1

The origin of structural DNA nanotechnology

Everyone knows that DNA is the genetic material of all living organisms. Its double helical structure has become an icon for our age. The publication of its double helical structure by Watson and Crick in 1953 revolutionized biology.[1.1] Its most prominent applications today are in clinical diagnosis of genetic diseases and pathogenic organisms and in forensics. The key element of DNA is that it contains *information* – information in a form that is easy to understand, utilize, and manipulate. The central feature of this information is that it is linearly encrypted in the sequence of the DNA polymer. There are four different elements to this information, known as A, T, G, and C. We'll get into the chemical details of what those letters mean a little later. The important thing is that the molecule is in its most stable state (has the lowest free energy) when A on one strand is opposite T on the other strand, and when G on one strand is opposite C on the other strand. As Watson and Crick famously put it, it did not escape their attention that this complementary pairing leads immediately to a mechanism for replication: an A on the parental strand means you put a T on the daughter strand, and vice versa; similarly a G on the parental strand means you put a C on the daughter strand, and vice versa. It is important to realize that strands exhibiting this complementarity can be put together *in vitro*, a fact first noted by Alexander Rich and David Davies.[1.3] A key and often unvoiced aspect of this mechanism is that the helix axis is linear, not in the geometrical sense of being a straight line, but in the topological sense that it is not branched. This book is about what happens when the helix axis is branched and how we can use it to make new and interesting molecules and materials on the nanometer scale.

The chemical details of the classical structure of DNA are shown in Figure 1-1, and the backbone structure of DNA is shown in Figure 1-2. The double helical structure has many interesting features that need to be mentioned. First, the backbones are antiparallel. What do we mean by that? There is a chemical polarity

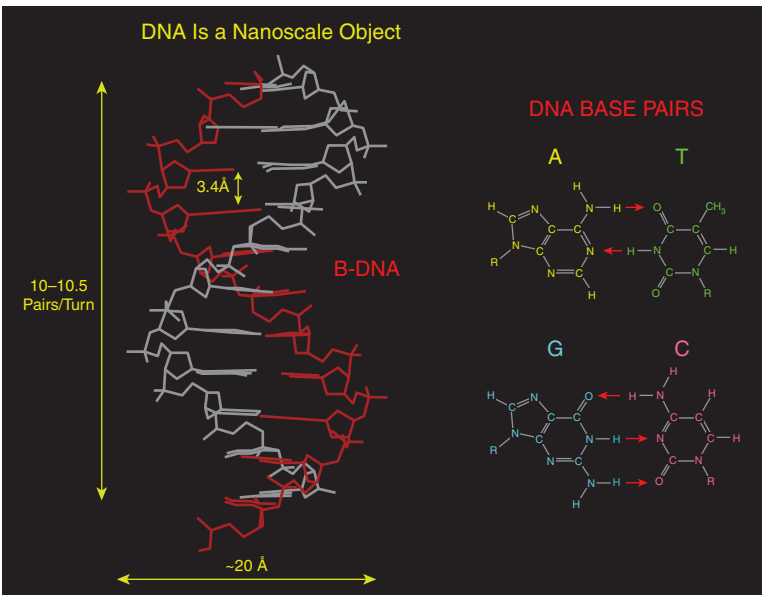

Figure 1-1 *The DNA molecule.* The double helical structure of DNA in the B form and its dimensions are indicated on the left. The details of the base pairing are shown on the right, where two hydrogen bonds are indicated between A and T, and three hydrogen bonds are indicated between G and C.

to the molecules, so that the strands have directionality. If you look at the left side of Figure 1-1, you will see pentagons every repeating unit (called a "residue") in both the gray strand and the red strand. They are pointing in opposite directions. For example, in the third step from the top of the helix, the pentagons are pretty much in the plane of the paper; the gray pentagon has a vertex pointing to the bottom of the page; by contrast, the pentagon opposite it on the red strand has a vertex pointing to the top of the page.

This arrangement is much like the opposite lanes in a two-lane highway with one lane going in each direction. Although the two strands of DNA are like the two opposite lanes of a highway, highways would not be very useful if they were like DNA: the road would go on forever, with no opportunity to turn off it. Branched DNA molecules can be thought of as intersections, where two roads meet. However, at the intersections in DNA, the lane is forced to turn and go off in a new direction, as shown in Figure 1-3. The lanes in Figure 1-3 are drawn in four independent colors, and the arrows of the same colors represent vehicles in the same lanes. Thus, the vertical red arrow pointing downward in the red lane is going south and the vertical cyan arrow opposite it pointing upward is going

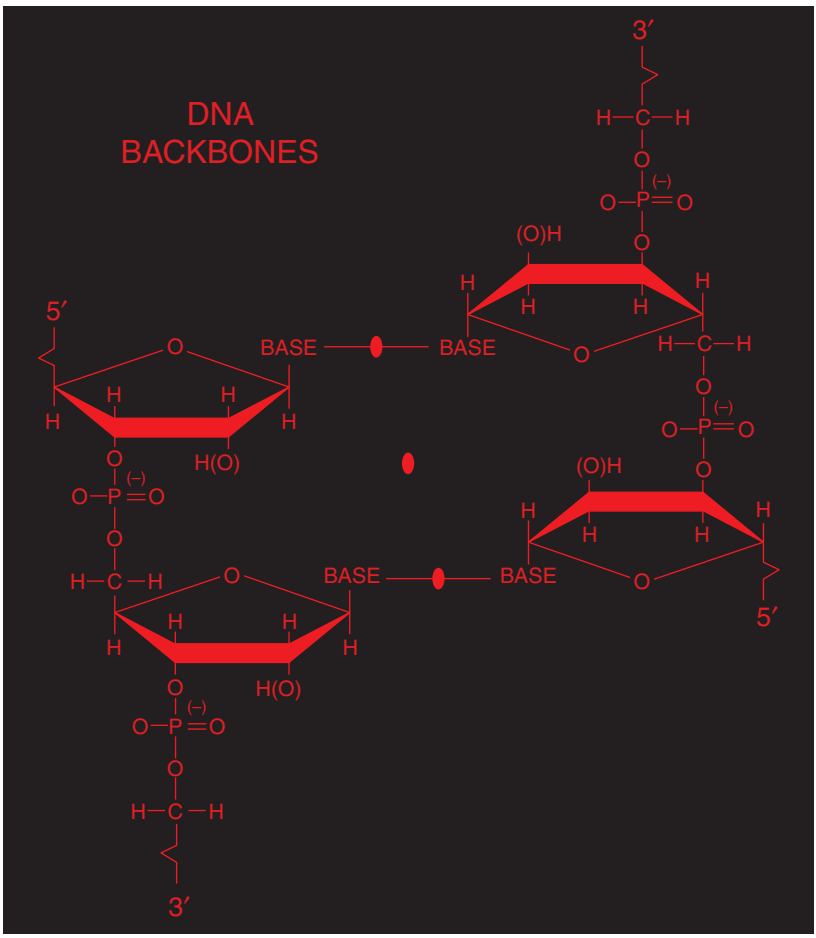

Figure 1-2 *DNA backbones.* The stereochemistry and directionality (5′ to 3′) are shown. The antiparallel nature of the backbone is evident from this representation. Note that there are two-fold axes indicated by little lens-shaped symbols. These are perpendicular to the helix axis and they relate only the backbones, and not the base pairs; there is one two-fold axis through each base pair plane, and a second one halfway in between them.

north. However, the red arrow makes a right turn at the intersection to go west, and after its right turn it is opposite the green arrow in the green lane going east. Similar relationships exist between the arrows in the other directions: all the lanes make right turns and change direction by 90°. The directions that they are going correspond to the movement of traffic in North America or in continental Europe, where vehicles drive on the right. Traffic in the UK, India, Japan, or

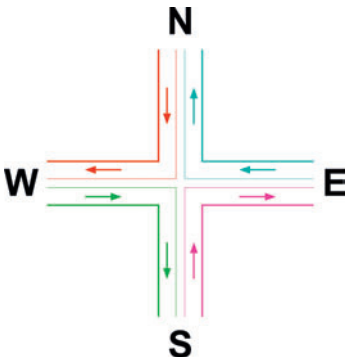

Figure 1-3 *A 4-way intersection.* Four different roads, coming from the north, west, south, and east, are shown, analogous to a 4-arm DNA branched junction. The flow of traffic through the intersection is color-coded, with the direction of traffic (in right-hand drive countries) indicated by arrows. Similar to DNA, each road has traffic going in two antiparallel directions. Reprinted with permission from Springer.[1.2]

Australia, where vehicles drive on the left, would be represented by a diagram that is the mirror image of Figure 1-3.

I started off talking about 4-way intersections, because that was what got me interested in branched DNA. The 4-way intersection or junction is analogous to an intermediate in genetic recombination, called the Holliday junction.[1.4] Although the possibility of branching had been mentioned earlier,[1.5,1.6] the Holliday junction was the earliest DNA junction that people thought about functionally, and in many respects it is still the most popular junction to make and with which to work, as we will see in later chapters. Of course, there are lots of kinds of intersections, not just those that consist of two roads crossing. It is possible to have three roads meet in a 3-way intersection,[1.7] or five or six,[1.8] or even more.[1.9] Generalizations of Figure 1-3 are shown in Figure 1-4, which illustrates 3-way and 6-way junctions.

Structural DNA technology got started one day in September 1980, when I went over to the pub on campus to think about nucleic acid junctions, which were analogous to 6-way intersections. Because they were easy to draw that way, I walked into the bar thinking of them as having snowflake-like six-fold symmetry. However, during my first beer, I suddenly thought about M.C. Escher's famous woodcut, *Depth*, which is shown in Figure 1-5. I realized that the fish in the drawing could be thought of as 6-arm junctions: there was no reason to draw 6-arm junctions as planar objects, like roadways usually are, but instead I could think of them in three dimensions. Furthermore, *Depth* doesn't show just a single fish, it shows a whole arrangement of fish. In fact, this

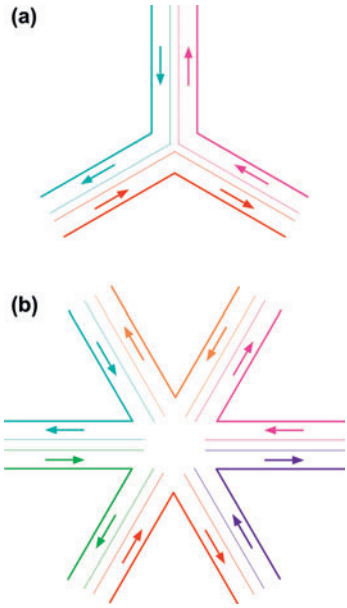

Figure 1-4 *Non-standard intersections.* (a) *A 3-way intersection.* (b) *A 6-way intersection.* In the same way that there are 3-way intersections and 6-way intersections, there are 3-arm DNA branched junctions and 6-arm DNA branched junctions. Junctions have been made with up to 12 arms. Reprinted with permission from Springer.[1,2]

arrangement is just like the arrangement of molecules in a molecular crystal: a periodic (repeating) pattern front to back, left to right, and top to bottom.

This was very meaningful to me, particularly at that time, because I was entering my fourth of five probationary years as an Assistant Professor. I had been hired to do crystallography on biological macromolecules, but had not by that time managed to crystallize any molecule of interest to myself or to others. Consequently, I was facing a fatal progression of "no crystals, no crystallography, no crystallographer." The fact that Escher had been able to organize these fish into a crystalline arrangement in a deliberate fashion made me think that I could organize nucleic acids the same way, and to get my crystals to form by a designed self-assembly process.

Of course, there was a catch! Escher was in complete control of every stroke in his artwork. As a natural scientist, I needed to work with a natural interaction to get branched DNA molecules to associate with each other and thereby form a crystalline lattice. The molecules themselves need to be connected to each other through some sort of chemical process. Fortunately, a process like this does exist, and has been used since the early 1970s to put simple double helices

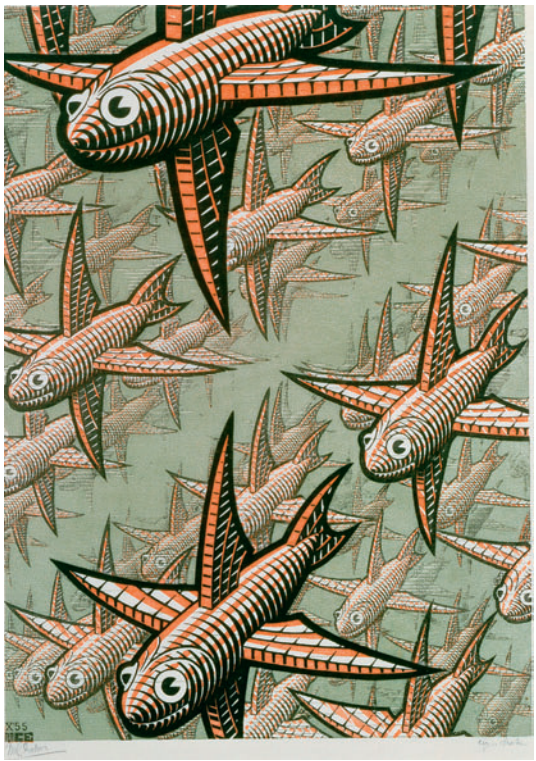

Figure 1-5 *Escher's woodcut, Depth.* This image inspired the field of structural DNA nanotechnology. The fish are analogous to 6-arm junctions: starting from their centers, they have a head, a tail, a top fin, a bottom fin, a right fin, and a left fin. In addition, they are organized like the molecules in a molecular crystal, demonstrating periodicity front to back, left to right, and top to bottom. Copyright (2015) The M.C. Escher Company – The Netherlands. All rights reserved. www.mcescher.com

together.[1.10] This process is based on a notion called "sticky ends." Sticky ends result when two strands of a double helical DNA molecule are not quite the same length. In the example shown in Figure 1-6, we see a red double helix and a blue double helix. Of course, the double helices don't look like helices here, because they have been unwound so that it is easy to display their sequences. Let's look at the two double helices at the top of the picture. The blue double helix has two strands, and the upper strand is four residues longer than the bottom strand. The overhang of four extra nucleotides is on the 3′ end, because it is near the arrowhead indicating the direction that the strand goes. The red double helix is similar, except that its overhang is on the 3′ end of the bottom strand. We always describe nucleic acid sequences from the 5′ end, even if we

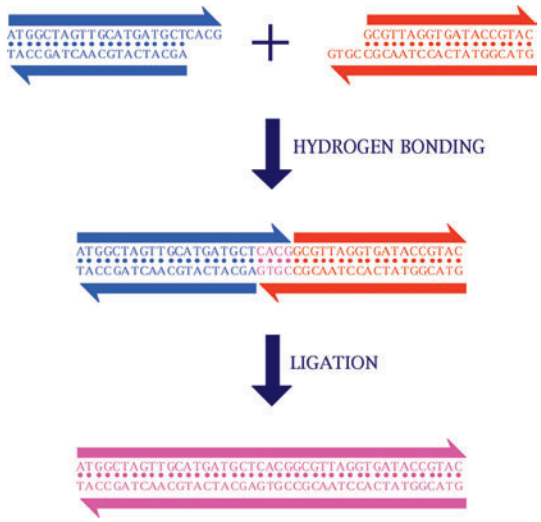

Figure 1-6 *Sticky-ended cohesion.* Two unwound double helices are shown at the top. Their strands are of slightly different lengths, creating overhangs that are called "sticky ends." If the sticky ends are complementary and conditions are proper, the two molecules can cohere, as shown in the middle. It is possible to join the ends covalently by enzymatic (or sometimes non-enzymatic) ligation.

have to read the picture backwards to do it. Given this rule, the sequence of the blue sticky end is CACG, and the sequence of the red sticky end is CGTG. The antiparallel nature of the double helix makes it a little tricky, because you have to read the sequence of one of the strands backwards to see it, but these two short strands are complementary to each other. Consequently, they can cohere with each other in just the same way that the two blue strands cohere or the two red strands cohere. This is evident in the middle part of the diagram, where the complementary bases are juxtaposed. The bases have been colored magenta, a color intermediate between red and blue, to indicate that they form a short segment of double helix all their own, part red, and part blue, just like the blue double helix and the red double helix. Under appropriate conditions (i.e., with sufficiently high concentrations of DNA and of cations, and at low enough temperatures), the structure in the middle of Figure 1-6 will cohere stably, and will form a single complex. If you want to make sure that the complex is really robust, it is straightforward to link the red strands to the blue strands covalently. This is done by ligating them enzymatically to form two total strands, rather than four. The ligated product is shown in the bottom panel of Figure 1-6, where there are now just two magenta strands, which are shown to be paired.

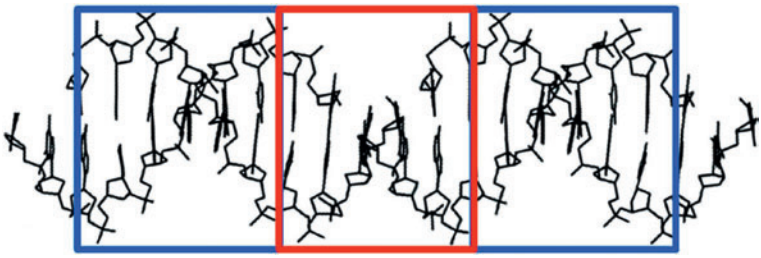

Figure 1-7 *A crystal structure showing sticky-ended cohesion.* This is the crystal structure of a self-complementary DNA decamer, but the dyad axis is displaced from the middle, creating a sticky end. Sticky-ended cohesion is shown in the red box in the middle, where the two gaps lacking phosphate groups are prominent. The important thing about this image is that the DNA in the red box has the same B-DNA structure as the DNA in the blue boxes, even though it is upside down, because it is a half-turn away. Thus, if one knows the coordinates of the DNA in the blue box on the right, one knows the coordinates of the DNA in the blue box on the left, even in solution. Reprinted by permission from Elsevier.[1.11]

The sticky-ended interaction is a very special affinity interaction. There are lots of affinity interactions in biology – for example, the binding of multi-subunit proteins to each other or the interactions of antigens and antibodies. However, all we know about those interactions is that there is affinity between the components of the complex. In every case, we need a crystal structure or some comparable analysis to know the details of which atoms are binding to which atoms.

Sticky-ended interactions are different in this regard, as shown in Figure 1-7. This drawing illustrates a crystal structure that forms an infinite DNA double helix in the horizontal direction.[1.11] It consists of two strands of DNA that are slightly offset from each other, so that the duplex is tailed in 2-base sticky ends. There are two gaps in the middle of the red box, places where one strand stops and the other strand hasn't started yet. These gaps delineate the position where the duplex is held together by sticky ends. The blue boxes outline roughly the same places in two successive unit cells. Their contents appears to be upside down from the material in the red box, but that is because they are a half-turn away. The key point, however, is that there is a great deal of similarity between the structure surrounding the place of sticky-ended cohesion and structures at the middle of the blue boxes (even if they are upside down). Consequently, if we know the coordinates of the atoms in the left blue box, we know the coordinates of the atoms in the right blue box. Thus, when two pieces of double helical DNA cohere, we know not merely that there is an affinity interaction but, on a predictive basis, what that interaction looks like structurally, without

having to do the crystal structure. The details of the structure may not be so well determined a double helical turn or two away, but, at least locally, we know what the local product structure will be.

This is a very powerful property, unique, to my knowledge, in macromolecular recognition. If we have six nucleotides in our sticky ends, there are 4^6 unique sticky ends that can be made from the conventional bases: 2048 pairs (less 64 self-complementary hexamers, or a total of 1984) of different interactions that can be programmed uniquely. When we combine the notion of sticky ends with branched DNA that can contain at least 12 branches, we can imagine making huge numbers of stick-figure objects and lattices of DNA, where the vertices correspond to the points of branching and the edges consist of double helical DNA.

A simplified example of this concept is illustrated for a quadrilateral in Figure 1-8. At the left of the picture is a branched junction whose four arms meet at right angles, just like an intersection. Each of the four arms is tailed in a sticky end X, its complement X′, Y, and its complement Y′. What the diagram shows is that four of these junctions can stick together to form a quadrilateral. In addition, there are many unsatisfied sticky ends on the outside of the quadrilateral. Thus, more junction molecules could bind to the outside of the

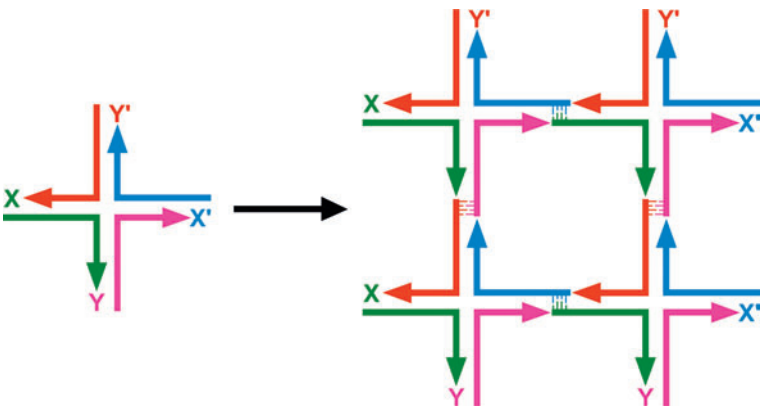

Figure 1-8 *The central concept of structural DNA nanotechnology: combining branched junctions into larger constructs.* The left side of the drawing contains a branched junction with four sticky ends, X, Y, and their complements, X′ and Y′. Four of these junctions have been combined to form a quadrilateral on the right. However, there are many sticky ends that are unsatisfied, so the construct could be extended, in principle, to form a lattice in 2D. In fact, since the ladder-like representation of DNA shown is not really appropriate to its double helical structure, it is possible to form 3D lattices from branched DNA molecules.

quadrilateral to form an infinite two-dimensional lattice. As we shall see, this system is not limited to two dimensions, but can in fact be extended to three dimensions, leading to infinite three-dimensional crystalline lattices. Figure 1-8 shows the fundamental notion of structural DNA nanotechnology: putting well-structured branched DNA molecules together to form lattices in two or three dimensions. There are many variations on this theme, going beyond lattices to devices and computation and other exciting topics. We shall explore them in the succeeding chapters of this book.

References

1.1 J.D. Watson, F.H.C. Crick, Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature* **171**, 737–738 (1953).
1.2 N.C. Seeman, Another Important 60th Anniversary, *Advances in Polymer Science* **261**, 217–228 (2013).
1.3 A. Rich, D.R. Davies, A New Two-Stranded Helical Structure: Polyadenylic Acid and Polyuridylic Acid. *J. Am. Chem. Soc*. **78**, 3548–3549 (1956).
1.4 R. Holliday, A Mechanism for Gene Conversion in Fungi. *Genet. Res.* **5**, 282–304 (1964).
1.5 J.R. Platt, Possible Separation of Intertwined Nucleic Acid Chains by Transfer-Twist. *Proc. Nat. Acad. Sci. (USA)* **41**, 181–183 (1955).
1.6 A. Gierer, Model for DNA and Protein Interactions and the Function of the Operator. *Nature* **212**, 1480–1481 (1966).
1.7 R.-I. Ma, N.R. Kallenbach, R.D. Sheardy, M.L. Petrillo, N.C. Seeman, 3-Arm Nucleic Acid Junctions Are Flexible. *Nucl. Acids Res.* **14**, 9745–9753 (1986).
1.8 Y. Wang, J.E. Mueller, B. Kemper, N.C. Seeman, The Assembly and Characterization of 5-Arm and 6-Arm DNA Junctions. *Biochem*. **30**, 5667–5674 (1991).
1.9 X. Wang, N.C. Seeman, The Assembly and Characterization of 8-Arm and 12-Arm DNA Branched Junctions. *J. Am. Chem. Soc.* **129**, 8169–8176 (2007).
1.10 S.N. Cohen, A.C.Y. Chang, H.W. Boyer, R.B Helling, Construction of Biologically Functional Bacterial Plasmids *in Vitro*. *Proc. Nat. Acad. Sci. (USA)* **70**, 3240–3244 (1973).
1.11 H. Qiu, J.C. Dewan, N.C. Seeman, A DNA Decamer with a Sticky End: The Crystal Structure of d-CGACGATCGT. *J. Mol. Biol.* **267**, 881–898 (1997).