

sequence such as CG to TG as noted in Chapter 11 may also contribute to such diversity). Finally, recombination occurs frequently in mature cells as part of the general mechanism for DNA repair: where a severely damaged stretch of bases on one chromosome may be cut away, and replaced by intact bases from a nearby homologous chromosome.

In relation to (c) nanotechnology, some scientists and engineers (for example the visionary Ned Seeman) are now trying to use the specific base-pairing properties of DNA in order to assemble electronic, mechanical or computing devices on a very small, near-atomic scale. Some progress has already been made at fabricating tiny rotary devices, by engineering DNA in the form of specific and controllable 'swivels'. But there are still fundamental difficulties to be overcome, when trying to build real, useful devices from such highly flexible building-blocks.

With regard to (d) evolutionary biology, it seems that recent developments in sequencing whole genomes of DNA have brought about a huge accumulation of information from hundreds of diverse species. This information can be used in principle to deduce historical relationships among many different species, on the assumption that genes move only from parent(s) to children; in other words, if we assume that hardly any genes have moved 'horizontally' *between* species. DNA sequencing can also provide details of the slight genetic variation found within any species, such as the human polymorphisms mentioned in Chapter 11. These data can be used to study recent historical events: for example the human settlement of different continents, or the strange 'bottleneck' of population size which occurred when *Homo sapiens* emerged about 100 000 to 200 000 years ago.

In summary, whether we consider the chemical basis of DNA structure, recombination, nanotechnology or molecular evolution, there is one overriding feature that distinguishes a well-understood field of science from a poorly understood one. In a well-understood field, there is a well-developed sense of the relations between things, and how they are structured with respect to one another. Often a single theory may explain hundreds of experimental observations. By contrast, in a poorly understood field there are many 'effects' and 'factors' and long, hard-to-remember words, but few relations among them in a theoretical sense. As helpful advice to a young student: the practitioners of any poorly-understood field may often be distinguished by their refusal to say the words 'I don't know!'

Our goal, therefore, should be to distinguish what is known from what is not known, and not ever to pretend that word-knowledge is a substitute for structural-knowledge. This important idea was

expressed clearly long ago by John Locke (1690), as a comment on his *Essay Concerning Human Understanding*:

We cannot but think that angels of all kinds much exceed us in knowledge; and possibly we are apt sometimes to envy them that advantage, or at least to repine¹ that we do not partake with them in a greater share of it. Whoever thinks of the elevation of their knowledge above ours, cannot imagine it lies in a *playing with words*, but in the contemplation of things, and having true notions about them; a perception of their habitudes and relations one to another. If this be so, methinks we should be ambitious to come in this part, which is a great deal in our power, as near them as we can.

Note

1. Repine: to fret, be discontented.

Further Reading

Protein-DNA theory

Oobatake, M., Kono, H., Wang, Y., and Sarai, A. (2003) Anatomy of specific interactions between lambda repressor and operator DNA. *Proteins* **53**, 33–43. A computer analysis may identify key features of the lambda repressor-operator DNA complex which are responsible for specific recognition.

Parry, D., Moon, S.A., Liu, H.H., Heslop, P., and Connolly, B.A. (2003) DNA recognition by the *Eco* RV restriction endonuclease using base analogues. *Journal of Molecular Biology* **331**, 1005–16. Chemical changes to base-pairs within the *Eco* RV recognition site GATATC are used to probe its specificity.

Tateno, M., Yamasaki, K., Amano, N., Kakinuma, J. *et al.* (1997) DNA recognition by beta-sheets. *Biopolymers* **44**, 335–59. Studies of general inter-atomic contact when protein beta-sheets are used to recognize DNA.

Recombination

Chen, L., Huang, S., Lee, L., Davalos, A. *et al.* (2003) WRN, the protein deficient in Werner syndrome, plays a critical structural role in optimizing DNA repair. *Aging and the Cell* **2**, 191–9. A human defect of recombination-repair enzymes can cause a disease known as Werner syndrome, which involves cancer and premature aging.

Mucha, M., Krol, J., Goc, A., and Filipinski, J. (2003) Mapping candidate hotspots of meiotic recombination in segments of human DNA cloned

- into yeast. *Molecular and Genetic Genomics* **270**, 165–72. The most recombinatorial human sequences may be identified by cloning into yeast as CG islands (Chapter 11) and repetitive ‘minisatellites’ (Chapter 10).
- Seidman, M.M. and Glazer, P.M. (2003) The potential for gene repair *via* triple helix formation. *Journal of Clinical Investigation* **112**, 487–94. Third-strand-forming DNA oligonucleotides can induce novel recombination and repair as a possible medical therapy.

Nanotechnology

- Okamoto, A., Tanaka, K., and Saito, I. (2003) Rational design of a DNA wire possessing an extremely high transport ability. *Journal of the American Chemical Society* **125**, 5066–71. Incorporation of a modified base ‘benzo-deaza-adenine’ enables better ‘hole transport’ through a DNA helix.
- Sussman, H.E. (2003) Nanodevices hold promise for gene therapy. *Drug Discovery Today* **8**, 564–5. New inventions from nanotechnology which could be used for DNA gene medicine.
- Zhu, L., Lukeman, P.S., Canary, J.W., and Seeman, N.C. (2003) Nylon/DNA: single-stranded DNA with a covalently stitched nylon lining. *Journal of the American Chemical Society* **125**, 10178–9. A novel composite of DNA and nylon polymer that may be useful for the construction of molecular devices.

Molecular evolution

- Anzai, T., Shiina, T., Kimura, N., Yanagida, K. *et al.* (2003) Comparative sequencing of human and chimpanzee MHC class regions unveils insertions/deletions as the major path to genomic divergence. *Proceedings of the National Academy of Sciences, USA* **100**, 7708–13. Detailed comparison of human and chimp DNA in an important genomic region.
- Cooper, G.M., Brudno, M., Green, E.D., Batzoglou, S. *et al.* (2003) Quantitative estimates of sequence divergence for comparative analyses of mammalian genomes. *Genome Research* **13**, 813–20. Use of DNA sequencing data to study the possible historical evolution of mammals.
- Thomas, J.W., Touchman, J.W., Blakesley, R.W., Bouffard, G.G. *et al.* (2003) Comparative analyses of multi-species sequences from target genomic regions. *Nature* **424**, 788–93. A study of molecular evolution by DNA sequencing among diverse vertebrates.

Appendix 1: Notes on the Derivation of Some Technical Terms

All of the terms in the following list have been introduced and explained in the text. The sole point of these notes is to explain the philological background of the words.

Archaea (or archae-bacteria). Micro-organisms, similar to ordinary bacteria, but different in their molecular organisation. Greek: *archaeo* = ancient, primitive.

Bacteria (singular: bacterium). From the Greek, a diminutive of *baktron*, a stick or staff.

Cell. A term first used by Robert Hooke in 1665, on seeing the structure of angular spaces in a thin section of cork in a microscope. The structure was similar to that of cells in a honeycomb, monastery or prison.

Circular. Like a circle in the sense of being endless: not necessarily even a plane figure (cf. linear).

Chromosome. Filamentous body in the cell nucleus which appears colored (Greek: *chromos* = color, *soma* = body) in the light microscope after treatment with chemical stains.

Enzyme. An organic catalyst, usually a protein. From the Greek *zumosis* = leaven, used to describe the process of fermentation.

Eukaryote. Organism in whose cells the DNA is contained within a nucleus. From the Greek: *eu* = truly, *karyon* = kernel, nut.

Gene. From the Greek *gen*, to produce. Cf. generator, generation, Genesis.

Homologous. Corresponding. From the Greek: *homo* = same, *logos* = word.

Interphase. From two Greek words: *phasis* = appearance and *inter* = between; used to describe an appearance in the middle of the cell cycle.

Linear. Like a line in the sense of having two ends and no branches: not necessarily straight, or even planar (cf. circular).

Metaphase. From two Greek words: *phasis* = appearance and *meta* = beyond or after; used to describe an appearance near the end of the cell cycle.

Nucleus. From the Latin *nux* = a nut; the central kernel of the cell. A term coined by the botanist Robert Brown in 1831.

Oligomer. A short example (Greek: *oligo* = few) of a long-chain molecule.

Polymorphic. Having many shapes. From two Greek words: *poly* = many, *morphe* = form.

Prokaryote. Organism (e.g. bacteria, archaea) in which the cell does not have a nucleus. Greek: *pro* = before, as if the nucleus came later.

Protein. From the Greek *proteios* = of the first rank. First applied to the large group of compounds by Muldner, acting on a suggestion made in 1838 by the famous chemist Berzelius.

Topoisomerase. From the Greek *topos* = a place and *isos* = equal. The ending -ase denotes an enzyme. An enzyme that changes the linking number *Lk* of DNA, to make different forms of DNA equal.

Toroid. A ring of the kind put through the nose of a bull: Latin *taurus* = bull.

Translation. From the Latin: *trans* = across, *latum* = past participle of 'bring'. Hence, movement of a body without rotation.

Appendix 2: The Chemical Theory of Base-stacking Interactions in DNA

In Chapter 2 we explained how the water-repelling hydrophobic quality of the bases could push a 'ladder' of DNA into a twisted, double-helical shape; and we called that a first-order effect on the structure of DNA. Then in Chapter 3 we invoked the same hydrophobic property in order to account for the 'propeller twist' which is usually found in the Watson–Crick base-pairs; and we described that as a weaker, 'second-order' effect. Propeller twist makes the stacking of base-pairs onto one another much less straightforward than for planar base-pairs, and it thereby provides a key to understanding some aspects of the conformational behavior of double-helical DNA. In Chapter 3 we also mentioned the contribution made to stacking by partial electric charges within the base-pairs, in order to explain more accurately some of the second-order structural effects. And we said that a fuller treatment of them would be given in Appendix 2.

Here, then, we shall explain more about the nature of partial electric charges in the organic bases of DNA, as they influence the overlap or stacking of adjacent base-pairs; and we shall describe some observations on stacking that can only be explained by their use. We shall also describe briefly some recent observations from the crystal structures of DNA, that show the important role of propeller twist in determining stacking arrangements.

A new detailed theory of chemistry, by which the base-pairs of DNA stack on one another specifically due to several kinds of chemical energy, including weak interactions between partial charges in the base-pairs, has been developed over the past decade by Chris Hunter. He had used the same theory previously with some success to describe the stacking behavior of large, flat molecules called 'porphyrins', which are used by the body for many different purposes, such as carrying electrons in the mitochondria and carrying oxygen in the blood. But the same theory is also very helpful in the study of DNA.

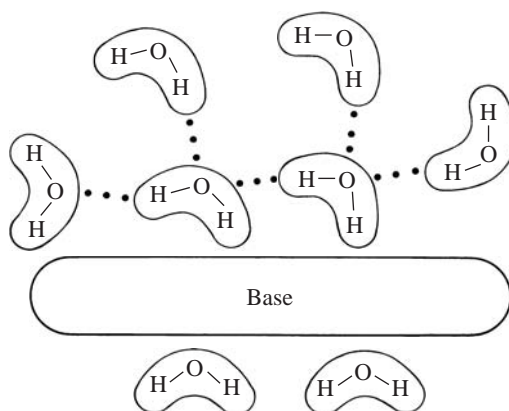


Figure A1 Schematic diagram showing how $\text{H}\cdots\text{O}$ bonds between water molecules are disrupted by a hydrophobic base. The loss of such hydrogen bonds can be minimized when two base-pairs stack directly on top of one another, as in Figs 2.4 and 2.5.

Hunter says that there are three general kinds of electrical force that control the interactions between molecules such as porphyrins, or among the base-pairs of DNA. The first force is the one that we have described already, namely a tendency for the base-pairs to stack on one another as fully as possible, in order to escape from contact with water. What is the basis of this effect, in molecular terms? As shown schematically in Fig. A1, water molecules in solution have a strong tendency to associate with other water molecules by means of hydrogen-to-oxygen or $\text{H}\cdots\text{O}$ bonds, which are electrical in nature. Normally any water molecule will have a total of four partners: two that place their hydrogen atoms in contact with the water oxygen, and two that place their oxygen atoms in contact with each of the water hydrogens. This intricate and ever-changing network of hydrogen-bonding between water molecules is responsible for the 'stickiness' of water, for example its tendency to form drops when falling slowly out of a tap. But when those water molecules contact an oily, water-insoluble DNA base, many hydrogen bonds of the kind $\text{H}\cdots\text{O}$ are lost, as shown in the diagram. Hence, the optimal arrangement of base-pairs in water is as shown previously in Figs 2.4 and 2.5, where the bases stack on one another as completely as possible in order to exclude water. In a solution of water mixed with ethanol (as in strong whisky) or, for example, at low water content in fibers of DNA, this kind of water-exclusion force becomes rather weak, so the base-pairs no longer need to stack so firmly onto one another.

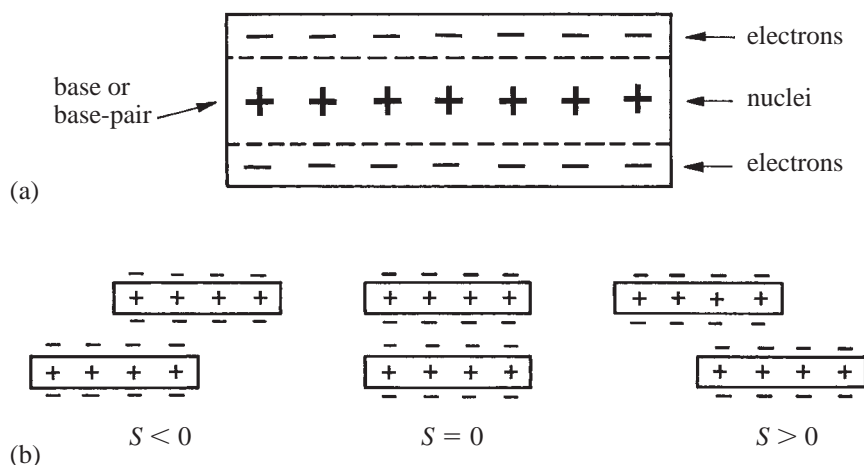


Figure A2 (a) Distribution of electric charge in a base or base-pair. (b) Positive or negative slide at a base-pair step reduces the repulsion between the negative charges, and produces some attraction between positive and negative charges in different base-pairs.

The second kind of force between base-pairs comes about because the upper and lower surfaces of any base (or base-pair) have a slight negative charge. As shown in Fig. A2(a), the negatively charged electrons that make up the 'aromatic' system of bonding between atoms in any base lie mainly above and below the main part of the ring. Yet the positively charged nuclei of the various atoms that join to form a base (such as carbon, nitrogen and oxygen) lie near the center of the ring, when viewed edge-on, as in Fig. A2(a). So the overall distribution of electric charge in any base looks somewhat like a 'sandwich', where the electrons are the bread and the nuclei are the cheese.

Now when two bases (or base-pairs) come into contact along their upper and lower surfaces, they must repel each other to some extent, because of a negative-to-negative charge repulsion. An idealized, vertical stacking of two base-pairs is shown in the center part of Fig. A2(b) for the case of slide $S = 0$. We said previously that a fully stacked geometry is favorable because it excludes water; but now we see that it is highly unfavorable when viewed in the context of our second force, which says that there is a strong negative-to-negative repulsion between base-pairs in a fully stacked geometry. Were it not for the water, the two base-pairs would repel one another like the North poles of two magnets. Generally, the base-pairs in DNA will slide away from one another in a left-to-right sense to escape this repulsion, because the flexibility of sugar-phosphate chains is greatest in that direction. Thus, as shown in

Fig. A2(b), the base-pairs can lie offset from one another at either negative slide S , as shown on the left-hand side of the drawing, or at positive slide S , as shown on the right-hand side. In either case, the negative-to-negative charge repulsion is lessened by the increased distance between pairs, and some positive-to-negative attraction may come about between nuclei in one base-pair and electrons in the other.

Thus the 'bread-and-cheese' effect is the exact opposite of the 'hydrophobic' effect, so far as the stacking of adjacent base-pairs on one another is concerned. Which of the two is the stronger? This depends critically on the amount of water which surrounds the bases. When conditions are wet, the hydrophobic effect wins out; but when conditions are dry, the bread-and-cheese effect is stronger.

These notions explain rather simply the structural behavior of double-helical DNA in fibers. Originally, in the 1950s, it was not possible to grow crystals of DNA, because scientists had not yet learned how to synthesize DNA chemically; so most investigators had to isolate chromosomal DNA from natural sources such as calf thymus or salmon sperm, and then pull it out into long fibers, in order to study its structure by X-ray diffraction (see Chapter 9). When these fibers were wet, the X-ray photographs showed a 'B' form structure with 10 base-pairs in a complete helical turn, and with base-pairs stacked vertically over one another at slide $S = 0$. But when the fibers were dry, either an 'A' form with 11 base-pairs per turn, or a 'C' form with 9.0 to 9.3 base-pairs per turn was observed. In the 'A' form, the base-pairs stack offset at a negative slide $S = -1.5 \text{ \AA}$, while in the 'C' form they stack offset at a positive slide $S = +1.0 \text{ \AA}$. Recall that slide S and twist T change in tandem, as described in Fig. 3.13.

Not all possible forms of double-helical DNA can be seen in fibers, because the packing of long, thin DNA molecules into fibrous bundles often restricts their helical repeat to integral values such as 10 or 11 base-pairs per turn. Yet when such structures are observed, we can understand their behavior as follows. When the fiber is wet the hydrophobic forces are dominant, so the base-pair stacking is one of low slide near $S = 0$, and the result is the 'B' form; but when the fiber is dry the hydrophobic forces are weaker, so the bread-and-cheese effect becomes the stronger of the two. Then the stacking involves a slide which is either negative – in which case the 'A' form appears; or positive – in which case the 'C' form appears. A left-handed 'Z' form, mentioned briefly in Chapter 2, can also be seen in fibers when the conditions are relatively dry; and in this 'Z' helix the base-pairs lie offset from one another, as expected from this theory.

Finally, according to Hunter, there exists a third important kind of electrical force between base-pairs, which depends on the identities and locations of individual atoms in the four different kinds of base ring. Certain atoms, mainly hydrogens that are attached to nitrogens as NH or NH₂, or to carbon as CH, can generate a partial positive charge in their close vicinity; while other atoms such as oxygen O or simple nitrogen N can generate a partial negative charge. Thus, the stacking of different base-pairs on one another can be 'fine-tuned' by the electrical interactions of individual atoms in the rings; and this feature will depend strongly on the sequence or ordering of bases at any step.

It follows from this that bases are not so free to slide over each other as we might suppose, on the basis of a model in which the surfaces of the bases are featureless, as we have assumed so far with our models in which base-pairs are represented by simple blocks. Instead, certain positions will be preferred – positions in which a positively charged atom from one base lies directly above or below a negatively charged atom from the other, and in which atoms with 'like' charges do not lie directly above or below each other. In other words, the preferred positions will maximize attractive juxtapositions of atoms and minimize repulsive ones. This so-called 'partial-charge' effect is well-known in chemistry; and it can have a striking influence in the positioning of large, flat molecules relative to one another.

Some partial charges for atoms in the DNA base-pairs, as calculated by Hunter, are shown schematically in Fig. A3. Roughly, they correspond to what you would expect from having oxygens and nitrogens negative, and hydrogens positive; but the computer calculations are much more accurate than any assignment of partial charges that could be made directly from chemical intuition. For clarity, no individual atom types are identified; but one may compare with Fig. 2.11(a) and (b) to identify the different kinds of atom at various places in the rings.

The most striking aspect of this figure is a dense accumulation of negative charge along the major-groove edge of the guanine ring, as shown at bottom left. Two negatively charged atoms, a nitrogen and an oxygen, lie there close together in space, and so generate a joint concentration of negative charge that is greater than that of any single atom elsewhere in the G–C or A–T base-pairs. There is also a concentration of positive charge on the major-groove edge of the cytosine ring. What are the implications of highly charged guanine and cytosine rings for the three-dimensional structure of DNA?

First, when two guanine rings stack over one another in a step GG/CC, they will not find it easy to stack vertically, one over the other at slide $S = 0$, owing to a strong repulsion between negatively