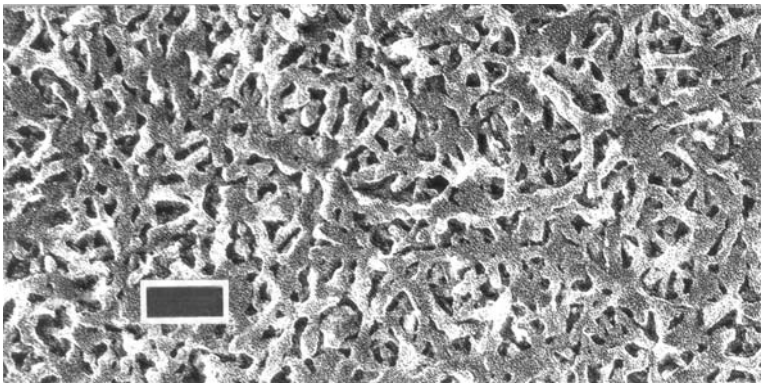

Figure 9.10 Electron micrograph of a portion of a 2% agarose gel, 1 mm × 0.5 mm overall: the small black rectangle is 1000 Å × 500 Å. Individual gel fibers are about 100 Å wide. Courtesy of Sue Whytock and John Finch.

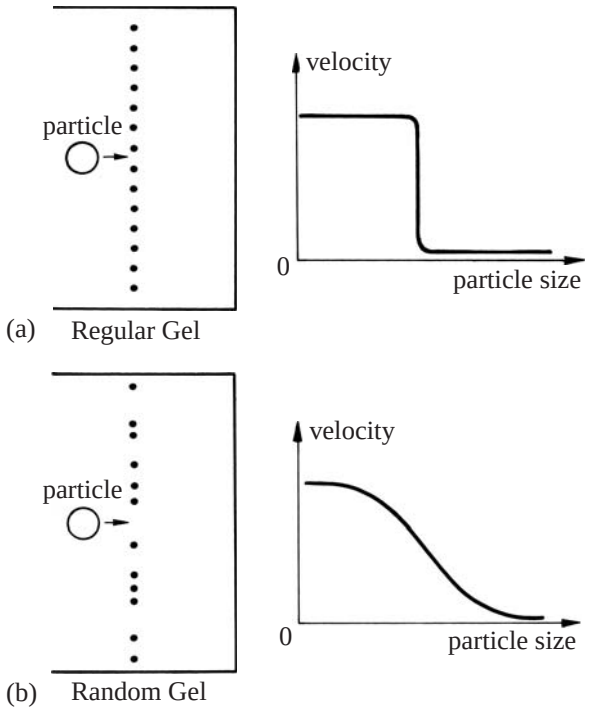

Figure 9.11 Sorting effect of (a) regular or (b) random-sieve gel.

the concentration of their gel so that the size of the molecule they wish to isolate lies about in the middle of the smooth curve shown in the diagram.

How does the DNA pick its way through a series of randomly sized passages in a gel? Obviously, there can be no regular pattern to its motion, and so its path is usually described through the use of statistics. There are several ways of thinking about this. One way is shown in Fig. 9.12(a), where a 'disc' must move from left to right across a plane field of randomly distributed point obstacles. In the free spaces it goes quickly, just like DNA free in solution; but where it overlaps one or more points it goes slowly, perhaps at only 5% or 10% of the speed of DNA in solution. The bigger the disc, the greater the fraction of its path where it contacts point obstacles: hence, big discs go more slowly through the field of obstacles than do small discs.

Another way of looking at the same thing is shown in Fig. 9.12(b). There we see cars traveling along a road, either singly or in caravans. The road is smooth for most of its length, permitting travel at 100 km/hr, but it is rough in parts, permitting travel at just 10 km/hr. A single car will average a speed of 92 km/hr along this road, if the rough patches amount to just 1% of the total road surface: recall that

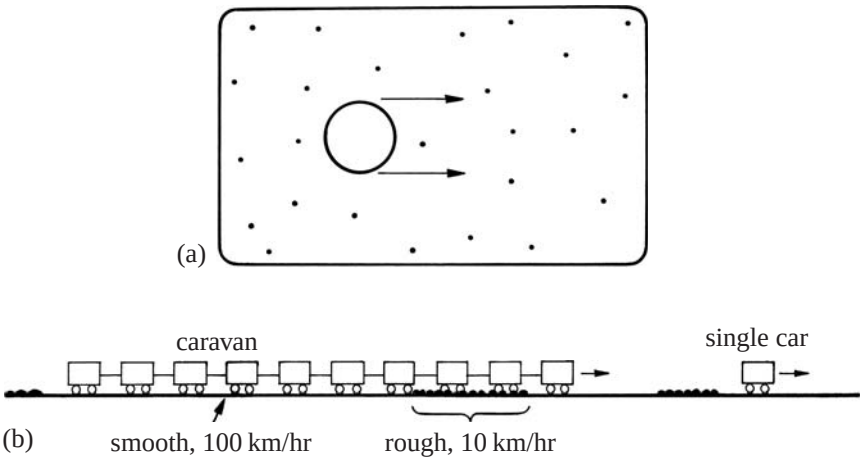

Figure 9.12 Two analogies for gel electrophoresis of DNA at constant field: (a) a disc passes through a two-dimensional 'gel' of randomly-located point obstacles; (b) a long caravan is impeded more than a single car by rough patches in a road.

average speed = (total distance)/(total time). But the long caravan will proceed at about 10 km/hr, because it always contacts a rough spot somewhere along its length. Both parts of Fig. 9.12 describe a process of non-uniform motion, where the DNA starts and stops a lot while going through a gel. It does not proceed steadily, but its speed is uniform when averaged over a long period of time – which is itself short in comparison with the total time of testing.

The probability that a DNA molecule will make contact with a gel fiber is given by statistical theory for several different hypothetical cases. For a gel made of long fibers, the DNA is slowed according to its surface area; while for a gel whose obstacles are points, such as the junctions between fibers, the DNA is slowed according to its volume. Many workers now agree that DNA is likely to be slowed in a gel according to its volume, as if the gel behaved like a set of points arranged randomly in space; but such fine points of gel theory are still not certain.

Many experiments show that *curved* DNA moves more slowly through gels than does *straight* DNA of the same size or length. From the discussion presented in Chapter 5, it seems obvious that curved DNA will occupy a larger volume, or have a larger effective surface area, than straight DNA of the same size, because it can coil around itself to include a lot of empty space into which the gel fibers cannot enter. Thus, from careful measurements by gel electrophoresis, one can estimate the increased volume or surface area of curved DNA compared with straight DNA. But one cannot learn anything definite about local structural parameters such as roll,

slide, or twist (see Chapter 3), because the gel does not sense such parameters directly.

Other experiments show that *supercoiled* DNA moves through gels with a velocity that depends on its linking number Lk. Some pictures of supercoiled DNA from bacteria were shown in Fig. 6.7. Circular DNA molecules with Lk near 0 move more slowly through gels than do molecules with a large value of Lk, either positive or negative, because the shape of the DNA becomes more compact as it gets more supercoiled. To be precise, it is thought that the cross-sectional diameter of an interwound supercoil decreases as Lk departs from zero. For example, in Fig. 6.5 the circle (a) with Lk = 0 has a larger cross-sectional diameter (or 'fatness') than circles such as (c) to (e) with Lk = −3. Given a mixture of supercoiled DNA molecules of identical size but different linking numbers Lk, one can determine the true linking number Lk of any one of them by running the total mixture through an agarose gel, and then counting the number of discrete bands from the slowest with Lk = 0 at the top of the gel to, say, Lk = −10 or −20 near the bottom (see Exercise 9.6).

In general, the method of gel electrophoresis as described above can separate DNA molecules of different length out to about 50 000 base-pairs. Above this length the DNA is never out of contact with gel fibers: it is like a very long caravan in Fig. 9.12(b), always impeded by rough patches on the road. Thus, all DNA molecules beyond a certain length travel at the same speed; and so the gel is of little use in separating them.

To separate very large DNA molecules according to their size, one can carry out gel electrophoresis in the presence of a regularly alternating electric field. Suppose that the voltage applied to a gel is reversed at regular intervals of time: how will that affect the motion of DNA through the gel? For example, the voltage from left to right in a gel may be set at +100 V for 10 s, then at −100 V for 5 s, then at +100 V again for 10 s, and so on. The DNA will move first in one direction for 10 s, then in the other direction for 5 s, then in the first direction again for 10 s. How will its net motion be altered? After an interval of 15 s, will it go simply $(10 - 5)/15 = 1/3$ as far as it would have gone at a constant 100 V? If the DNA is small, say less than 1000 base-pairs, then periodic reversal of the voltage has precisely this effect: the DNA goes 1/3 as far as it would have gone at constant voltage. But if the DNA is very large, say 100 000 to 10 million base-pairs, then periodic reversal of the field has a dramatic and unexpected effect: the large DNA cannot 'turn around' in the gel as quickly as small DNA, when the voltage is reversed, and so the large DNA proceeds much more slowly than expected. In other

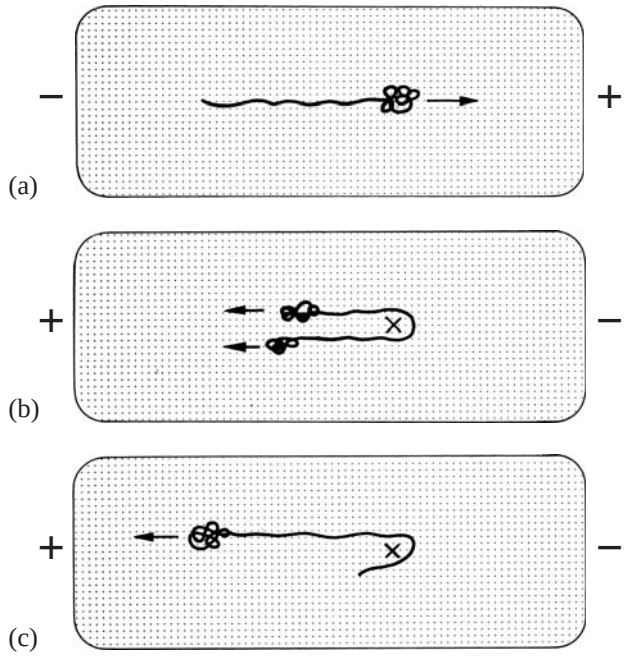

Figure 9.13 Movement of a very long DNA molecule through a gel, in alternating-field electrophoresis: (a) forward motion; (b) 'tie on coathook' situation; (c) reverse motion.

words, there is a 'time delay' for the large DNA to change direction in the gel; and consequently it does not go as far through the gel as expected. In the absence of this effect, all of the very large DNA molecules would go through a gel at the same speed.

Several approximate models for the behavior of long DNA in gels with an alternating electric field have been described by Carlos Bustamante, Bruno Zimm, and others. One possible mechanism of this time-delay is shown schematically in Fig. 9.13. During the first 10 s at +100 V, the DNA proceeds from left to right at the expected speed. Studies by light microscopy show that the long DNA is sperm-shaped, with most of its mass concentrated into a 'head', while a small part follows as a 'tail'. The large head of DNA must force its passage between gel fibers, like an icebreaker or bulldozer, because there are few if any pre-existing passages in the gel large enough to accommodate it. This kind of motion of DNA through the gel does not depend much on DNA size. But when the voltage is reversed, to −100 V, the DNA must reorganize itself to create a new 'head' in the opposing direction. For a brief interval, the DNA is suspended over the gel fibers in motionless equilibrium, like a tie over a coat-hook. After this brief interval or 'time-delay', the DNA falls from its

unstable position and forms a 'head' that can proceed in the reverse direction. The time required for the DNA to 'fall off' the coat-hook depends quite strongly on DNA size. The same time-delay will occur when the voltage is switched back to +100 V. This interesting and unexpected behavior of large DNA in gels was discovered by David Schwartz and Charles Cantor in 1984. It enables one to separate in size, by means of gel electrophoresis, DNA to a length of 5 or 10 million base-pairs. Thus, one can separate in a gel all the chromosomal DNA molecules from a simple organism such as yeast, where the chromosomes do not exceed 10 million base-pairs. But one cannot yet separate by this method the chromosomal DNA molecules from a human being, which are on the order of 50 to 400 million base-pairs in length.

The key to the separation of large DNA molecules in gels, by this method, is to adjust the times for the forward and reverse pulses of voltage to be of the same order as the time to 'turn around', for any given length of DNA. To a first approximation, the time required for a DNA molecule to 'turn around' in the gel increases as the 2/3 power of its size. This is not hard to understand in principle: short ties fall from coat-hooks much more frequently than long ties.

One can also understand why DNA larger than 5 to 10 million base-pairs cannot be separated in size: when a tie gets very long, say the distance from your floor to your ceiling, then the time required for it to fall from a coat-hook, by random vibration, will not depend very much on its length. Often one can improve the separation of such DNA in gels by fiddling with the voltages, or by reversing the voltage at an angle slightly less than 180° (say 120°); but the fundamental difficulty of separating 50 to 400 million-base-pair DNA in a gel remains. Whoever can solve that problem will be a great hero to molecular biologists, because he or she will for the first time make possible the detailed study and manipulation of human chromosomes. There seem to be two possible paths to a solution of this problem: (a) to change the structure of the gel so that its 'coat-hook' properties apply to larger DNA; or (b) to reduce the contour length of the DNA uniformly, perhaps by wrapping it around proteins, to make the DNA seem smaller relative to the structure of the gel. Here is an interesting, if extremely difficult, problem for a clever student.

So far we have described methods to study the properties of large numbers of DNA molecules in solution (or in a gel or a crystal). Any properties so measured will represent the average over many different individual values. For example, there will typically be 10^{12} DNA molecules within any gel sample described above, all of which will move through the gel along slightly different paths and at slightly different rates!