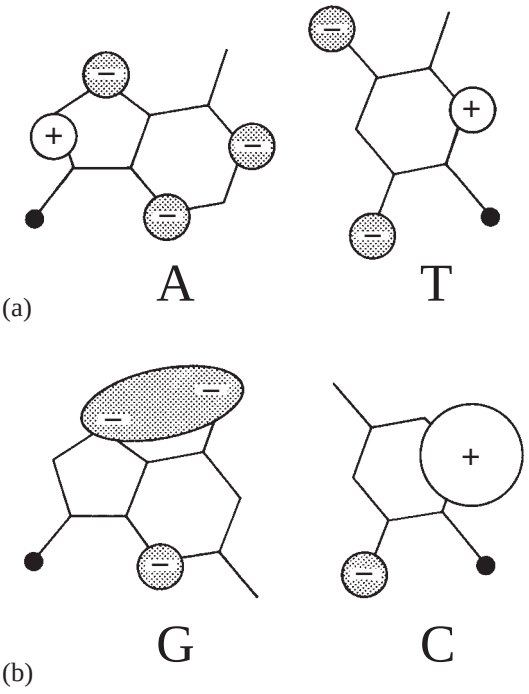

Figure A3 Regions of 'partial charge' for A–T and G–C base-pairs. The base-pairs have the same relative orientations as in Fig. 2.11(a) and (b), where atom types H, C, N, O can be identified.

charged regions; instead they will prefer to lie offset from one another, either at low slide, $S = -1$ to -2 Å or else at high slide, $S = +1$ Å, in order to minimize charge–charge repulsions. And indeed, we saw precisely that effect in Fig. 3.12(c), where the GG/CC step showed an unoccupied middle range of slide from $S = -1$ to 0 Å. Second, when a G–C base-pair stacks over another such pair in a step GC/GC or CG/CG it will favor certain stacking geometries over others, because of interactions between the strong negative and positive charges on the guanine and cytosine rings, as shown in Fig. A3.

In contrast to the G–C base-pairs, the A–T base-pairs have no strong, joint concentrations of partial electric charge. And so the strong stacking preferences that we have just described for steps made up from two G–C base-pairs should not act so strongly when there is an A–T base-pair in any step, say for the step AC/GT.

A careful and detailed examination of all kinds of dinucleotide step in many crystallized oligomers of DNA has been made by Mustafa El Hassan. Figure A4 shows his generalized plot of roll R against slide S, in the same style as the three examples shown in Fig. 3.12, but with preferred regions now marked for nine of the

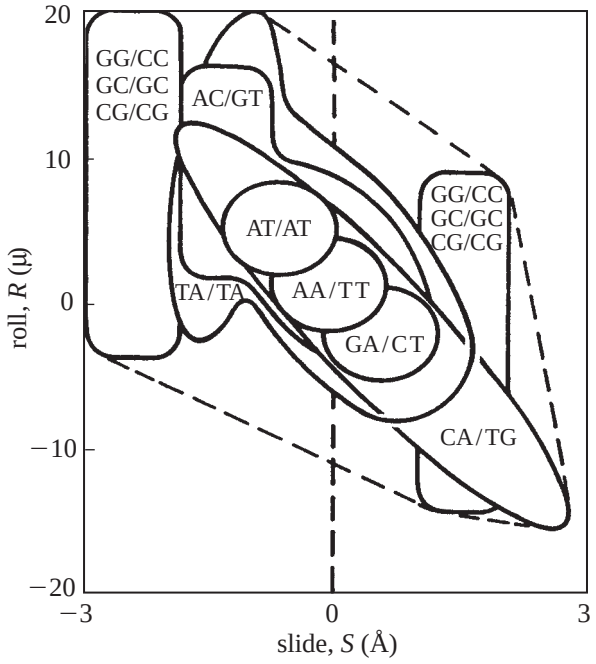

Figure A4 Schematic plot of roll *R versus* slide *S* for nine types of base-pair steps as seen in crystallized DNA oligomers. (Data for three particular step types are shown in more detail in Fig. 3.12.) Adapted from El Hassan and Calladine (1996) *Endeavour* **20**, 61–7.

ten kinds of sequence step, for which data are plentiful and clear. The partial electric charges, as described above, might be expected to play a clear part in the stacking of those steps where the constituent base-pairs are both G–C. Indeed, for all of the steps GG/CC, GC/GC and CG/CG, it is found that the conformation of the step prefers to lie in either of two separate regions of roll/slide, as may be seen from detailed inspection of Fig. A4.

Figure 3.12(a) also showed an example, AA/TT, of a step which is 'rigid' in the sense that each of roll *R* and slide *S* is fixed within narrow limits. Figure A4 shows that the steps AT/AT and GA/TC behave in the same way, but adopt somewhat different chosen conformations.

While the partial-charge effects decribed above thus enable us to understand some of the empirical features shown in Figs 3.12 and A4, they cannot provide predictions of the actual conformation of dinucleotide steps without the introduction, somehow, of constraints imposed by the backbones. Now the simple idea of a 'rigid' backbone link, as shown in Fig. 3.13, provides a first-order simulation of backbone constraints. But it is too crude to predict many of the

details shown in Fig. A4, because the sugar–phosphate backbone, as explained in Chapter 2 and shown in Fig. 2.8, acts as a rather more complicated linkage, because some of the inter-atomic bonds can rotate a little.

However, Hunter and colleagues have devised a simple and successful way of accounting for backbone behavior. First, they take slide and 'shift' as their main working variables for a dinucleotide step. Shift is one of Euler's six degrees of freedom of a rigid block with respect to another: it is an in-plane translation, like *slide*, but measured along the 'front-to-back' axis shown in Fig. 3.7. In general, shift does not vary nearly so much as slide; and for many purposes – as indeed throughout much of this book – its variations may be ignored. But it turns out to be important in the calculation of the conformation of dinucleotide steps. Hunter's idea is to endow the simple backbone link with an empirically-derived energy function of slide and shift; and when this is coupled with the chemical-energy computations, the conformations of individual dinucleotide steps of given composition are predicted well.

These same workers have recently investigated a further important feature of dinucleotide step conformation, which we have not mentioned previously: it concerns a conformational 'communication' between neighboring steps. Now a detailed study of *tri*-nucleotide and *tetra*-nucleotide fragments from El Hassan's database shows that both slide and shift are strongly correlated from step to step, whereas all other conformational variables depend only on the composition of the step in question: for example, slide, which has an overall range of some 6 Å (see Fig. A4) rarely differs by more than 1 Å from one step to the next. Shift, on the other hand, tends to be anti-correlated, in the sense that its values generally go high-low-high-low along the molecule.

It is clear that any such communication of conformational features from one step to the next must be made *via* the backbones, and probably through the single sugar ring that is shared by two consecutive backbone segments. Again, by further endowing the backbone with simple, semi-empirical energy expressions that depend on the *difference* of slide from step to step (in order to penalize variation), and likewise the *sum* of the shift values, Hunter and colleagues have been able to predict, by economical computations, a range of observed context-dependent features in DNA oligomers. For example, G–C rich regions tend to be *bi*-stable between 'A' and 'B' or 'C' forms (low and high slide, respectively) – a feature that may indeed be seen qualitatively in Fig. A4 when it is recalled that slide varies but little from step to step. Also, the role of AA steps in inhibiting the 'A' form over the surrounding region is

clarified by this work. Finally, these recent computations have given encouraging estimates of the sequence-dependent flexural/torsional stiffness of DNA.

Now we may recall that the more-or-less unique conformation of AA/TT steps was attributed in Chapter 3 to a 'locking' or stabilizing effect, on account of the high propeller twist in both of the A–T base-pairs (recall Fig. 3.6). More detailed data from DNA crystals show that there is in fact a very strong empirical correlation between the average propeller twist in the two base-pairs of any step, and the total range of slide which is allowed. Thus, all three steps in Fig. A4 which show the narrowest range of slide values prefer high values of propeller twist, while those steps that show the widest range of slide values, particularly CA/TG, tend to have the least propeller twist. One can make a simple physical model, as suggested in Exercise 3.5, to show that a small propeller-twist offers little hindrance to slide.

The data shown in Fig. A4 have all been derived from X-ray diffraction studies of 'free' DNA oligomers, which are typically packed with 40 to 60% water into a crystalline state. Different ranges of conformation in individual steps may therefore be induced in some cases, by the moderate but non-negligible intermolecular forces which arise when these oligomers are deformed so as to pack together into the regular arrangements known as 'crystal lattices'.

But when DNA is bent strongly around proteins, much wider ranges of roll-slide-twist are seen in X-ray studies. For example, AA/TT steps (which in Fig. A4 appear 'rigid') show high roll and low twist in the sequence TATATAAA, when it is bound to the protein TBP (Chapter 4); while certain CA/TG steps bound to the histone octamer (Fig. 7.2) again show a wider range of conformation than would be deduced from Fig. A4.

In summary, the stacking interactions of base-pairs in DNA are influenced by three kinds of electrical force: (1) maximizing base-to-base overlap to avoid contact of the bases with water; (2) reducing base-to-base overlap to avoid repulsion of negatively charged surfaces; and (3) maximizing attraction and minimizing repulsion between partial charges on individual atoms in the base-pair rings. And all of these effects are subject, of course, to various subtle constraints imposed by the sugar–phosphate chains of DNA, and by propeller twist in the base-pairs; and so the base-pairs will not always be able to attain their otherwise optimal local configurations.

There do exist, of course, a wide range of other constraints which may combine to influence local step conformations, and thereby produce severe helical distortions. These include imposed curvature: which may be mild as in the binding of 434 repressor, or severe as in the binding of TBP, and may involve interaction of the

sugar–phosphate chains with the protein; and there is also the intimate intrusion of intercalating drugs, or hydrophobic amino acids from contacting proteins, into the space between neighboring base-pairs, which can produce curvature and striking local distortions of geometry.

We think that a deeper understanding of all of these effects will be of considerable use in understanding the actions of DNA in biology; but a complete discussion is beyond the current scope of our book.

Further Reading

El Hassan, M.A. and Calladine, C.R. (1995) The assessment of the geometry of dinucleotide steps in double-helical DNA: a new local calculation scheme. *Journal of Molecular Biology* **251**, 648–64. How to calculate roll, slide and twist accurately for any base-pair step, by a method which does not rely on an overall helix axis.

El Hassan, M.A. and Calladine, C.R. (1996) Propeller-twisting of base-pairs and the conformational mobility of dinucleotide steps in DNA. *Journal of Molecular Biology* **259**, 95–103. The propeller twist of base-pairs seems to influence their ability to slide along the long axis of any base-pair step.

El Hassan, M.A. and Calladine, C.R. (1998) Two distinct modes of protein-induced bending in DNA. *Journal of Molecular Biology* **282**, 331–43. When strongly bent by proteins, DNA dinucleotide steps move outside the range of conformations shown in Fig. A4.

Gardiner, E.J., Hunter, C.A., Packer, M.J., and Willett, P. (2003) Sequence-dependent DNA structure: a database of octamer structural parameters. *Journal of Molecular Biology* **332**, 1025–35. A systematic computation of seven-step oligomers, which provides a clear description of context-dependent behaviour, including bi-stability.

Hunter, C.A. (1993) Sequence-dependent DNA structure: the role of base stacking interactions. *Journal of Molecular Biology* **230**, 1025–54. A clear description of the chemical theory which underlies base-stacking preferences in DNA, and some calculations of these effects.

Matsumoto, A. and Olson, W.K. (2002) Sequence-dependent motions of DNA: a normal mode analysis at the base-pair level. *Biophysical Journal* **83**, 22–41. Accurate calculations of DNA sequence-dependent structure in solution, using preferred values of roll-slide-twist as seen in crystals.

Packer, M.J., Dauncey, M.P., and Hunter, C.A. (2000) Sequence-dependent DNA structure: dinucleotide conformational maps. *Journal of Molecular Biology* **295**, 71–83. Theoretical analysis of base-pair stacking energies in terms of roll, slide and twist at a dinucleotide level; and a prediction of rigidity for AA/TT, but flexibility for CG or CA/TG or TA, in accord with observed preferences of dinucleotide sequences.

Richmond, T.J. and Davey, C.A. (2003) The structure of DNA in the nucleosome core. *Nature* **423**, 145–50. Details of *roll*, *slide*, *twist*, etc. in DNA bent around proteins, showing even more extreme conformations of CA/TG steps than in Fig. A4.

Suzuki, M., Amano, N., Kakinuma, J., and Tateno, M. (1997) Use of a 3-D structure database for understanding sequence-dependent conformational aspects of DNA. *Journal of Molecular Biology* **274**, 421–35. A careful analysis of free-DNA crystal structures in terms of roll, slide and twist at individual base-pair steps.

Wang, L., Hingerty, B.E., Srinivasan, A.R., Olson, W.K., and Broyde, S. (2002) Accurate representation of B-DNA double helical structure with implicit solvent and counterions. *Biophysical Journal* **83**, 382–406. The development of refined energy parameters to calculate DNA sequence-dependent structure in solution, including water and cations.

Appendix 3: How to Modify Gene Expression using Anti-sense Oligonucleotides, Ribozymes, or si-RNA

In Chapter 10 we explained how specially designed DNA or RNA molecules may be used in 'gene therapy' to correct the aberrant expression of particular genes within diseased animals or plants. Here we shall elaborate on three current methods that use DNA or RNA to inhibit gene expression. First, there is the 'anti-sense' approach, where small RNA (or DNA) oligonucleotides are designed so that they will form 20 to 30 specific base-pairs with some small part of a long messenger-RNA, and thereby block its translation into protein by several possible mechanisms. Second, there is the 'ribozyme' approach, where small RNA oligonucleotides are again designed to form specific base-pairs with some long messenger-RNA; but in addition they now carry a catalytic domain for RNA cleavage, which cuts the message so that full-length protein cannot be made (similarly for 'DNAzymes', which are composed of DNA rather than RNA). And third, there is the 'small interfering RNA' approach, where small RNA oligonucleotides once again are designed to form specific base-pairs with some long messenger-RNA; yet in this case they guide to the message a cellular ribonuclease complex, which cuts the message so that it cannot be translated into protein.

Since many of these methods use RNA, we shall first describe a few of its relevant structural properties. As we showed earlier in Fig. 10.3, the only chemical difference between RNA and DNA backbones consists of an OH group on the sugar ring at carbon 2'. DNA usually forms a double-helical structure; and indeed RNA is also able to form a double-helical structure, which is similar to the

'A' form of DNA (see Chapter 2). But whereas DNA is nearly always found in Nature in continuous double-helical form through Watson–Crick base-pairing, cellular RNA is usually single-stranded. Nevertheless, it can still form base-pairs between different parts of the same single strand. Thus, RNA often folds into complicated structures like that of the transfer-RNA molecule shown schematically in Fig. 2.14. Another example is the elaborate fold of the RNA found within the large assembly of ribosomal RNA molecules (called 28S, 18S and 5.8S in animals). As mentioned in Chapter 1 (see Fig. 1.12), this molecular machine translates information encoded by base triplets in the messenger-RNA into a linear sequence of amino acids within a polypeptide chain.

Anti-sense therapeutics

Now let us explain what is meant by the word 'anti-sense'. At the level of double-stranded DNA, the coding sequence of any gene is contained within the 'sense' strand, whereas its base-paired complement lies within the 'anti-sense' strand. Thus when messenger-RNA is made from any DNA gene by RNA polymerase, the newly-made RNA represents a 'sense' strand, whereas the DNA template which was 'read' by the polymerase represents an 'anti-sense' strand. How then might some short piece of anti-sense RNA or DNA, which presumably binds to long messenger-RNA after it has left the nucleus and entered the cytoplasm, impair the expression of that message into protein?

Intuitively, one might suppose that steric hindrance by a short antisense molecule could prevent proper 'reading' of a long messenger-RNA at the ribosome, where proteins are made, as shown in Fig. A5(a) on the right. This expectation has proved true in some cases; yet an equally important influence seems to be the action of a cellular enzyme known as 'RNase H' which cuts RNA–DNA hybrids. The normal function of RNase H is to degrade small pieces of RNA in RNA–DNA hybrids, that serve as primers for replication of the DNA during cell growth. Yet when short anti-sense DNA binds to a long messenger-RNA, that process also creates an RNA–DNA hybrid, which may be degraded by the nuclease RNase H, as shown on the left of Fig. A5(a).

Any 'anti-sense' nucleic acid will be liable to quick destruction by cellular enzymes that cut nucleic acids: such enzymes are known as 'nucleases'. In order to protect the full-length anti-sense molecule from the action of nucleases, certain modifications must be made. One approach by chemists has been to replace an oxygen atom on the phosphate group with a nuclease-resistant sulfur (this

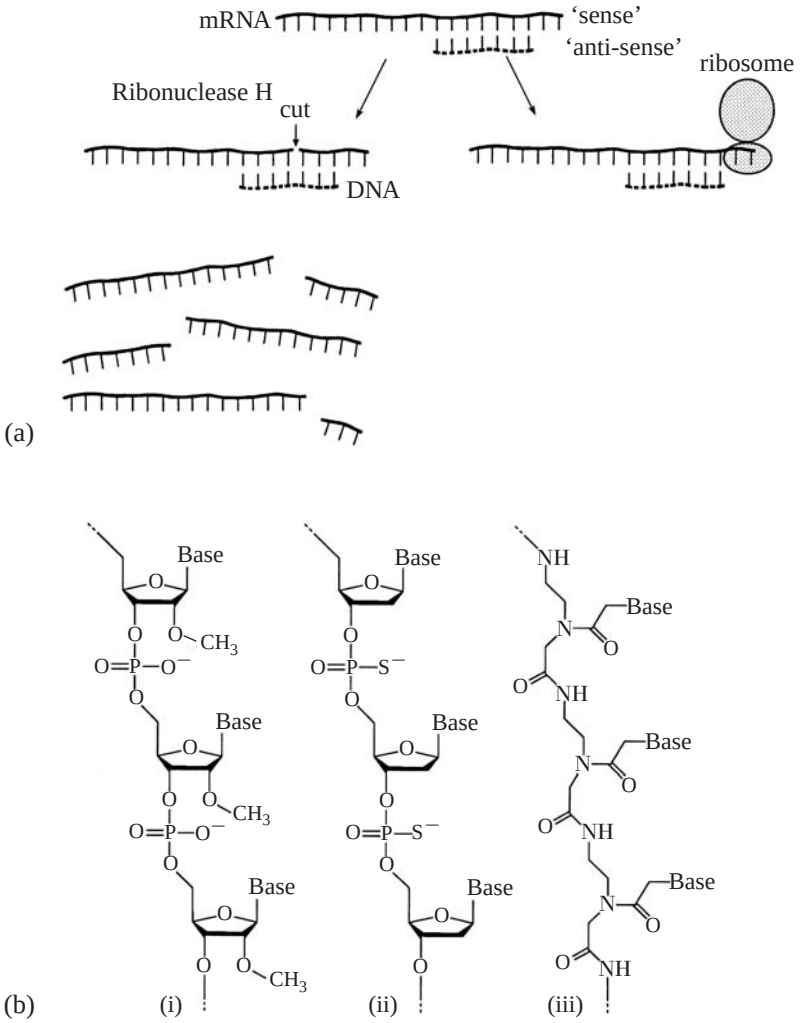

Figure A5 (a) Two common mechanisms by which anti-sense DNA can reduce gene expression: steric hindrance of translation at the ribosome (right), or destruction of the transcript by ribonuclease H (left). (b) Representative oligonucleotides used in the anti-sense approach. In the DNA molecule on the left, one of the oxygens on the phosphate group has been replaced by sulfur, to form a phosphoro-thioate that is resistant to nucleases. In the center, an anti-sense RNA molecule can be protected against destruction by placing a CH_3 group on the 2′ OH of the sugar. And on the right, the nucleic acid can be protected by replacing the sugar–phosphate backbone entirely by N-(2-amino-ethyl) glycine linkages, which are like peptides. There are many other stabilizing modifications, not shown here, that have also been useful for the anti-sense approach.

is called a phosphoro-thioate): see Fig. A5(b), center. Anti-sense RNA can also be protected against nucleases by adding a $-CH_3$ group to the 2′ OH on the sugar: see Fig. A5(b), left. Using these and other protective modifications, carefully designed anti-sense

oligomers can often exert a potent influence on specific gene activities in living cells. Indeed, some anti-sense molecules are now undergoing clinical trials for the control of diseases such as viral infection or cancer.

In order to achieve even greater nuclease resistance, Peter Neilsen and colleagues have made a stable form of oligonucleotide by synthetic chemistry, which has *peptide* linkages between adjoining bases, in place of the sugars and phosphates found in natural RNA and DNA. That 'peptide nucleic acid' or 'PNA' is shown on the right in Fig. A5(b). Because its backbone is electrostatically neutral, PNA tends to form PNA-DNA or -RNA hydrids that are more stable under low-salt conditions than natural DNA or RNA, whose phosphate backbones contain many repulsive negative charges. PNA has turned out to be very important for diagnostic work, where different DNA or RNA molecules are detected on microarrays (see Fig. 10.5), and great chemical stability is required. However, PNA has not yet become the panacea which many people once expected for the control of gene expression in cells, because the PNA-RNA hybrid is not cut by RNase H; and also because simple steric hindrance is usually not enough for any PNA to prevent the expression of some messenger-RNA into protein at the ribosome. The ribosome is a very powerful engine for making proteins, and it often manages to read through an RNA message, even when some anti-sense DNA or PNA is present.

PNA and other anti-sense nucleic acids also show promise for correcting the occasional mistakes that arise in the routine processing of messenger-RNA. Normally, certain sections of the RNA that do not code for protein (these are called 'introns') are 'spliced out' before translation into the protein. But in certain diseases caused by mutation, some of the splicing of a particular messenger-RNA may not occur; or else an alternative set of splice sites within the RNA might be used instead. This splicing of RNA at the wrong site has been prevented in cells by use of anti-sense oligonucleotides that are complementary to the incorrect splice junction. Also, proper splicing has been restored in cells where loss of a normal splice-site has occurred, by linking to an anti-sense PNA that binds close to the splice site a small peptide, which activates splicing.

RNA ribozymes and DNAzymes

Ribozymes were discovered in the 1980s, with the surprising finding that certain RNA molecules could cleave either themselves or other RNA molecules. Subsequently it was found that ribozymes can also catalyze other chemical reactions, in addition to cutting. We can learn something about the catalytic potential of RNA from