Chapter 2. Weaver, 2/e. Mol Biol X107A.

Ch 2. The molecular nature of genes.

Reading notes

Also read (required): First sub-section of Sect 5.1, on gel electrophoresis. See Sect J, below.

We will spend no class time on Section 1 (The nature of genetic material), and little time on the basics in Section 2 (DNA structure). These are topics that should be part of your background, prior to taking this course. Look over these sections for things that may be new to you, and feel free to ask questions as needed. Some of the key points are highlighted below in this handout, to guide you. The homework will include some questions on this material, as a check.

Clark & Russell. Ch 4.

⇒ Please bring your DNA model to class.

A. Introduction

Highlights:

- The transforming principle, and the identification of DNA as the genetic material. Griffith; Avery et al; Hershey & Chase.

- The idea of informational macromolecules.
- What is DNA? - an introduction. Nucleotide monomers; double helix. Watson & Crick.
- There is a basic structure of DS DNA, called the B form.
- There are other structures. The Z and H structures and cruciforms are particularly distinctive.
- Supercoiling; a higher order structure.
- DNA structure has micro-heterogeneity, dependent on base sequence and on conditions.
- DNA structure is dynamic.
- Gel electrophoresis as a tool for separating DNA by size.
- Denaturation and renaturation of nucleic acids; secondary structure of SS nucleic acids.

B. Abbreviations and conventions for nucleic acid structures

⇒ For reference.

SS = single stranded; DS = double stranded; RH = right-handed; LH = left-handed; CCC = covalently closed circular; bp = base pairs; kbp = kilobase pairs. (kb = kilobases. It is appropriate for SS nucleic acids, but it is sometimes used carelessly when kbp is really intended.) Nucleic acid sequences (DNA or RNA) are written 5′→3′, unless noted. (For DS nucleic acids, the upper strand is 5′→3′.) You are encouraged to show strand polarity explicitly. In addition to the usual symbols for the bases, Py and Pu are used for pyrimidine and purine, respectively. If one letter codes are needed, Y and R are used.

C. String exercises

Exploring the properties of a piece of multi-stranded helical string will help you visualize the DNA structures and some of their topological properties (Sect F-H, below). We will provide you with a piece of RH string, and go through some exercises in class.

For this Ch, the major goal is to recognize the helix handedness, using the ‘“right hand rule”. I will also show you some features of DNA supercoiling using the string, but you are not responsible for that at this time.

D. DNA is the genetic material

⇒ No class time on this section, except as questions come up.

Almost absolutely true. The only exceptions are some viruses that use RNA as the genetic material (Sect 2.3). *DNA and RNA are informationally equivalent. (Important idea.)*

Two classic experiments led to the identification of DNA as the genetic material. The first was Avery's identification of Griffith's transforming principle as DNA -- a rigorous chemical determination. The second was the Hershey & Chase experiment with a bacterial virus.

> Term: Transformation. In bacterial genetics, transformation is the transfer of genetic information from one cell to another by using free DNA. (In cell biology, transformation has a different meaning; it refers -- loosely -- to acquiring the ability to grow like cancer cells.)

Griffith (Fig 2.2). The mutant organism that Griffith used presumably had a single mutation, but that mutation led to three distinct macroscopic phenotypic changes. The following table compares the phenotypes of Griffith's wild type and mutant organisms.

	wild type	mutant
virulence	yes	no
colony form	smooth	rough
capsule	yes	no
enzyme (to make capsule)	yes	no

Li et al (2001) is an example of recent work on natural bacterial transformation.

Hershey & Chase. Fig 2.4. They used a very simple system: a bacterial virus. They could label the virus by growing it in a medium containing radioactive phosphate or sulfate. It is good to follow the logic of this labeling procedure, how each label tracks one type of molecule.

> In each part of Fig 2.4 the labeled macromolecule is red. Key issue: Where is the red part at the final step (after infection)?

See Sect M for a web site that may be a useful supplement for this material.

E. Informational macromolecules

Macromolecules are made by polymerizing small molecule subunits. The biosynthesis of these subunits is the subject of traditional biochemistry.

Informational macromolecules: there are different kinds of monomer subunits, and their sequence is important.

There are two problems to address: How do we put the subunits together? How do we choose which subunits to use?

F. DNA

DNA is an informational macromolecule.

There are four kinds of subunits: the nucleotides A, T, G and C. Each nucleotide contains three parts: a phosphate group, a sugar, and a base. The phosphate group and the sugar are the same in all the DNA nucleotides. There are four different bases. The nucleotides contain a common region (the phosphate and sugar) and a variable region (the base).

> Nucleotides with one phosphate are also called nucleoside monophosphates; that should seem logical. Weaver uses the term nucleotide to mean "at least one phosphate"; Fig 2.9 and Glossary. I think that is somewhat confusing, but he is not entirely alone. In any case, the terms nucleoside and nucleotide are often mixed up; beware!
>
> Note the designations α, β, and γ for the three phosphate groups of a nucleoside triphosphate; Fig 2.9, right side. The α phosphate is the one that is retained in the polymer.

Useful diagram of the structure of one DNA strand:

```
—S—P—S—P—S—P        S = sugar, P = phosphate
   |     |     |
   B     B     B        B = base (A,T,G,C)
```

Fig 2.11 shows another type of shorthand for DNA; these shorthand structures show different levels of detail.

The polymerization reaction itself involves only the common part, the "sugar-phosphate backbone." The special (informational) properties are in the variable part, the bases.

Each nucleotide, and therefore also the polymer chain, is polarized, with a 5′ end and a 3′ end (Figs 2.7 & 10). DNA chains grow by addition of the 5′ end of the incoming nucleotide to the 3′ end of the growing chain (p 22; also see Fig 3.13). This is called growth from the 5′ end, or growth from 5′ to 3′.

The actual precursors used for DNA synthesis are the nucleoside 5′-triphosphates (Fig 2.9 right hand part and Fig 2.11a). The energy of the high energy (acid anhydride) bond in the precursor is retained in the polymer.

> [Not in book: The other product, pyrophosphate (PP_i) (shown in Fig 3.13a), is then hydrolyzed to two phosphates (P_i). The cellular concentration of PP_i is very low, thus preventing reversal of the biosynthetic reaction. Ho et al (2000) discuss another possible biological role for pyrophosphate.]

The common DNA structure is a double helix. The two strands run in opposite directions (i.e., are antiparallel; Fig 2.14). The charged and polar backbone of the DNA chains is on the

outside, and the bases are on the inside. The bases are "stacked" (hydrophobic interactions) and "paired" (hydrogen bonding). (We will discuss hydrogen bonding and hydrophobic interactions in Ch 3.)

RNA is similar to DNA. The sugar in RNA is different from the sugar in DNA, and one of the bases is different. However, *RNA and DNA are informationally equivalent.* (The other remarks about DNA hold generally for RNA, except that RNA is more commonly found in the single stranded form. This reflects its biological roles, not anything inherent in the polymer.)

Schöning et al, 2000, discuss the sugar part of the nucleic acid structure, from a chemistry viewpoint.

Choosing the subunits. The sequence of bases in the DNA chain dictates the sequence of bases in the new DNA chain (or new RNA chain). Base A must pair with T (or U, in RNA), and G with C. The pairing is based on hydrogen bonds. (See Matray & Kool, 1999, for an unusual variation.)

> [The sequence of bases in the DNA (and RNA) is used to determine the sequence of amino acids in the proteins. We will discuss the mechanism of this complex translation process later (Ch 17-19; briefly noted in Ch 3). Base pairing is part of the process.]

Developing the DNA structure. Figuring out a molecular structure from X-ray data is a trial and error process: propose a structure, calculate the predicted X-ray pattern, then compare it with the data. What makes it practical is good guessing. (Lucas et al, 1999, discuss the X-ray analysis.)

Watson & Crick had the X-ray data (Fig 2.12), which indicated a regular helical structure. They knew the general composition of DNA, the structure of the nucleotides, and their propensity to hydrogen bond. As their work was in progress, they learned how the nucleotides are linked together.

But in addition, Watson & Crick had the insight that complementarity would be a good basis for replicating genetic material. And they knew that A = T, G = C.

General scheme for replication. The proposed structure implied a scheme for replication, called semiconservative (Fig 2.15). The term means that each new DNA molecule is half new, half old. The Meselson & Stahl experiment demonstrated that DNA replication is indeed semiconservative (Fig 20.2).

G. DNA structures

Fig 2.16 shows the first three nucleic acid structures we discuss: the B, A and Z forms.

B-DNA. The basic structure of DNA is the B form; most DNA most of the time is in (or near) the B form. You should have a general understanding of the structure of B-DNA; most discussion of DNA structure uses B-DNA as a reference point.

As noted earlier, the two strands are antiparallel; this holds for all common DS DNA structures.

The major and minor grooves of DS DNA are defined in the context of B-DNA. Weaver introduces these grooves in Fig 2.14bc.

> You might also peek ahead to Fig 9.19, which shows that the base pairs are more visible and more easily discriminated in the major groove. Later, we will see that proteins that recognize specific DNA sequences most often (but not always!) read the major groove. Henikoff & Vermaak (2000) discuss drugs that bind to the minor groove.

A-DNA. The general form is similar to B-DNA. Think of a B helix, then squash it a little. The helix is shorter and fatter. Also, the bases are more tilted.

Although the general form may look very similar, the structural details are quite different. From the point of view of a DNA-binding protein looking for specific contacts, A and B DNA are quite distinct.

> [The 2′-OH group of RNA precludes the B conformation. Therefore, DS RNA and (probably) RNA-DNA hybrids assume the A conformation. p 28.]

Z-DNA. This DNA structure is radically different: the helix is LH. The backbone has an unusual zigzag structure. There are 12 bp per turn.

> [Not in book.] The structural basis of the zigzag backbone is that alternate sugars are rotated differently with respect to the base. [The base is always (approximately) perpendicular to the helix axis.] In chemical terms, alternate nucleotides are in the syn conformation (instead of anti). [I'll show a model.] As a result, the structural repeat is two consecutive base pairs; X-ray analysis shows 6 repeating units per turn.

[Not in book.] Factors that promote the formation of Z-DNA…

- Certain base sequences, especially alternating Pu-Py. (Only purines, especially G, can assume the syn conformation, which is required at alternate positions.)

- Methylation (to 5-methyl-C), within the above sequences. (The reason for this is understood. The extra methyl group is better removed from the surrounding water in the Z structure.)

- The stress of underwinding (see Supercoiling, below).

- Z-binding proteins might stabilize Z-DNA.

- High salt concentrations. (Useful in vitro.)

Schwartz et al (1999) suggest they have uncovered a natural role for Z-DNA. Mao et al (1999) have invented an unnatural role for Z-DNA, making use of its sensitivity to salt concentration.

See Sect M for more about left-handed DNA. ☺

Cruciforms [not in book]. Cruciform structures may form when there are inverted repeats (palindromes; see Glossary) in the DNA sequence. Cruciform structures are another way to relieve the stress of underwinding.

> (When a palindrome sequence is transcribed into RNA, the SS RNA forms a hairpin, which is a "half-cruciform".)
>
> (You may see cruciform-like structures form in your string model when it is underwound.)

Triplex DNA; H-DNA [not in book]. Polypurine:polypyrimidine sequences have unusual properties, including sensitivity to nucleases considered specific for SS DNA. These sequences can form a novel and distinctive structure, now called H-DNA. Its characteristic feature is a triple strand - single strand structure. This provides a "hinge" in the DNA. Wells et al (1988) review some of the other factors, along with the polyPu:polyPy sequence, that affect whether DNA is H or B. Their Fig 1 (attached, p 22) shows a diagram of H-DNA, and the hydrogen bonds that are involved. Note that the H-bonds to the third strand involve different groups, but still provide the same AT and GC pairings as the common Watson-Crick rules. These pairings involve the 5-membered ring of the purine bases, and are called Hoogsteen pairings. (Vasquez & Wilson, 1998.) Also note that the two pyrimidine strands in the triplex are anti-parallel to each other.

The general idea of triple-stranded (triplex) DNA structures is now well accepted, although the natural significance is not clear. Not all triplex DNA is H-DNA. Other triplex structures for DNA are known (or proposed) (Vasquez & Wilson, 1998).

Sakamoto et al (1999) may be an example of the relevance of triplex DNA -- or more complex related structures -- to human disease. Wright (2000) discusses the possible role of unusual DNA structures in mutation. These two papers may deal with fundamentally related issues.

Allemand et al (1998), Sakamoto et al (1999) and Wolf et al (1999) describe novel DNA structures. Nielsen (1999) discusses the useful synthetic nucleic acid analog PNA.

Braun et al (1998), Seeman et al (1998), Mao et al (1999), Clelland et al (1999), Mao et al (2000) and Giese et al (2001) explore issues that may bear on possible technological applications of DNA. Even if the applications may appear fanciful, these papers are often fun, and reveal fascinating aspects of DNA. (One question on last year's test -- your sample test -- is another example of a "technological" use of DNA; the reference is on the answer key.)

H. Supercoiling

Weaver briefly introduces DNA supercoiling on p 34; also see Fig 2.25.

Supercoiling is a higher order DNA structure. It is most easily understood with CCC DNA, as shown in Fig 2.25. Each of the two DNA strands is a closed circle, and the two strands are intertwined. If the twist of the main helix is changed, something else must change, by an equal amount in the opposite direction, to compensate. That something is "supercoiling".

DNA supercoiling will come up from time to time during the course. As a preview, I will show you a little about supercoiling, mainly using the string.

There are two forms of supercoiling:

- In toroidal (or solenoidal) supercoiling, the main DNA helix wraps around a cylindrical core. (Also called restrained.)

- In interwound (or plectonemic) supercoiling, the main DNA helix winds up and back down around a supercoil axis. (Also called unrestrained.)

Weaver does not explicitly address the two types of supercoiling. He introduces supercoiling in Fig 2.25; this is interwound supercoiling. He also discusses wrapping of DNA around the histone core in nucleosomes in Ch 13, and briefly notes that the wrapping is LH (p 383); we will probably not discuss this much.

DNA is commonly underwound in the cell. That is, if we isolate CCC DNA from cells, it is negatively supercoiled. In bacteria, an enzyme called topoisomerase (Ch 6) is probably responsible for this. In higher organisms, the common negative supercoiling results largely from the toroidal supercoiling on the nucleosomes.

Although nucleosomes provide the most common form of toroidal supercoiling, we will also see examples of DNA regions wrapping around an individual protein. This, too, induces supercoiling. Kimura et al (1999) discuss that this is useful. More in Ch 9.

With linear DNA (or circular DNA that is not covalently closed on both strands), it is less obvious that there should be unrestrained supercoiling. After all, the closed constraint does not exist. Nevertheless, it is clear that DNA is commonly supercoiled. Apparently other constraints exist, presumably due to the organization of the genome into topologically isolated domains.

Supercoiling, Z-DNA, and cruciforms. Z-DNA is LH; Z-DNA is underwound compared to ordinary B-DNA, because it is wound in the opposite direction. Thus negative supercoiling and Z-ing are alternative ways to relieve the stress of underwinding in DNA. Similarly, cruciforms are another alternative to relieve the stress of underwinding.

I. DNA structure is variable

We have hinted above how the structure of DNA may vary from one condition to another. However, we need to make an even stronger point… Even under "uniform conditions", DNA structure is not entirely constant along the length. This is chemically reasonable, since the DNA polymer is a heteropolymer. Although normal DNA may be more or less B-form, the details vary depending on the base sequence. The properties listed in Table 2.2 (p 28) are averages. DNA-binding proteins may make use of such variation to make sequence-specific contacts.

Further, as we would expect for any real structure held together with "weak bonds" (such as hydrogen bonds; Ch 3), the structure may fluctuate with time. DNA breathes and bends, and may even shift from one basic structure to another.

Rohde et al (1999) is an example of such DNA structural variability or flexibility.

Perspective. All of the non-B structures are probably less stable than B-DNA -- for unstressed DNA. But much DNA is stressed. The non-B structures can be considered as alternative ways to relieve the stress of supercoiling. It then follows that anything that changes the stress (the supercoiling) may also change the DNA structure, with whatever implications that has.

J. Gel electrophoresis

Gel electrophoresis is a major tool for analyzing DNA. Weaver mentions it on p 34, but describes it in Sect 5.1; that sub-section is assigned. This procedure separates DNA by size, and is calibrated by running "markers" (samples of known "size" = length = molecular weight). Figs 5.1-2.

> [Electrophoretic mobility is affected by shape as well as by size. For example, circular DNA runs faster than linear DNA of the same size, and circular DNA that is supercoiled runs even faster (it is more compact). Further, some DNA sequences have non-standard shape, and therefore run anomalously. It is a tacit assumption in most analyses that the gel is measuring only size.]

The gel electrophoresis conditions used in DNA sequencing allow single base resolution; e.g., Fig 5.18-19.

Pulsed field gel electrophoresis (PFGE) is a technical variation of the basic method that allows analysis of much larger DNAs. Fig 5.3. PFGE reduces the problem of the DNA chains tangling with the gel matrix.

Han and Craighead (2000) describe a novel system for DNA separations.

K. Denaturation and renaturation of nucleic acids

Weaver spends considerable time on this topic, pp 29-36. Historically, $c_o t$ analysis was an important way to determine genome sizes. It also revealed the existence of large amounts of repetitive DNA, which -- along with other non-coding DNA -- accounts for much of the "excess" DNA content of some organisms (C-value paradox, p 35; mentioned in Ch 1 handout; also see Petrov et al, 2000). But $c_o t$ analyses and repetitive DNA analyses are not priority topics for us.

What we want to get from this is a general sense of how nucleic acid denaturation and renaturation occur. In one form or another, these are still very important techniques.

A good place to start is Fig 2.17. This illustrates a nice well-behaved melting curve for a homogeneous DNA sample. T_m, the melting point of the DNA -- the midpoint of the transition, is a well-defined parameter, which depends on the base composition of the DNA (Fig 2.18) and the conditions used for the measurement (especially ionic strength). Fig 2.17 also illustrates one easy way to measure denaturation, using the hyperchromic shift (p 29), the increase in absorbance upon strand separation.

The denaturation reaction is, in principle, fully reversible. The reverse reaction is variously called renaturation or annealing or -- especially when carried out with a mixture of different kinds of nucleic acid -- hybridization. Logically, the renaturation reaction is the same whether DNA strands, RNA strands, or a mixture of DNA and RNA strands are used.

The renaturation reaction is bimolecular, and thus has two characteristics not shared with the denaturation reaction.

- The rate of the renaturation reaction depends on concentration of the nucleic acid sequences.

- The reaction is subject to competition.

The concentration dependence is the key point behind $c_o t$ analysis. The larger the genome, the fewer molecules are present in a given mass of DNA. Therefore, the rate of renaturation is inversely proportional to genome size (Fig 2.20; Fig 2.23 with "problem 3", p 32). Further, repetitive DNA sequences renature faster, precisely because their concentration is higher (e.g., Fig 2.21).

The T chosen for renaturation is important, because of the complexity of the nucleic acid interactions. A useful rule of thumb is to anneal at T_m - 25 (in °C); Weaver explains why on p 30.

> If hybridization is done at a lower T, there is more chance for imperfect hybrids to be retained. We call such hybridization conditions "non-stringent." (See Glossary for "stringency".) Sometimes we do this intentionally, because we want all "related" sequences.

Relevant here is the structure of SS nucleic acids. At low T, SS nucleic acid chains can fold back and form a variety of structures. Some are just a nuisance, as in the context of annealing. Some are important, as when we look at the structure of tRNA, and even other RNA species. Schwartz et al (1999) illustrate the importance of mRNA secondary structure.

The bigger point behind this is that renaturation/hybridization has become a major tool for molecular biologists, to measure how well two nucleic acid samples match. We will see numerous applications of one or another variation as we continue.

L. Further reading

R D Wells et al, The chemistry and biology of unusual DNA structures adopted by oligopurine-oligopyrimidine sequences. FASEB J 2:2939, 11/88. Review. H-DNA. See Fig, p 22.

K M Vasquez & J H Wilson, Triplex-directed modification of genes and gene activity. TIBS 23:4, 1/98. Discusses various kinds of triple-stranded DNA, and possible application by adding a third strand that complexes with DS DNA in a sequence-specific manner. Such triplex formation can affect transcription directly. It can also be used to target a mutagen that is attached to the incoming third strand. They mention work suggesting that triplex DNA has been found in normal cells.

E Braun et al, DNA-templated assembly and electrode attachment of a conducting silver wire. Nature 391:775, 2/19/98. Engineers use DNA, too. In this case, they identify two electrodes that are a few μm apart by tagging them with DNA "probes". They then connect the electrodes with a piece of DNA, with appropriate complementary sequences at each end. Finally, they deposit silver along the DNA, to make a wire connecting the electrodes. Also see a short note by S Bains, Optoelectronics: Double helix doubles as engineer. Science 279:2043, 3/27/98. See hw. [As you compare this work with that of Giese et al (2001), note that the latter deal with the intrinsic conductive properties of DNA. In contrast, Braun et al merely use the DNA as an engineering tool; conductivity is due to the silver.]

N C Seeman et al, New motifs in DNA nanotechnology. Nanotechnology 9:257, 9/98. Many are fascinated by the possibility of using DNA in nanotechnology, exploiting the base pairing specificity of DNA to help assemble structures. A cube, an octahedron, knots (of defined structure), Borromean rings. Some may seem frivolous, but the goal is learning how to make useful structures with DNA, making use of the complementarity. A key issue they address is trying to make rigid structures; they report some success with "double crossover"-type molecules. In doing some work with RNA, they found that one topoisomerase can act on RNA. If nothing else, cute pictures. From an X107A student. Also see Mao et al (1999), below.

J F Allemand et al, Stretched and overwound DNA forms a Pauling-like structure with exposed bases. PNAS 95:14152, 11/98. (+ Commentary, Rich, p 13999.) P-DNA is made by stretching a molecule of overwound B-DNA. The P-DNA has the bases sticking outward -- a structure once proposed (incorrectly) by Pauling. A similar structure has been found in a virus. Rich's commentary more broadly discusses work done by manipulating and taking physical

measurements on single DNA molecules (also recall Vale and Milligan, 2000, Ch 1 FR; and see Liphardt et al, 2001, below).

C Mao et al, A nanomechanical device based on the B-Z transition of DNA. Nature 397:144, 1/14/99. Here, they make a mechanical switch, based on the conformational change between B and Z DNA, induced by change in salt conditions. Also see Seeman et al (1998), above.

A A Lucas et al, Revealing the backbone structure of B-DNA from laser optical simulations of its X-ray diffraction diagram. J Chem Educ 76:378, 3/99. They describe an optical simulation of the X-ray analysis of DNA, to help students visualize how the X-ray data revealed features of the DNA structure.

N Sakamoto et al, Sticky DNA: Self-association properties of long GAA·TTC repeats in R·R·Y triplex structures from Friedrich's ataxia. Mol Cell 3:465, 4/99. A hot subject is the role of triplet repeats in human diseases. Changes in the number of repeats may have severe consequences. The change in number of repeats is probably due to slippage during replication. It has been shown, in one case or another, that the relevant repeats can form stable triplex structures or hairpins. Now they show that the same features that lead to triplex structures can entangle more than one region of the chromosome.

C T Clelland et al, Hiding messages in DNA microdots. Nature 399:533, 6/10/99. (+ Erratum, 402:750, 12/16/99.) Do you analyze your mail with PCR (polymerase chain reaction, Sect 4.2)? Hm, they should have arranged to publish this on page 007.

T Schwartz et al, Crystal structure of the Zα domain of the human editing enzyme ADAR1 bound to left-handed DNA. Science 284:1841, 6/11/99. Weaver notes that the natural presence or role of Z-DNA remain unclear. Here, Rich's group explores the structure of an enzyme, DS RNA adenosine deaminase, that binds specifically to Z-DNA. The enzyme deaminates A residues in DS RNA, including in hairpin regions of mRNA. The Z-DNA binding activity could couple this enzyme to nascent mRNA, by binding to the Z-DNA region that may form behind the RNA polymerase for certain genes; Ch 6. Since deaminated A has the coding specificity of G, this action could result in translational variation. (In general, RNA editing involves the post-transcriptional changing of individual bases in an RNA molecule. Weaver introduces one type of RNA editing in Ch 16; we won't discuss this. DS RNA adenosine deaminase is part of a different type of RNA editing process, in which A's are deaminated to make I's.)

T J Matray & E T Kool, A specific partner for abasic damage in DNA. Nature 399:704, 6/17/99. Although we normally think of hydrogen bonding as the major issue in determining which base gets inserted during replication, simple steric factors are also important. Bases with poor H-bonding are incorporated with good specificity, if they fit well. Here, they study a new "partner" that pairs specifically with "nothing" -- an abasic site, with no base. The new "partner" is very large, and essentially occupies the entire base pair space. They use this new partner to determine abasic sites by "ordinary" sequencing.

P E Nielsen, Peptide nucleic acids as therapeutic agents. Current Opinion in Structural Biology, 9(3):353-7, 6/99. Review. This paper deals with PNA = peptide nucleic acid, a synthetic compound with nucleic acid bases attached to an uncharged, peptide-like backbone.

Interestingly, PNA molecules form DNA-PNA hybrids and PNA duplexes -- that are stronger than natural DNA duplexes. They can also form triplex structures with DNA. This increased strength is presumably due to the lack of charges on the backbone. Partly because of the strength of hybrids, PNAs are getting attention as possible “anti-sense” agents to regulate gene function by binding to a natural nucleic acid in the cell.

S G Wolf et al, DNA protection by stress-induced biocrystallization. Nature 400:83, 7/1/99. The Dps protein of E. coli co-crystallizes with DNA -- in vitro and in vivo. Since this protein is induced by stress, they suggest that the crystallization may be a protective response.

K Kimura et al, 13S condensin actively reconfigures DNA by introducing global positive writhe: implications for chromosome condensation. Cell 98:239, 7/23/99. An example of DNA wrapping around a protein. In this case, the protein induces restrained supercoiling, and this is part of the process for condensing the DNA into the complex eukaryotic chromatin (Ch 13).

J R Rohde et al, The Yersinia enterocolitica pYV virulence plasmid contains multiple intrinsic DNA bends which melt at 37 °C. J Bact 181:4198, 7/99. An example of the importance of DNA microstructure. The virulence gene becomes active as the T is raised to 37 °C. This happens because a specific DNA site “melts”.

G Poinar, Ancient DNA. Amer Sci 87:446, 9/99. A discussion of the issues involved in trying to analyze DNA from samples that are hundreds or thousands of years old.

D A Petrov et al, Evidence for DNA loss as a determinant of genome size. Science 287:1060, 2/11/00. (+ News, Capy, p 985. Also see a related but somewhat broader news item: A S Moffat, Genetics: Transposons help sculpt a dynamic genome; Science 289:1455, 9/1/00.) We mentioned the C-value paradox in the Ch 1 handout, Sect C; Weaver discusses it on p 35. Why do some organisms have large amounts of “junk” DNA? It might involve rates of acquisition or rates of loss -- or both. Here, they show in one direct comparison (crickets vs fruit flies) that the organism with more junk DNA has a demonstrably slower rate of elimination of dead transposons.

R F Service, Tracking DNA when the chips are down. Science 288:427, 4/21/00. Part of a larger article, discussing some items from a recent ACS meeting. This item deals with DNA labeled with a radio receiver. See homework.

J Han and H G Craighead, Separation of long DNA molecules in a microfabricated entropic trap array. Science 288:1026, 5/12/00. They suggest that this system is an order of magnitude faster than PFGE for separating large DNAs.

B E Wright, A biochemical mechanism for nonrandom mutations and evolution. J. Bacteriol 182:2993-3001, 6/00. Minireview. An unusual phenomenon, discovered about a decade ago, is that mutations sometimes appear to be non-random. The basis for such environment-affected mutations is still not entirely clear, but the general idea may be that conditions promote a burst of hypermutation that produces increased variability. One part of this may relate to unusual DNA structures.

A M Ho et al, Role of the mouse ank gene in control of tissue calcification and arthritis. Science 289:265, 7/14/00. (+ News, Hagmann, p 225.) Why is a paper on arthritis listed in the DNA chapter? They provide evidence that the levels of pyrophosphate -- a natural chelator -- may be involved in arthritis. They suggest that the mutant mice are unable to transport pyrophosphate into the joints, thus allowing calcification to proceed.

C Mao et al, Logical computation using algorithmic self-assembly of DNA triple-crossover molecules. Nature 407:493, 9/28/00. Example of a developing field of using molecules for computation.

K M Vasquez et al, Specific mutations induced by triplex-forming oligonucleotides in mice. Science 290:530, 10/20/00. Here they use a triplex-forming oligonucleotide (TFO) to induce targeted mutations in vivo. Mutations are apparently induced during repair of the novel structure. Many of the mutations are small insertions or deletions in regions already considered slippage-prone (e.g., repetitive bases). Thus this system targets mutagenesis to a small region, but does not induce a specific type of mutation.

K-U Schöning et al, Chemical etiology of nucleic acid structure: the α-threofuranosyl-(3′→2′) oligonucleotide system. Science 290:1347, 11/17/00. (+ News, Orgel, p 1306.) They explore the base pairing properties of nucleotides based on a variety of sugars. The underlying question is why Nature chose ribose and deoxyribose. In an earlier paper, they showed that nucleotides based on arabinose would have stronger hydrogen bond pairings. Here, they show that nucleic acid structures could be formed with tetrose sugars. At first, this was a surprising result, since it seemed that the backbone would be too short. But it works. And the simpler tetrose sugars might well have been more abundant in the “primordial soup”, so this result may have implications for the nature of early life.

S Henikoff & D Vermaak, Bugs on drugs go GAGAA. Cell 103(5):695, 11/22/00. Minireview. Discusses drugs that bind to the DNA minor groove, with some sequence specificity. The title alludes to some recent work in which one such drug caused effects in Drosophila similar to homeotic mutations.

E Nir et al, Pairing of isolated nucleic-acid bases in the absence of the DNA backbone. Nature 408:949, 12/21/00. We would generally not expect to see base-pairing interactions between isolated bases, i.e., in chains of length = one. This is because the strength of individual H-bonds is quite low, and there is competition with the solvent. Here, they do observe such interactions -- in the gas phase. Their method also creates complexes of 1 or 2 bases with a small number of water molecules.

A Aderem & D A Hume, How do you see CG? Cell 103(7):993-996, 12/22/00. Minireview. How does the mammalian body recognize a bacterial invader? One part of the story is that immune cells recognize unmethylated CG sequences, which are relatively common in bacteria but relatively uncommon in mammals.

Y Li et al, Natural genetic transformation of Streptococcus mutans growing in biofilms. J Bacteriol 183:897-908, 2/01. An example of recent work on natural transformation. S mutans is a major organism in dental plaque. The bacteria use a peptide “pheromone” to induce

competence (ability to be transformed) in the population; this induction is most effective in the dense biofilm.

J Liphardt et al, Reversible unfolding of single RNA molecules by mechanical force. Science 292:733, 4/27/01. (+ News, Fernandez et al, p 653.) Another example of single molecule biophysics (Allemand et al, 1998, above), this time with RNA. They show that as one increases the force pulling on an RNA hairpin, it suddenly unfolds cooperatively. That is, the hairpin sequence seems to exist in only two states, entirely folded and entirely unfolded.

B Giese et al, Direct observation of hole transfer through DNA by hopping between adenine bases and by tunnelling. Nature 412:318, 7/19/01. The DNA double helix contains a stack of π-electron systems -- arguably not a very good stack. Can it conduct electricity? Experimental results vary, with experimental design, and theory is confusing. But the answer is at least maybe and sometimes. This paper shows that there are two distinct mechanisms of charge propagation, depending on the base sequence. Why do we care? Electron conduction through DNA could be relevant to DNA damage from excited electrons; is the damage only local, or can it be transmitted some distance away from the original activated site? Further, conducting DNA might have technological uses.

M. Computer resources

Visit the NanoTheater, from Digital Instruments (Santa Barbara, CA).
http://www.di.com
DI makes atomic force microscopes (AFM) and such, and shows off what such instruments can see in this theater. Images include individual atoms, DNA molecules, and more.

RasMol is a program for viewing 3-dimensional molecular structures on your own computer. It was originally written for viewing protein structures, but now thousands of files for RasMol are available -- of proteins, nucleic acids, and small organic chemicals. You can even make your own files for RasMol.

I have a page at my web site, “RasMol - An Introductory Guide”. This page will help you get RasMol and get started using it, with both small molecules and proteins. It includes some good sources of files to view. And it includes a section on how to make your own files for RasMol -- using free software (only practical for small molecules, but it is fun to do).

Also, my page “Internet - Miscellaneous” has links to sources of RasMol and files to view in RasMol, in the section “Chemistry: Drawing and viewing molecules”.

I encourage you to get RasMol, and get started with it, as we do Ch 2-3. You will find it useful from time to time throughout the course. Please let me know of any problems.

http://www.profiles.nlm.nih.gov
A new feature, at the National Library of Medicine… “Profiles in Science”, featuring the papers of great 20th century scientists. The first scientist featured here is relevant to this Ch,

Oswald Avery (Sect D). The second is Joshua Lederberg, who made pioneering discoveries in bacterial genetics. Currently, there are six Profiles; more are added from time to time.

http://www-lecb.ncifcrf.gov/~toms/LeftHanded.DNA.html
☺ This site discusses misprintings of the normal DNA structure that make it seem left-handed. It shows some examples. This is at Tom Schneider's site, at the National Cancer Institute. If you will step back to his home page, you will also find information on the 'logo' presentation of DNA consensus sites that we will show later (for ribosome binding sites and promoters). [This site was listed in a web feature in Science (Nov 13, 1998, p 1223) -- with a wrong address.]

N. Homework

Work on the questions at the end of the chapter. Consistent with the coverage discussed above, #8-9 are not class material. Bring any questions you have. We may discuss some of them in class, but may not unless you have questions. In any case, come to class prepared.

A few of my own questions…

1. Nucleic acids contain 4 bases: A, T (or U), G, C. We can describe the base composition of a nucleic acid by giving the (mole) percentage of each base.

a. For a single-stranded nucleic acid, such as ordinary RNA, what is the minimum number of bases for which we must state the composition in order to know the complete base composition?

b. How about for a double-stranded nucleic acid, such as ordinary DNA?

c. A particular DNA contains 60% GC (i.e., 60% G + C). Calculate the percentage of each base in this DNA.

2. The normal infection cycle of a bacterial virus such as T2 (Fig 2.4) ends after about 30 min when the infected cells lyse, liberating about 100 progeny virus per infected cell. However, it is possible to break open the infected cells artificially, and then assay for viruses in the artificial lysate.

Assume that you have 10^8 bacteria infected with T2. If you lyse the bacteria at, say, 29 min after infection, you will find about 10^{10} viruses in the artificial lysate. If you do this at 5 min after infection, you will find very few viruses, far fewer than the 10^8 that you know you started with. Why?

3. There are many proteins that recognize DS DNA with great sequence specificity. Many do this via the major groove. Would you expect this to work better for the A-form or the B-form? Explain, based on the structures (not on how common they are).

4. Consider the E. coli chromosome.

a. How many turns of the DNA helix are there in the chromosome?

b. In DNA replication the two strands must unwind from each other to produce separate daughter helices. Assuming for the moment that there is only one swivel point at which this unwinding occurs, what unwinding rate would be required, in revolutions per minute? Use 40 min as the time for one round of replication.

5. Ordinary B-DNA has ~10 base pairs per helix turn (Table 2.2; in fact, the average is closer to 10.5 bp/turn). If you find that a particular piece of DNA has 11 bp/turn, is it "overwound" or "underwound", compared to normal B-DNA? Explain.

6. Ethidium bromide (EB) is a planar molecule that can bind to DNA by intercalating between the bases, thus unwinding the DNA a bit as it binds. Which should bind more EB, relaxed DNA or CCC DNA? Explain.

7. We referred to renaturation reactions as being bimolecular. However, in a sense that isn't true if the reaction involves two regions of the same single strand.

a. Give an example of a SS sequence which could do such a reaction. (To focus on the idea, use a four base region that forms the hybrid. In practice, that usually too short to be effective.)

b. What is the term for such a structure? For the DS DNA sequence that leads to this?

c. How do the features of this reaction compare with those stated above (Sect K) for ordinary renaturation reactions?

8. In a SS nucleic acid, is it possible that a G base might H-bond with another G base? Is it possible that it might H-bond with more than one other base at the same time? Explain. Give a specific example of how two G's might H-bond together.

9. You are studying "an interesting gene", and are testing whether DNA of other organisms will form hybrids with DNA for your gene. If you wanted to make sure that you got hybrids from genes that are only "similar" (as well as from those that are "identical"), would you use a somewhat higher OR lower hybridization temperature? Explain.

10. Surprisingly, T-C mismatches are moderately stable, and do not cause a distortion of the backbone. Suggest how these two bases might pair without backbone distortion, given a piece of DNA in aqueous solution. (Use Fig 2.13 to guide you. Don't worry about exact dimensions, but try to show a plausible structure with approximately the right spacing of the bases.)

11. In higher organisms, mRNAs are occasionally modified after being made. One type of modification changes C to U or A to I. (Weaver shows the structure of inosine, Fig 19.32, in the context of tRNA where I is common.)

a. What type of chemical reaction is this?

b. Is it oxidative?

12. A continuing trend is miniaturization. As equipment becomes smaller, it is harder to figure out how to position things specifically. As a molecular biologist it might occur to you that you could use DNA to position individual devices on a microscopic surface. Describe how you would do this. You can assume that you can attach specific DNA chains to things "as needed".

13. A company has announced a procedure for attaching a radio transmitter to a DNA molecule. Why would this be useful to you in a molecular biology lab? (For those who want to think about this beyond the superficial… think about why it might be easier to use DNA with radio transmitters attached than to use gene chips.)

14. a. A "bit" is a basic unit of information, representing one binary (yes/no) choice. How many bits of information are contained in one DNA base?

b. Compare the amount of information in the E. coli genome to the storage capacity of a common personal computer disk. (If you're not into information theory, feel free to skip down to the answer.)

15. Improve the chapter summary. If you were going to add one more thought, what would you add? Would you change or delete anything? (This question, which can be considered for any chapter, is a chance to review our perspective on each chapter.)

O. Partial answers

1. a. We want to know four base percentages. However, there is one relationship between them: A + T + G + C = 100%. Thus we need to state only three of the base percentages; we can then calculate the fourth.

b. Now there are two additional relationships: A = T and G = C. That is, there are three relationships among the four base percentages. Thus we need specify only one base percentage, and all the others can be calculated.

c. This is an example of the point made in part b. The question must refer to DS DNA; it wouldn't make sense for SS. G + C = 60%, but G = C; therefore G = C = 30%. A + T must be the rest -- the other 40% -- and they are equal; so A = T = 20%.

> %GC is a common way to express DNA composition. It is, in effect, the percentage of GC base pairs. Some easily measured properties of DNA vary linearly with %GC. These include the temperature at which the strands separate ("melting temperature", Fig 2.18), and the density of the DNA in a CsCl density gradient (the kind used by Meselson & Stahl; Fig 2.19).

2. The virus life cycle begins by injecting only the DNA into the host cell. Thus there are no intact viruses early in infection -- not even the infecting viruses. (This stage of the viral life cycle is called the eclipse.) This contrasts with ordinary cell replication, in which cells grow and divide, but remain viable at all stages.

3. B-form. There is very little groove in the A form; Fig 2.16.

> Recall that DS RNA is commonly in the A form. No sequence-specific DS RNA-binding proteins are known.

4. Need some data? The helix parameters are in Table 2.2; remember that we are dealing with "normal DNA" -- the B helix. The size of the E. coli genome is in the Ch 1 handout.

The final answer to part b is about 10,000 rpm! As will become clearer in Ch 20-21, unwinding is not limited to a single swivel.

5. Underwound.

If you said overwound, because 11 > 10… that's not the correct comparison. You need the reciprocal of those numbers. 10 bp/turn (i.e., per 360°) is 36°/bp. 11 bp/turn is only 33°/bp, less wound than normal B DNA.

If the math doesn't convince you, try it… Take the string and overwind it, i.e., wind it "tighter". Then look; you will see that the strands wrap around the helix axis more rapidly, corresponding to fewer bp per turn.

6. EB unwinds DNA as it binds. If the DNA is CCC, the unwinding must be compensated for by a positive increase in supercoiling. As a result, there is a limit to how much EB CCC DNA can bind (until it can't supercoil anymore). Therefore, relaxed DNA can bind more EB.

The reduced binding of EB can be used to separate CCC DNA from relaxed DNA. EB reduces the density of DNA in a CsCl density gradient (Fig 2.19). Since less EB binds to the CCC DNA, its density is shifted less.

> EB is commonly used to "stain" DNA in gel electrophoresis. See Fig 5.3 for an example. Fig 5.1b probably also uses EB.

7. a. AAGGxxxxCCTT. The four x represent a spacer, so the bend isn't too sharp. Watch directions. Remember that the DS reign that is formed must have antiparallel strands, so the two regions shown here that will form the hybrid must run in opposite directions.

b. hairpin; palindrome

c. Not concentration dependent, since the two regions forming the hybrid are in the same molecule.

Subject to competition in principle. However, given that the intra-molecular complement is so close, it is likely to dominate.

> Practical application: Primers made for PCR must be designed so that they are not self-complementary.

8. Yes and yes. The normal base pairing rules are for DS helices, which provide a structural constraint. Free of that structural constraint, lots of things can happen. (The structure of H-DNA illustrates the latter point.)

Once free of the constraint of the double helix, all you need to do is to identify H-bonding donor and H-bonding acceptor groups in the G. The H-bonding donors are the H on the $-NH_2$ group and the H on the ring N just "above" that in Fig 2.5. The H-bonding acceptors are the lone pairs on the =O or on any of the N. Fig 2.13 shows you examples of these groups; all you need to do is to make a version involving two G's.

9. Lower. The poorer the match, the lower the T_m. This is the essence of the distinction between "stringent" and "non-stringent" hybridization (see Glossary).

Think about… why might you choose this strategy, to do a search at low T_m?

10. If you're having trouble, start reading this for an idea, but break away from it as soon as possible and try to continue on your own. Logical flow… Two pyrimidines aren't supposed to pair because they are too small. Of course, they might pair if the helix backbone dimensions could be distorted, but that's not allowed here. Or they might "pair" without backbone distortion if something bridged them, to form a "ternary pair". A clue as to what might do the bridging is in the wording of the question: that omnipresent but oft-forgotten… H-O-H.

This is based on S R Holbrook et al, Crystal structure of an RNA double helix incorporating a track of non-Watson-Crick base pairs. Nature 353:579, 10/10/91. The work was actually done with RNA, but they argue likely relevance to DNA. The purpose of the question is to emphasize the role of the solvent in macromolecular structures. (Bridging waters have also been implicated in protein structures. See Fig 9.18 for a bridging water between a protein and DNA)

11. The reaction changes an amino group to a carbonyl group. It is not oxidative, but you have to look carefully at the hydrogens to see that. First, replace the $-NH_2$ with -OH; clearly a non-oxidative reaction. But U (for example) has a carbonyl group (not -OH) replacing the amino group. Look carefully, and you will that the H shifts (a tautomeric shift), and is not lost (as it would be if this were an oxidation). That is, the H count is what you would expect for replacement of $-NH_2$ by -OH.

Schwartz et al (1999) discuss one of these "RNA editing" events.

12. Tag each device with a specific SS DNA strand, and tag the desired location with the complement. They will find each other.

Tagging the desired locations makes use of the idea of DNA arrays, or gene chips. A page on gene chips is at my web site.

See Braun et al (1998) for a simple example. The Science note mentioned there discusses preliminary work for more complex constructions.

13. The simple answer is so you know where the DNA is. A company proposes attaching a specific transponder to each DNA molecule, so that each one sends out its own identifying signal. Now, hybridize the test sample -- labeled with fluorescence tags -- to this "radio-labeled" DNA. Run the resulting mixture through a cell sorter type of apparatus and look for "hits", where the fluorescence signal matches a radio signal. See Service (2000).

14. a. There are only four choices for each base. It takes two binary decisions to choose among four choices; $2^2 = 4$. Thus each base = 2 bits.

b. The genome size corresponds to 8×10^6 bits of information (from part a). There are eight bits to a byte, so that is 10^6 bytes, or 1 Megabyte.

Disks come in different sizes, but common floppy disks hold ~1 Megabyte. The point is, for perspective, that the E. coli genome size is similar to the capacity of one disk.

To extend this a little, a single-spaced page of typing contains about 3000 bytes of information. Thus the coli genome is equivalent to about 300 pages of single-spaced text -- considerably less than any human-written book that purports to explain what that genome does. (The human genome is approximately equivalent to 1 CD.)

Figure: H-DNA, from Wells et al, 1988. I have added labels on the strands near top.

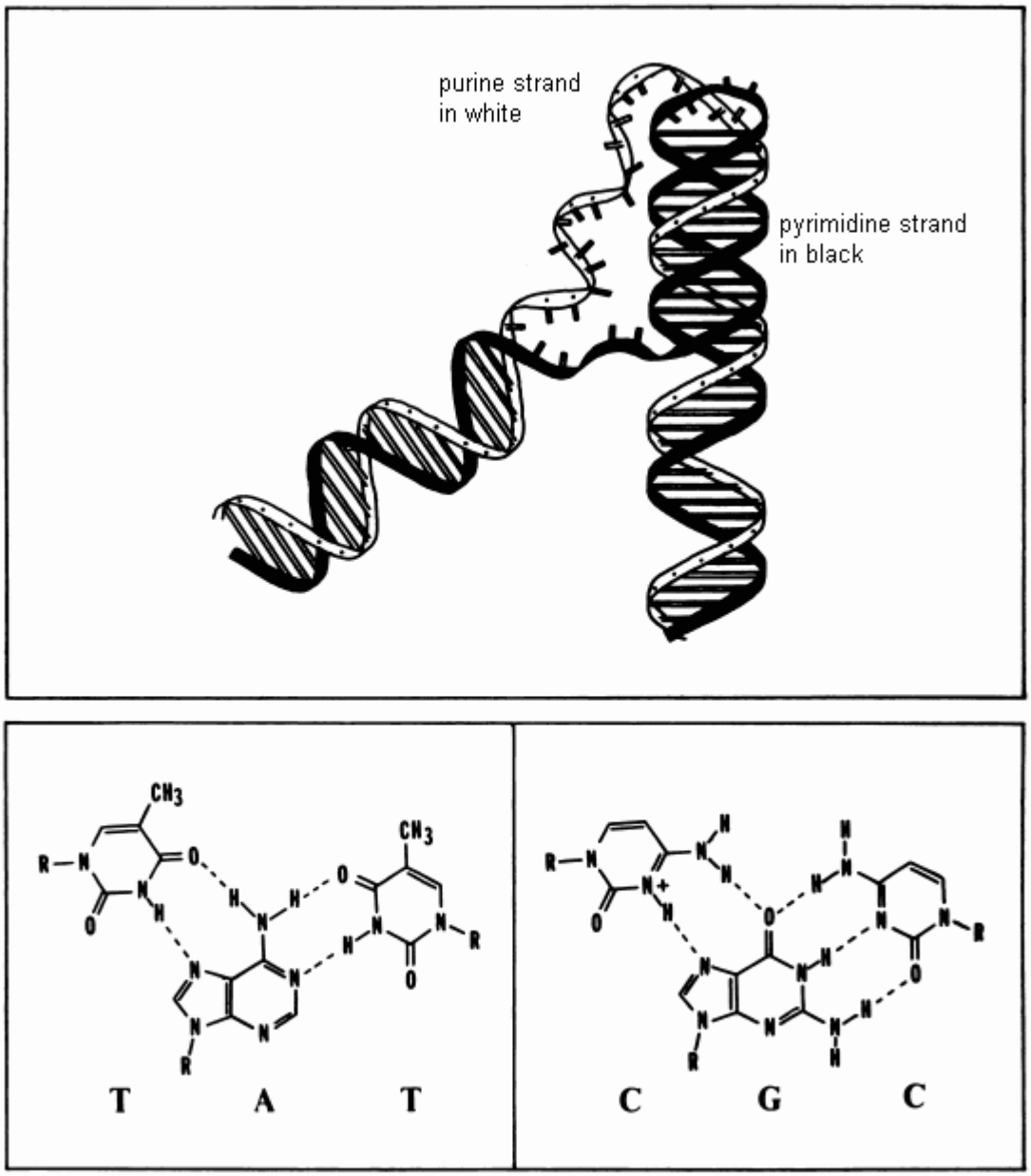

Figure 1. Model for intramolecular triplex (top panel). The Watson-Crick (W.C.) paired duplexes of the vector are shown at the bottom right and bottom left. The strand containing the pyr-rich sequence (upper half of figure) is shaded (black); the strand containing the pur-rich sequences is unshaded (with dots). The 3'-half of the pur strand (termed second strand) of the insert remains W.C. paired with the 5'-half of the pyr strand. The 3'-half of the pyr strand (termed third strand), after dissociating from its W.C. complement, occupies the major groove of the duplex region and is specifically complexed with it through Hoogsteen base pairs (bottom panels). The mirror repeat and the oligopurine·oligopyrimidine nature of these sequences enable this specific association. The structure of the nonpaired 5'-half of the pur-rich strand is uncertain. Reprinted with permission from ref 50.