NUCLEIC ACIDS

1. Introduction

Nucleic acids are polymeric materials formed from nucleotides that are essential to all organisms. Deoxyribonucleic acid (DNA), a double-helical biopolymer, encodes the genetic information contained in each cell (1). Ribonucleic acid (RNA) constitutes a more diverse class of biopolymers that are able to adopt both helical and other more complex tertiary structures. The structural diversity of RNAs enables these molecules to carry out a variety of intracellular functions, including transmitting the genetic message to the site of protein synthesis (see PROTEINS), functioning as enzymes, gene silencing and gene regulation (2). Both DNA and RNA interact with a host of other molecules, eg, proteins, drugs, as well as other RNAs and DNAs. The specificity of these interactions, key to all biological processes, is related to recognition of specific structural elements including sites for hydrogen-bond formation and regions where sequence-dependent structural variation complements the binding molecule.

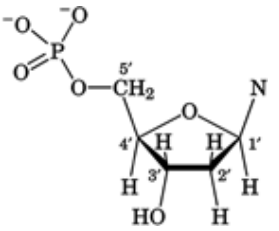
Development of techniques to synthesize oligonucleotides (3), ie, short, well-defined sequences of DNA or RNA, has provided the opportunity to study nucleic acid structure in detail. In addition, oligonucleotides are essential for a number of scientific methods (PCR, DNA sequencing, DNA Microarrays or DNA chips, and affinity chromatography) that are used in genetic engineering (qv), protein engineering (qv), and forensics, and medicine (see CHROMATOGRAPHY; FORENSIC CHEMISTRY). The unique ability of nucleic acids to bind to self-complementary sequences has been exploited in the design of oligonucleotide probes, new drug development and as a framework for constructing molecular devices (nanotechnology). The completion of the Human Genome Project has ushered in a new era where techniques and products based on oligonucleotides will become increasing important.

2. DNA Structure

The structure of DNA is characterized by its primary sequence, secondary helical structure, and higher order structure or topology. The primary sequence of DNA refers to the atomic connections required to construct the polynucleotide chain. The helical conformation of these polynucleotide chains constitutes the secondary structure of DNA. Sequence-dependent structural diversity and flexibility are important DNA characteristics and play a crucial role in biological processes. The organization of helical DNA in topologically distinct three-dimensional (3D) conformations represents the higher order structure. Higher order structural features, in particular the supercoiling of DNA, have a profound influence on the dynamic processes and biology of nucleic acids within living cells (1).

The DNA double helix was first identified by Watson and Crick in 1953 (4). Not only was the Watson–Crick model consistent with the known physical and chemical properties of DNA, but it also suggested how genetic information could be organized and replicated, thus providing a foundation for modern molecular biology.

The primary structure of DNA is based on repeating nucleotide units, where each nucleotide is made up of the sugar, ie, 2′-deoxyribose, a phosphate, and a heterocyclic base, N. The most common DNA bases are the purines, adenine (A) and guanine (G), and the pyrimidines, thymine (T) and cytosine (C) (Fig. 1). The base, N, is bound at the 1′-position of the ribose unit through a heterocyclic nitrogen.



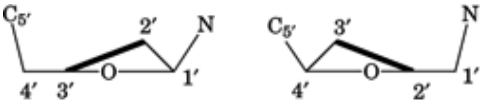
The nucleotides are linked together via the phosphate groups, which connect the 5′-hydroxyl group of one nucleotide and the 3′-hydroxyl group of the next to form a polynucleotide chain (Fig. 1**b**). DNA is not a rigid or static molecule; rather, it can adopt a variety of helical motifs.

2.1. B-DNA. In B-form DNA, two self-complementary polynucleotide strands associate with one another to form a right-handed double helix (Fig. 2**b**). The two polynucleotide chains are antiparallel. One chain extends in the $5' \rightarrow 3'$ direction, the other in the $3' \rightarrow 5'$ direction. The phosphate groups of each polynucleotide chain lie on the surface of the double helix, constituting a phosphodiester backbone, which defines two concave regions on the surface of the double helix known as the minor and major grooves. The polynucleotide chains are associated with one another via hydrogen bonding between complementary A–T or G–C base pairs (see Fig. 1) lying perpendicular to the helical axis at the center of the helix. This double-helical structure is known as B-DNA and is the conformation adopted by most of the DNA found in cells.

The original structural model for B-DNA was based on X-ray diffraction of heterogeneous DNA fibers (4). Because heterogeneous DNA was used, the Watson-Crick model characterizes the average structure of B-form DNA, which has helical parameters of 10.5 base pairs per turn of the helix, a rise of 0.34 nm/base pair, and a helical width of 2 nm. The X-ray structure of a B-form oligonucleotide also has these average parameters, but slight differences between bases are evident (Fig. 2**b**). In the double helix, the hydrophobic bases are oriented toward the interior of the molecule where the bases are stabilized by hydrogen-bonding and base-stacking interactions with one another. The hydrophilic sugar-phosphate backbone characterizes the surface of the double helix and gives B-DNA the water-solubility properties required in a biological environment.

2.2. A-DNA. Other helical DNA structures have been identified from fiber diffraction data. Most important among them is A-form DNA (Fig. 2**a**), which is characteristic of DNA fibers in a low (75%) humidity environment and is the common helical conformation adopted by double-stranded RNA and RNA–DNA hybrids. The A duplex is more common in DNA with alternating G:C pairs of sequences that have higher G:C content. The A–DNA helix (2.6 nm across) is wider than B-form and requires 11 base pairs per turn of the

helix with a helical rise of 0.25 nm/base pair. Moreover, the base pairs are tilted and off center with respect to the helix axis. Unlike B-DNA, where the minor and major grooves resemble each other in depth, the A-DNA minor groove is wide and shallow, and its major groove is narrow and deep. The characteristic difference between these two helices, however, is the conformation of the sugar ring. Sugar puckering is $C_{2'}$-endo for B-DNA and $C_{3'}$-endo for A-DNA (5).



$C_{2'}$ - endo $C_{3'}$ - endo

2.3. Crystallographic Analyses of Oligonucleotides. The Watson-Crick model of DNA was devised based on interpretation of fiber diffraction data, and therefore, is representative of the average structure of DNA. Single-crystal X-ray analysis of well-defined DNA sequences (see X-RAY TECHNOLOGY) reveals that the local structure of DNA varies depending on nucleotide sequence and crystallization conditions (6). Such studies became feasible in the late 1970s when methods were developed for synthesizing short oligonucleotides having the purity required to obtain crystals suitable for structural analysis. The first single-crystal structure of B-form DNA was of the self-complementary dodecamer, d(CGCGAATTCGCG), in 1980 (7) (Fig. 2**b**). Since that time, hundreds of other B-form oligonucleotides have been characterized by X-ray crystallography and the structural information is available on the Nucleic Acid Database (NDB) established in 1991(8). Although generally consistent with a B-DNA conformation, the crystal structures of these oligonucleotides show considerable variation in local helical parameters. For example, the average helical twist angle between consecutive base pairs is virtually identical to the Watson-Crick value, but actual twist angles range from 20 to 55° for individual base steps. The rise per base pair averages 0.34 nm, but actual values range from 0.25 to 0.44 nm and sugar pucker is broadly distributed around the average C2′-endo configuration. The crystal structures also reveal that the base pairs orientation may deviate from the average planarity. These deviations are described using the parameters twist, roll, tilt, slide and propeller twist (Fig. 3). Based on comparisons of several crystal structures these parameters appear to vary in a sequence-dependent manner, but they do so in a probabilistic way. In other words, certain sequences have structural tendencies, but may adopt different structures in different chemical environments. There appears to be no definite set of rules for these variations, but a few general observations can be made. First, A:T base pairs, having only two internucleotide hydrogen bonds compared to 3 for G:C base pairs, demonstrate far more variability in propeller twisting. As a result, the overall dimensions of the minor groove fluctuate more in regions were there is higher A:T content. Bending of the helical axis almost always is the result of roll and, again, sequence has an influence. Pyridine:purine steps, particularly T:A, is most likely to undergo rolling in the positive sense resulting in contraction of the major groove and expansion of the minor groove of the DNA helix. Purine: purine steps, especially A:A, are least likely to roll and therefore are the stiffest.

Some oligonucleotides adopt an A-form helical structure (Fig. 2**a**) (9). The average structural parameters have been found consistent with the fiber diffraction model, but, as for B-form DNA, considerable variation is apparent among individual base pairs. The A and B structures represent the two extremes in a continuum of right-handed helical structures. The DNA may adopt either of these helical forms as well as intermediate forms given the right set of conditions. Further evidence for this interpretation comes from structural studies of DNA in solution. Oligonucleotides that adopt A conformations when constrained in a crystalline environment often are in the B conformation in solution. An interesting example of the interplay between these two helical forms is demonstrated by the crystal structure of d(CATGGGCCCATG). This duplex has structural features that are intermediate between A and B DNA (10). These observations have led to the proposal that the A conformation in crystallized oligonucleotides is an artifact of crystal packing forces rather than an inherent structural characteristic of the DNA sequence.

2.4. DNA Bending and Flexibility. The results of structural investigations via X-ray analysis and solution methods such as NMR (11) demonstrate important sequence-dependent variation in oligonucleotide conformation. The structures reveal variation at the base pair level and localized bending of the helical axis. One example of an intrinsically bent oligonucleotide is the decamer, d(CATGGCCATG) having a helix bent by 23° over the central four base pairs (12). Other examples of bent DNA include oligonucleotides having phased A:T tracts inserted at regular intervals within the sequence. These A:T tracts impart a regular curvature to the molecule either by being intrinsically curved themselves or by spacing curved junctions between A:T tracts and ordinary B-DNA at regular intervals (13). Structural analysis of d(GGCAAAAAACGG) bound to its complimentary sequence reveals that only 5° of the overall 19° bend occurs at the A tract implying the pyrimidine:purine junctions are responsible for most of the bending (14). Recent molecular dynamics simulations of DNA are consistent with this junction model implying that the junctions are regions of flexibility (15). Bending of the helical axis may be important in orienting long stretches of DNA to allow interactions between nonconsecutive sequences, but the more important role for helical bending and flexibility is in binding of proteins that interact with DNA.

2.5. DNA–Protein Interactions. Of the ~30,000 genes within the human genome, 13% are thought to code for proteins that bind to DNA (16). These include proteins that bind to all DNA irrespective of sequence as well as proteins that recognize specific sequences. Proteins that bind to DNA in a largely nonspecific manner include topoisomerases, histones, polymerases, and some nucleases. These proteins generally bind via electrostatic interactions between positively charged basic amino acid side chains and the negatively charged backbone. Many of these proteins also interact extensively with the minor groove where there are few sequence dependent features. Regulatory proteins involved in modulating transcription of DNA and some nucleases recognize specific sites within a DNA chain millions of base pairs long. According to the Watson-Crick model, the only feature distinguishing one DNA region from another is the base-pair sequence, which suggests that DNA binding proteins recognize specific DNA sites by direct readout of the sequence. In the direct readout model, hydrogen-

bonding acceptor groups in proteins take advantage of characteristic hydrogen-bonding donors located in the major groove of the helix. Both purine and pyrimidine bases have unique heterocyclic nitrogen atoms and carbonyl groups in the major groove that can serve as points of recognition. A survey of numerous DNA–protein complexes has shown that site specific proteins have more specific contacts in the major groove of the helix. The most frequent DNA–protein contact is a hydrogen bonding between a G:C base pair and a basic amino acids such as arginine (17).

Although the direct read out mode of recognition is important, the ability of a protein to identify a specific sequence based on its shape and flexibility is also a factor. This mode of recognition is called the indirect readout model. The regulatory protein trp repressor recognizes its operator sequence even though it has no direct contacts with the DNA bases (5). Several studies involving cocrystals of protein bound to DNA reveal that proteins often induce bending of the DNA helix. A dramatic example of protein-induced bending is provided by the regulatory protein CAP and its 20 base pair operator sequence, where two 40°-kinks are introduced at symmetrically disposed (CA)·(TG) base pairs in the bound protein–DNA complex (18). The bending in this complex is reminiscent of the junction model described above for bending of DNA in the unbound state. Histones are another class of proteins that induce DNA bending required to condense DNA in the cell nucleus.

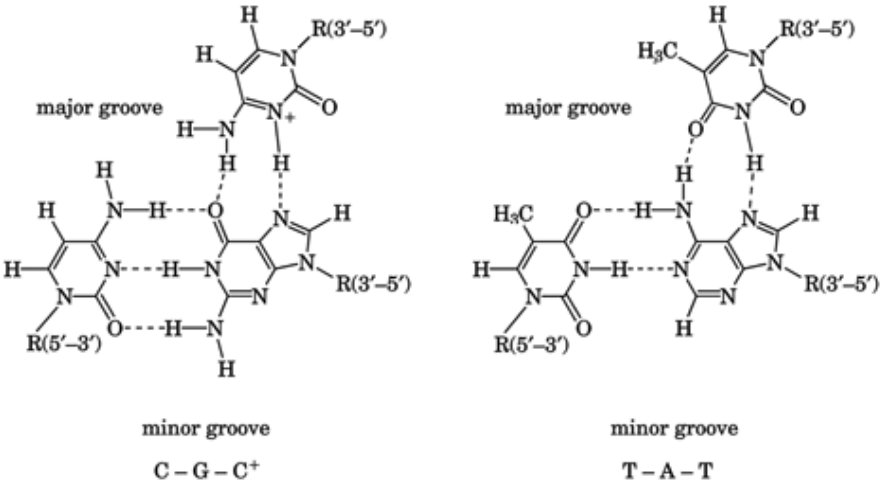
2.6. Z-DNA. In addition to A- and B-form DNA, several other helical conformations have been identified. Among these, one the most well studied is Z-DNA, a left-handed helix first characterized by X-ray crystallographic analysis of the oligonucleotides d(CGCGCG) (Fig. 2**c**) (19). Alternating purine–pyrimidine sequences, in particular alternating CG sequences, have been shown to adopt the Z-conformation at high ionic strength (20).

This helix is named for its zigzagging sugar-phosphate backbone, caused by the syn–anti conformation adopted by the glycosyl linkages to G and C, respectively. The Z-helix requires 12 base pairs per turn and has a diameter of 1.8 nm; the major groove of the helix is essentially flat, and the minor groove is narrow and deep.

The biological role of Z-DNA is not well characterized and has been the subject of much investigation. Early work showed that short segments of poly(dG-dC) incorporated within plasmids, or circular DNA, adopt the Z-conformation under negative superhelical stress. This result suggested that left-handed DNA may be important in genetic control serving as a signal for transcription, particularly in prokaryotes (21). Recently experiments have shown that an RNA editing enzyme forms a complex with Z-DNA and a similar binding domain was discovered in a protein associated with tumors (22). These results suggest that Z-DNA may be significant in living systems, perhaps having a role in regulatory processes.

2.7. Triple Helices. Triple helices occur when a double helix, containing primarily purine bases on one strand, binds to a third polynucleotide containing polypyrimidines. The polypyrimidine strand binds on one side of the major groove of the double helix, forming Hoogsteen hydrogen bonds with the purine bases along one side of the duplex. The single strand may be oriented in either a parallel or an antiparallel fashion. The T-AT helix is more likely to occur under

physiological conditions due to the requirement that C must be protonated in C^+-GC for the correct Hoogsteen hydrogen bonding to take place. The triple helix has structural features of an A-form double helix with the third strand binding in the major groove opening it wider with concomitant reduction of the minor groove width (23).



Much of the interest in triple helices is related to attempts to synthesize oligonucleotides capable of forming triple helices *in vivo*. Since most sequence specific DNA binding proteins interact with the major groove, triple helices could block protein binding and may provide a means of altering the transcription of genes or introducing a mutation (24). In both instances the triple helix would act as an anti-gene agent. A major impediment to the success of this strategy is the inherent weak binding of the third strand to the duplex as well as the fact that native oligonucleotides are easily degraded and poorly transported into the cell. An additional drawback is that target sequences within the genome were thought to be rare but recent bioinformatics studies have demonstrated that these sequences may be more abundant than was once thought (25). Design of robust oligonucleotides that bind strongly to their target sequences is an active area of research (26).

2.8. Guanine Quadruplexes. Eukaryotic chromosomes are linear DNA molecules that end in a series of short guanine tracts regularly distributed along the strand oriented in the $5' \rightarrow 3'$ direction. The very end of the chromosome is a single-stranded region 100–200 base pairs long composed of predominantly guanine residues. These guanine-rich sequences, called telomeres, are highly conserved and bind to specific proteins called telomerases *in vivo*. Telomeres serve to preserve the integrity of the chromosome by protecting it from nuclease digestion. In somatic cells, these single-stranded regions are shortened with each cell division, eventually leading to senescence when the telomere becomes too short. In contrast, tumor cells do not undergo teleomere shortening on division and can therefore grow indefinitely. Interest in telomeres derives from their importance in the life cycle of the cell and their importance in maintaining tumor cells.

Structural studies of single-stranded guanine rich sequences *in vitro* demonstrate that they form intramolecular four-stranded complexes in which the guanine bases associate with one another via Hoogsteen-type hydrogen bonds. The quadraplex structures occur in a variety of parallel and anti-parallel looped arrangements and appear to be quite stable. While these structures form spontaneously *in vitro* it is not known if these structures form in the cell. Thus, the biological relevance of quadraplex DNA is unknown at this time (27).

2.9. DNA Topology. DNA must be highly compacted and organized in the cellular environment. For example, the Escherichia coli chromosome, 1.5-mm long, must be contained within a cell having a diameter of < 1 μm. All naturally occurring DNA molecules, including plasmids, chromosomes, as well as mitochondrial and chloroplast DNA, achieve compaction in ways that depend in part on supercoiling. Whereas compaction is the most obvious objective of supercoiling, it plays a more important role in mediating DNA-associated biological processes. Natural DNA molecules are negatively supercoiled. Supercoiled DNA is a high energy conformation, and this energy contributes to the binding of proteins involved in replication, transcription, and recombination. In addition, supercoiling may induce the formation of some of the higher energy helical structures such as Z-DNA (28).

Supercoiling is a topological property of closed-circular DNA molecules. Circular DNA molecules can exist in various conformations differing in the number of times one strand of the helix crosses the other. These different isomeric conformations are called topoisomers and may be characterized in terms of the linking number, Lk. A linear DNA molecule having N base pairs and h base pairs per turn of the helix, if joined end to end, has the following:

$$Lk^{\circ} = N/h \tag{1}$$

where Lk° refers to the linking number of the relaxed state of the circular molecule. If torsion is applied either to underwind or to overwind the DNA helix before the ends of the linear DNA are joined, the linking number of the circular DNA changes. The linking difference, ΔL, is as follows:

$$ath\Delta L = Lk - Lk^{\circ} = Lk - N/h \tag{2}$$

When the helix is underwound prior to rejoining the ends, ΔL is negative and the circular DNA molecule is negatively supercoiled. Likewise, when the DNA molecule is overwound, ΔL is positive and the circular DNA is positively supercoiled.

The linking difference can be considered in terms of changes in the twist, Tw, and writh, Wr, of the molecule. Twist describes how the DNA strands are coiled around each other; writh describes the coiling of the helical axis. The change in linking number is related to changes in twist and writh:

$$ath\Delta L = \Delta Tw + \Delta Wr \tag{3}$$

In a DNA molecule, constraints imposed by duplex structure resist underwinding and overwinding of the helix about its axis, ie, they resist changes in the

twist of the helix. In other words, energy is required to change the number of base pairs per helical turn. The circular DNA molecule may accommodate a change in linking number by coiling the helical axis, which is equivalent to a change in writh.

Relaxed closed-circular DNA may be represented as shown in Figure 4**a**. The black line on the surface of the circle indicates that $\Delta Lk = 0$. Laboratory tubing is often used as a model for DNA supercoiling. If the circular tubing is opened at one point and the end is twisted around four times in a left-handed sequence before the tubing is reconnected, the linking number changes to $\Delta Lk = \Delta Tw = -4$ (Fig. 4**b**). The parameter ΔLk can also be accommodated by a change in writh, $\Delta Wr = -4$, where there is no change in winding about the axis. When this occurs, the axis coils around and crosses itself four times resulting in supercoiling (Fig. 4**c**). In an actual model, however, a change in linking number can generate some change in both twist and writh (Fig. 4**d**).

Electron micrographs of supercoiled DNA *in vitro* demonstrate that they molecules have a structure intermediate between Figure 4**c** and **d**, where changes in the linking number are accompanied by approximately a 75% change in Wr and a 25% change in Tw. This conformation is called plectonemic supercoiling.

Many naturally occurring DNA molecules, including plasmids, bacterial chromosomes, and mitochondrial as well as chloroplast DNA, are examples of closed-circular DNA in which much of the linking difference is manifested in plectonemic supercoiling. In the cellular environment, however, DNA is organized by binding to the surface of proteins. When proteins interact with DNA, the supercoiled structure is better described by a toroidal model (Fig. 4**e**), where the DNA helix is wound around the surface of a torus created by DNA-binding proteins. This model is particularly important in the organization of eukaryotic chromosomes where the DNA is bound to nucleosomes and organized in a higher order structure.

2.10. Chromatin. Chromosomal DNA, found in the nucleus of eukaryotic cells, is bound to histone proteins to form the nucleoprotein complex called chromatin (29). Chromatin assumes an organized structure based on sequentially repeating units called nucleosomes, which serve the dual purpose of condensing the chromosomal DNA, and organizing it so that it is accessible to the cellular machinery required for transcription, replication, and mitosis. A nucleosome consists of a core of eight histone proteins. A 146 base pair stretch of DNA binds to the histones by wrapping 1.7 times around the histone core to form a left-handed toroidal supercoil. The nucleosomes along a DNA strand resemble a string of beads, a structure known as the 10-nm fiber. The DNA between nucleosomes is bound to a linker histone that serves the dual role of protecting the DNA and organizing the nucleosomes in higher order fibers with a solenoidal structure (Fig. 5). The negative supercoiling of the DNA is manifested by writhing the helical axis around the nucleosomes. Chromatin structure is an example of toroidal winding; whereas eukaryotic chromosomes are linear, the chromatin structures, attached to a nuclear matrix, define separate closed-circular topological domains.

The fibers are further organized into folded domains, called chromonema fibers, that may be thought of as the nucleic acid equivalent of tertiary struc-

tures. The organization of these structures is a consequence of chemical modification by histone modifying complexes. These complexes mediate higher order chromatin structure by adding or removing methyl, acetyl, and other chemical groups to the histones. Each nucleosome has two tail domains that can undergo chemical modification and modification of the tail regions is thought to be most important in chromatin folding.

Chromatin is a dynamic structure that must provide access and sequester DNA in response to the needs of the cell. The ATP dependent chromatin remodeling complexes carry out a number of these functions including changing the position of the DNA along the nucleosome, histone exchange, and histone removal. Chromatin modification and dynamics are central to replication and transcription in the eukaryotic genome (30).

2.11. Topology of DNA in Biological Systems. DNA is necessarily supercoiled when it condenses in chromatin and the supercoiling changes whenever the DNA is unwound during transcription, replication, or recombination. The fidelity of these processes requires that DNA topology be maintained at a certain optimal level. Management of DNA topology is accomplished by a class of enzymes called topoisomerases that relax supercoils or introduce them in a process that is coupled to ATP hydrolysis (gyrases). Supercoiling can be altered by either breaking one strand on the duplex and passing the other strand through it (type I topoisomerases) or breaking both strands and reconnecting them (type II topoisomerases). These enzymes have been of significant interest because a number of drugs have been discovered that act by preventing topoisomerases from reconnecting the DNA after it has created the single or double-stranded break. Many of these drugs are used clinically as antimicrobials and anticancer chemotherapeutics (31).

3. RNA Structure and Function

3.1. Introduction. RNA serves a variety of functions within a cell and, given the new roles for RNA that have been discovered recently, the importance of RNA in cellular processes is just beginning to be fully appreciated. Functional diversity of RNA is related to structural diversity both in terms of their lengths and their folding patterns. The larger biologically important types of RNAs include messenger RNAs (mRNA), which serve as an intermediaries for carrying genetic messages from the DNA to the ribosomes where protein synthesis takes place, ribosomal RNAs (rRNA), the major structural and functional molecules in the ribosome, and transfer RNAs (tRNAs), molecules that translate the mRNA instructions in protein synthesis (2,5). Other RNA molecules, called ribozymes, function as enzymes to catalyzing chemical transformations the RNA phosphodiester backbone (32). Small double-stranded RNAs that are complementary to mRNA sequences have been found to silence the genes in a process call RNA interference (RNAi) (33) and small single stranded RNAs (aptamers) have been found that bind to metabolites to form complexes (riboswitches) that mediate gene transcription and translation (34). In addition to the primary functions described above, many of these RNA molecules are able to play a host of secondary roles in biological systems (35). The multifunctional character of RNA, par-

ticularly the involvement of RNA in enzymatic and regulatory processes, supports the hypothesis that life on earth evolved from RNA. In the RNA world, RNA had both genetic and catalytic functions commonly associated with DNA and proteins, respectively, in modern organisms (2).

RNA Structure. The primary structure of RNA is similar to that of DNA, but with a few notable exceptions. First, in RNA, instead of thymine, the pyrimidine base uracil (U) occurs, forming a complementary base pair with adenine in regions of double-stranded RNA (Fig. 6). Also, a wide variety of ribonucleotides having modified or minor bases are found in naturally occurring RNA, one of the most common of which is pseudouridine. In human tRNAs, as many as 25% of the bases are nonstandard. Over 80 modified bases have been characterized in naturally occurring tRNA; although the role of base modification is not clear, it may be important for biological recognition (2,5).

Natural RNAs are usually single-stranded molecules that fold, allowing different regions of the ribonucleotide to form distinct secondary structural elements (35). When self-complementary regions of the RNA strand are aligned, duplex regions, which often have Watson-Crick base pairs, are formed. In contrast to DNA, double-stranded regions in RNA are much more likely to have unusual base pairing between noncomplementary bases and to incorporate non-Watson-Crick base pairing. Owing to the steric requirements of the 2′-hydroxyl group on the ribose sugar, these duplex regions are constrained to an A-form helix with a 3′-endo sugar conformation. The constraints imposed by the 2′-hydroxyl group make the A-RNA double helix stiffer and less able to bend than the corresponding DNA helix. Although double-stranded RNA has the general features of an A-form helix, actual duplex characteristics, such as rise per base pair, groove dimensions, and base pair displacement from the helical axis, may vary.

Extensive duplex RNA regions exist, but this folding back of the strand and noncomplementary sequences necessarily give rise to single-stranded RNA regions. These latter regions form structures such as hairpins, internal loops, bulges, and junctions (Fig. 7). A hairpin forms when a single strand doubles back, creating a duplex stem and a single-stranded loop. Hairpin loop regions range in length from 2 to 10 ribonucleotides. Internal loops form when base mismatches occur on both strands on an otherwise self-complementary RNA duplex. Whereas these regions are called loops, the noncomplementary bases actually adopt unusual base pairing schemes known as wobble, reverse-wobble, Hoogsteen, reverse-Hoogsteen, and reverse-Watson-Crick hydrogen bonds (Fig. 8). These looped regions retain an overall A-form helical geometry, but are more open and flexible than A-form DNA, facilitating binding interactions with proteins and other RNAs. Bulge loops occur when noncomplementary bases are present on one strand of an otherwise complementary duplex. Branched loops or junctions connecting multiple hairpin structures are the functional domains of ribozymes. All of these structural elements are combined and folded to yield RNAs having diverse tertiary structures.

Efforts to understand the rules that govern RNA tertiary structure and folding to formulate a predictive algorithm are currently being pursued. The approach depends on characterizing the 3D structures of hundreds of RNA structures and using a structural database to identify structural motifs and relation-

ships. This field parallels similar efforts in the area of protein structure. Success in this area would be invaluable in designing functional RNA molecules for therapeutic purposes (36).

RNA Function. The functional diversity of RNA is directly related to its structural diversity. In contrast to DNA, RNA molecules are synthesized as single-stranded oligonucleotides that fold to give complex tertiary structures. These structures, which incorporate hairpins, loops, bulges, and junctions between single-stranded and double-stranded regions, exhibit long-range interactions within the folded tertiary structure. Long-range intramolecular interactions serve to stabilize RNA structures. In cells, RNA is usually associated with proteins or other nucleic acids, and these interaction influence the structure of the RNA. Dissociation of one protein and reassociation of another can radically alter RNA conformation. These interactions are essential for proper folding and maturation of RNA molecules and dynamic processes such as splicing and ribosome biogenesis (37).

3.2. Transfer RNA. Over 2000 sequences of tRNAs have been determined and several tRNA crystal structures have been solved . The overall structure of tRNA, as well as some specific nucleotides within the tRNA sequences, are highly conserved. The secondary structure (Fig. 9**a**) conforms to a clover leaf model. Crystal structures show that tRNA is an L-shaped molecule characterized by stem and loop regions (Fig. 9**b**). These structural regions participate in specific interactions with mRNAs, rRNAs, and proteins.

Transfer RNAs participate in protein synthesis by translating the genetic information encoded in the base sequence of mRNA. Each tRNA is specific for an amino acid, which is covalently attached to the 3′-terminus to give an aminoacyl–tRNA (see AMINO ACIDS). The anticodon, a base triplet contained in the anticodon loop of the tRNA, recognizes the corresponding mRNA triplet codon and forms a complex by complementary base pairing. When this complex forms, the amino acid is transferred from the 3′-end of the tRNA to the growing polypeptide chain. Amino-tRNA synthetases, the enzymes responsible for attaching the amino acids to the tRNA, recognize specific sites along the acceptor stem and the anticodon loop, ensuring fidelity in protein synthesis. Transfer RNA interacts with as many as 30–40 proteins and participates in other processes in addition to protein synthesis. The conserved structure of this molecule is essential for its biological function.

3.3. Ribozymes. In 1982, the first example of catalytic activity in an RNA molecule was reported. The self-splicing preribosomal RNA from ciliate *Tetrahymena* removes an intron from a primary RNA transcript and joins the two exons. Since this discovery hundreds of ribozymes have been discovered in organisms ranging from bacteria to humans all carrying out reactions involving the RNA phosphodiester backbone facilitated by nucleophilic attack by the 2′-hydroxyl on the backbone phosphate (Fig. 10**a**). Catalytic activity, as well as the many other essential biological processes involving RNA, supports the hypothesis that the first living organisms were RNA based with RNA serving as both a catalyst and transmitter of genetic information, with metal ions, amino acids and other small molecules acting as cofactors (32).

The ribozyme from ciliate *Tetrahyemena* is an example of a group I self-splicing intron. Over 1500 examples of this type of ribozyme have been characterized

making it the most common. Group II introns carry out a similar reaction using a different mechanism. There are about one-half as many known group II introns as there are group I. One significant similarity between these two ribozymes is that they both require divalent metal ions for catalytic activity. The crystal structure of a group I intron demonstrates this explicitly (38). Since nucleic acids have few functional groups with pK_a's values in the range required to facilitate general acid–base catalysis, metal ions fulfill this function for these two types of ribozymes. In addition, metal ions are probably important in mediating the folding of the ribozyme so that it can assume a catalytically active conformation.

The 40 nucleotide hammerhead ribozyme has been studied extensively and it was the first to be characterized by X-ray crystallography (Fig. 10**b**,**c**). The ribozyme catalyzes site-specific self-cleavage of the phosphodiester backbone of RNA at the phosphate, generating a 2′3′-cyclic phosphate and a 5′-hydroxyl terminus. The reaction is very efficient enhancing the rate by a factor of 10^9 over the uncatalyzed reaction. Base pairing between the ribozyme and the target is the means by which the cleavage target is recognized. The crystal structure of the ribozyme shows that it has a wishbone-shaped tertiary structure and three A-form helical stems (39). The stems meet at a central core encompassing a conserved CUGA sequence that makes up the cleavage site. Stems 1 and 3 are bound to the substrate sequence and hold the sequence at a bent position at the cleavage site. The ability of the RNA to orient the substrate cleavage site is a critical element in catalysis.

In an RNA world, ribozymes would be required to catalyze a far greater range of reactions than has thus far been observed in modern organisms. The lack of these functions may result from a gradual replacement of RNA enzymes with more robust protein enzymes through an evolutionary process, rather than from an inherently small enzymatic repertoire for ribozymes. Evidence for this hypothesis comes from *in vitro* selection methods that have been employed to create ribozymes having a wide range of chemical activity. The method involves randomly testing thousands of RNA molecules for ability to carry out a specific chemical reaction. Efficiency of activity is maximized via successive rounds of selection and the best catalyst is transcribed into DNA and amplified by the polymerase chain reaction (PCR). Using this approach, ribozymes able to catalyze the synthesis of nucleotides, amides, Michael adducts, and acyl–coenzyme A have been obtained (32).

3.4. RNA Interference. RNA interference (RNAi) refers to the ability of double-stranded RNA to silence genes containing the same sequence. This mode of gene silencing may have evolved as a means of defense against parasitic transposons, transgenes and viruses, or as a cellular mechanism for eliminating defective mRNAs and regulating genes. Of great interest recently, however, is the use of RNAi as a method to silence genes. The Human Genome Project (HPG) and sequencing initiatives involving other organisms have identified millions of genes whose protein products are uncharacterized. RNAi provides a means of removing the protein product from a cell without influencing other pathways. By observing the consequences, it is often possible to infer the function of the protein encoded by the silenced gene. In addition, this gene silencing strategy may provide the basis for an entirely new class of pharmaceuticals (40).

Several enzyme-mediated pathways are important in the RNAi process (Fig. 11). First, double-stranded RNA must be cleaved to yield 22 base pair double stranded fragments called small interfering RNAs (siRNA) This cleavage is carried out by a ribonuclease called dicer. The siRNAs then bind to a protein–RNA nuclease called RISC that unwinds the siRNA and uses one of the strands as a recognition element to generate the active form of the enzyme, RISC*. The RISC* enzyme carries out gene silencing in a number of ways. The most direct is RISC* binding to mRNAs preventing translation of the message. A second mechanism is cleavage of mRNA at the sequence complementary to the 22 base pair recognition element. Other possible functions of RISC* is to associate with chromatin remodeling factors and direct their action to the location specified by the recognition element. In this scenario, transcription of the gene would be altered. Also RISC* may be able to amplify the siRNA. These multiple pathways demonstrate that RNAi is complex, but its widespread use in genomics makes understanding this process extremely important (33).

Other small RNAs of cellular origin may also be important players in the RNAi process. The first of these to be discovered were micro RNAs (miRNA) encoded by the *lin-4* and *let-7* genes of *C. elegans*. These genes encode 70 base pair miRNAs that fold into hairpin structures that are subsequently processed by DICER to generate 22 base pair siRNAs that interfere with translation as described above. Since this initial discovery numerous genes for miRNAs have been identified in metazoan genomes and it is believed that mammalians genomes encode these regulatory miRNAs as well. These examples are the first of many noncoding RNAs that have been found to be biologically significant (41).

3.5. Riboswitches. Small folded domains in noncoding regions of mRNA have been found to bind to metabolites. Binding results in regulation of the synthetic pathway for that particular metabolite molecule. These regulatory structural elements are called riboswitches. The mechanism employed by riboswitches is an allosteric one, involving a structural change on metabolite binding that directly influences the transcription, translation, splicing, or stability of mRNA. The first riboswitch mechanism to be characterized involves feedback inhibition of one of the proteins required for synthesis of coenzyme B-12 in *E. coli*. A 315-nucleotide region of the 5′ untranslated region of the mRNA was found to bind to coenzyme B-12. This riboswitch operates at the trancriptional level terminating the elongating mRNA prematurely when coenzyme B-12 is bound. Termination of mRNA synthesis occurs because a structural change in the RNA forms an intrinsic terminator stem that causes transcription to be aborted. When no coenzyme B-12 is bound, transcription proceeds normally to produce the mRNA required for coenzyme B-12 synthesis. Riboswitches regulate several other coenzyme pathways including thiamine pyrophosphate, flavin mononucleotide, as well as guanine, adenine, S-adenosylmethionine, glucosamine 6-phosphate, and lysine. Additionally, several other regulatory systems have been identified where RNA combined with a protein component carry out the regulatory process (42).

Prior to the discovery of riboswitches, it was thought that only proteins had the structural features required for small molecule recognition and allosteric regulation. Riboswitches demonstrate that RNA is fully capable of distinguishing subtle differences in the structure of small molecules. Methods based on *in*

vitro evolution have been developed to synthesize RNAs called aptamers having the ability to recognize and bind to a wide range of small molecule ligands and proteins (43). A database cataloging these aptamers and their properties is currently available (44). The binding properties of aptamers could potentially be exploited in developing pharmaceuticals and analytical methods.

4. Oligonucleotide Synthesis and Applications

4.1. Oligonucleotide Market and Applications. Synthetic oligonucleotides are widely used in scientific research and are required for many experiments and analytical procedures. The most important application of synthetic deoxyoligonucleotides is their use as primers in the PCR (45), an important analytical technique having commercial applications in diagnostic medicine, genetic engineering (qv), and forensics (see FORENSIC CHEMISTRY). Demands for PCR primers have increased rapidly to support large-scale DNA mapping and sequencing efforts such as the human genome project (16). Also, synthetic oligonucleotides are used as linkers in gene cloning (47), to introduce site-directed mutations in genes (48), for structural biochemical, and biophysical studies of DNA and RNA and for examining the association of proteins (50) and small molecules (51) with nucleic acids. Modified oligonucleotides conjugated to cross-linking agents and molecular probes are widely used in molecular biology. In addition, gene-silencing techniques requiring RNA oligonucleotides are becoming increasingly important. Both RNAi and antisense mechanisms are currently being exploited for these studies.

Though the quantity of oligonucleotide needed in individual applications is relatively small, the high demand has given rise to an industry devoted to their preparation. The global market for all oligonucleotides is expected to grow from $ 340 million in 2003 to $ 776 million in 2010 (52). Some of this growth can be attributed to the trend among research laboratories is to obtain these reagents from specialized suppliers rather than by in house preparation. Also, custom suppliers are increasing able to provide a wide range of modified and bioconjugated oligonucleotides. Custom synthesis of oligonucleotides is automated, cost effective, and widely available from a number of sources. The leaders in DNA oligonucleotide synthesis include Integrated DNA Technologies (IDT), invitrogen, MWG Biotech, and Sigma-Genosys. These suppliers also offer modified oligonucleotides but the industry also includes numerous specialty companies. For example, leaders in the RNA synthesis include Dharmacon, Qiagen-Operon, and Ambion specialize in RNA oligonucleotide synthesis. Other specialty oligonucleotides houses include Proligo, Synthegen, and Eurogentec.

Oligonucleotides synthesized for use as antisense pharmaceuticals has created the need to produce very large quantities (kilograms) of single oligonucleotide products for clinical trials and commercial products. The approach used in these preparations is essentially the same as in small scale oligonucleotide synthesis, eg, phosphoramidite chemistry employing a solid support (see below). Substantial gains in cost effectiveness have been realized in the past decade with the cost per gram of oligonucleotide decreasing from > $100,000/g prior in 1990 to $500/g in 1995. Ultimately, the price may reach $50/g or less. The

major challenge facing the pharmaceutical industry in developing cost effective processes is to lower the cost of the solid support. The solid support is the most expensive material used in the process. Efforts to either create a reusable support or do without it altogether are currently being actively pursued by the pharmaceutical industry (53).

The development of DNA microarrays or biochips represents one of the most important technical advances of the 1990s and represents a major growth area for synthetic DNA and RNA. Microarrays consist of hundreds to millions of oligonucleotides covalently attached to a surface. While some microarrays incorporate biologically derived oligonucleotides, the trend is to synthesize the sequences of interest directly on the solid support (*in situ* synthesis) (54). To prepare microarrays, phosphoramidites used in solid-phase DNA synthesis are deposited on a solid support using for example inkjet technology. Approximately 10,000 unique oligonucleotides may be synthesized on a single array by delivering the base monomers in the correct sequence. The oligonucleotides are designed to recognize complementary DNA or RNA in a biological sample via hybridization. After incubation with a biological sample, analysis of the hybridization reactions can provide a profile of the genes expressed in the cell (55). Microarrays are powerful tools for monitoring gene expression. Other applications include the study of single nucleotide polymorphisms (SNPs), mutation detection, and sequencing. There is tremendous potential for this technology in the areas of medical diagnostics and electronics. Instrumentation for microarray preparation and analysis is commercially available from Agilent Technologies, DNA Technologies Inc., Axon Instruments, and Perkin-Elmer Life Sciences.

4.2. DNA Synthesis. The first procedures for oligonucleotide synthesis where carried out in solution and were based on phosphotriester (56) and *H*-phosphonate (57) chemistry. These approaches may still have applications in some large-scale syntheses and in syntheses of various oligonucleotide analogues, but most modern procedures rely on solid-phase phosphoramidite chemistry (58). Automated oligonucleotide synthesizers are commercially available, as are the required reagents and phosphoramidites. Together these permit the rapid production of custom oligonucleotides and oligonucleotide analogues with sequences specified by the consumer. The chemistry involved in the three most frequently used oligonucleotide syntheses is described herein.

The Phosphoramidite Method. The phosphoramidite method is the most efficient and widely used protocol for syntheses of both unmodified and modified oligonucleotides (3,53,58). Using this chemistry, oligonucleotides having chain lengths up to ~200 nucleotides may be obtained. Phosphoramidite technology is useful for generating both nanomolar quantities of DNA or RNA for use in genetics experiments and molar quantities of oligonucleotides for the pharmaceutical industry.

The synthetic scheme involves chain-extending addition of protected mononucleotides to a nucleoside bound covalently at the 3′-hydroxyl to an inert solid supports such as cross-linked polystyrene or controlled pore glass (CPG) (Fig. 12). The initial base-protected 5′-*O*-dimethoxytrityl (DMT) deoxynucleoside is attached to the solid support via a molecular linker yielding a 3′-terminal nucleoside attached to the solid support (**1**). Chain elongation requires the removal of the 5′-DMT protecting group by trichloroacetic acid (TCA) hydrolysis.

The chain is extended in the $3' \rightarrow 5'$ direction by tetrazole-catalyzed addition of the 5′-hydroxyl group to a protected 3′-phosphoramidite to generate a phosphite triester. The β-cyanoethyl-*N,N*-diisopropylphosphoramidite (**2**) is most commonly used for this purpose. Unreacted 5′-hydroxyl groups are acylated to prevent them from reacting in the next condensation cycle in a step called capping. Chain elongation at these hydroxyls can generate oligonucleotides having one base deleted from the target sequence, thus complicating the purification of the product oligonucleotide. The phosphite triester is oxidized using I_2 to form the stable phosphate triester (**3**). Treatment with TCA frees the 5′-hydroxyl of the growing chain so that chain elongation can proceed by repeating the condensation step and using the next phosphoramidite base in the sequence. The coupling reaction has been optimized to produce yields of >99% in 1 min/synthetic cycle. The exocyclic nitrogens on cytosine and the purine bases must be protected during the synthesis. The search is ongoing for protecting groups that are subject to fewer side reactions and that can be removed more easily in the final deprotection step (53). When the synthesis is complete, the protecting groups are removed and the oligonucleotide is liberated from the solid support by treatment with ammonium hydroxide. The oligonucleotide is then purified by high performance liquid chromatography (hplc) (59) (see CHROMATOGRAPHY; ELECTROCHEMICAL PROCESSING, INTRODUCTION).

Phosphoramidite chemistry is almost always carried out using an automated DNA synthesizer. The first successful instrument was introduced in 1982 by Applied Biosystems, Inc. in conjunction with Leroy Hood at the California Institute of Technology. While the instrumentation has seen significant modification and improvement, Applied Biosystems remains the major manufacturer of research scale oligonucleotide synthesis equipment.

The Phosphotriester Method. The phosphotriester method was the first to be used to prepare oligonucleotides on a solid support and, therefore, represents a technological milestone in nucleic acid synthesis (60). The approach, however, is limited due to technical difficulty and poor yields (90–95%), and interest in this chemistry is primarily historical. As in the other methods, synthesis proceeds in the $3' \rightarrow 5'$ direction with the 3′-hydroxyl bound to a solid support. The oligonucleotide is obtained by sequential condensation of protected nucleoside 3′-phosphodiesters. The 5′-hydroxyl group of the growing chain is activated by an arenesulfonic acid condensation reagent. Commonly used arenesulfonic acids include 2,4,6-triisopropylbenzesulfonic acid (TIPS) and 8-quinolinesulfonic acid (QSNT).

The time required for each coupling cycle in the phosphotriester method compares favorably to that in the phosphoramidite method, but yields for the former method are generally lower. Although the coupling yields are lower, this synthetic method has the advantage of generating the phosphotriester directly, thereby eliminating the oxidation step in each coupling cycle.

The H-Phosphonate Method. In the late 1980s, an early synthesis of oligonucleotides based on *H*-phosphonates (**4**) (Fig. 13), reemerged (61). This method eliminates one of the steps in the coupling reaction, resulting in fast coupling times. Unfortunately, the more reactive reagents often generate undesirable side products so this method is not widely used. The procedure involves the activation of 5′-DMT-deoxynucleoside 3′- *H*-phosphonates with a sterically

hindered carbonyl chloride, followed by condensation with the 5′-hydroxyl group of a protected nucleoside. The phosphite diester (**5**) is stable, and, following the required capping and deprotection steps, chain elongation can proceed without oxidation until the desired oligomer is obtained. Upon completion of the synthesis, the phosphite diester linkages are converted to phosphodiesters (**6**) by a single oxidation using *tert*-butyl hydroperoxide or iodine. In addition to simplifying the oxidation, this method eliminates the need for a phosphate protecting group. Because coupling reactions proceed in high yield, the method has been used successfully to synthesize oligonucleotides in excess of 100 nucleotides in length. While not commonly used in solid phase synthesis, this chemistry is being investigated to produce large quantities of oligonucleotides using a solution-phase approach (62).

4.3. Ribonucleotide Synthesis. The same general methods used in deoxynucleotide synthesis can be applied to the synthesis of ribonucleotides provided that the 2′-hydroxy group is suitably protected. Coupling between the 3′ and 5′ hydroxyl groups requires that the reactivity of the 2′ group be eliminated. Ultimately, the protecting group must be removed, but removal is complicated by the inherent reactivity of RNA. While the structural difference between RNA and DNA is seemingly minor, the 2′-hydroxyl group makes RNA much more susceptible to base hydrolysis and phosphodiester backbone isomerization. When solid-phase phosphoramidite chemistry is used, the protecting group must be resistant to the acidic reagents needed to deprotect the 5′-hydroxyl in the chain elongation step and, in the end, easily removable using conditions that will not damage the ribonucleotide.

Many protecting groups have been investigated for use in oligoribonucleotide synthesis and development of new protecting groups is ongoing. Currently, the most commonly used groups are 2′-*O*-silyl moieties because they are compatible with phosphoramidate chemistry. These groups were developed by Kevin Ogilvie and are now commercially available (63). Large groups, such as the *tert*-butyldimethylsilyl and triisopropylsilyloxymethyl, are resistant to both acid and base hydrolysis, but can be removed under neutral conditions by treating the reaction mixture with fluoride ion. The bulky group does make the coupling reactions slower, but that problem has been addressed by longer reaction times and more reactive phosphoramidite activators (64). Using this approach large tRNAs and ribozymes have been prepared.

Stephen Scaringe at the University of Colorado developed another promising synthetic route to RNA oligonucleotides (65). Instead of using 5′-DMT phosphoramidites requiring acid deprotection, Fluoride liable 5′-silyl groups were used. This modification allowed 2′-hydroxyl protection with the acid sensitive 2′-O-bis(2-acetoxyethoxy)methyl orthoester (ACE) group. This group is very stable through all of the steps required to prepare the nucleosides and oligonucleotides. After the synthesis is complete, the protecting group is hydrolyzed to the 2′-*O*-bis(hydroxyethoxy)methyl orthoester in the same step where the base protecting groups are removed. The resulting orthoester is still resistant to RNase and hydrolysis but is even more acid labile than ACE. The oligoribonucleotide can be purified and then deprotected with very mild base to give the RNA product in high yield. Reagents supporting solid-phase synthesis using this method are commercially available.

5. Modified Oligonucleotides

5.1. Applications. Much of the interest in modified or unnatural oligonucleotides has been inspired by their potential use as antisense agents (66). Antisense agents are typically short (15–30 base pairs in length) oligonucleotides having sequences that are complementary to coding or regulatory regions within mRNA, although some antisense oligonucleotides have also been designed to target DNA (antigene strategies). The antisense sequence recognizes and binds to a complementary sequence via the formation of a double-stranded duplex that have normal Watson-Crick base pairing or via Hoogsteen base pairing in the case of triple helix formation, thereby interfering with gene expression. The potential to design oligonucleotides having the ability to recognize and inhibit specific genes makes the antisense approach promising in the development of new therapeutic agents. In addition, antisense oligonucleotides can be used in research to elucidate gene function by providing a mechanism for regulating a gene artificially. This later application is especially important for characterizing genes whose function is unknown. While straightforward in theory, there are many practical challenges to development of effective antisense pharmaceuticals. These challenges include stability of the oligonucleotide in the cellular environment, transport of the oligonucleotide through the cell membrane, target affinity and toxicity. Despite he difficulties, a few antisense drugs have been approved for clinical use and several others are in the phase I and phase II clinical trials. This progress suggests that antisense technology will be a very important growth area for the pharmaceutical industry in the early twenty-first century.

The antisense mechanism is a natural mechanism used in both prokaryotic and eukaryotic cells to regulate gene expression (67). In one example of the naturally occurring processes, the genome encodes antisense sequences complementary to cellular mRNA. When these sequences are transcribed, antisense RNA hybridizes with the complementary mRNA sequences to form a double-stranded RNA–RNA duplex. Formation of the hybrid duplex inhibits expression of the gene corresponding to that particular mRNA by an enzyme-mediated process, in which cellular enzymes recognize and chemically modify the double-stranded RNA sequences. When modified, the mRNA cannot be translated. Thus, naturally occurring antisense mechanisms repress gene expression by interfering with the translation of mRNA. Antisense oligonucleotides are usually designed to inhibit gene expression by interfering with the translation of mRNA. This type of inhibition involves binding the oligonucleotide to the translation–initiation sequences of the mRNA, which prevents ribosome association and protein synthesis. Another potential mechanism involves hybrid formation at some other sequence within the mRNA, thus impeding translocation of the ribosome along the mRNA strand by steric blocking (66). These two mechanisms are based on blocking a sequence of RNA by double-stranded duplex formation using a specific antisense oligonucleotide.

Many oligonucleotides that form hybrids with mRNA also activate RNase H, an enzyme that catalyzes single-strand cleavage of RNA (68). In the RNase H mechanism, the duplex formed by the antisense oligonucleotide and the target

RNA is a substrate for RNase H. The enzyme cleaves the RNA at the complexes site rendering the RNA vulnerable to further degradation and inactivation by cellular exonucleases. The oligonucleotide, which is probably not a substrate for RNase H, can target multiple copies of complementary RNA. Antisense oligonucleotide action mediated by the RNase H mechanism has been shown to be a potent inhibitor of gene expression.

Catalytic nucleic acids (69) and small interfering RNAs are two other oligonucleotide strategies with great potential as pharmaceuticals affecting gene expression. Ribozymes recognize their target through base pairing, cleave a specific mRNA sequence, release the product RNA fragments and go on to carry out multiple rounds of this process (70). These RNAs, called trans-cleaving ribozymes, are currently in clinical trials targeting infectious diseases and cancer. RNA interference, a process in which short double-stranded RNA inhibits gene expression (see discussion above under RNA), is another promising strategy for the development of drugs.

In another approach called the antigene strategy, modified oligonucleotides are designed to bind to double-stranded DNA by forming a triple helix. Triple-helix formation is a sequence-specific interaction in which the single-stranded oligonucleotide binds in the major groove of homopurine–homopyrimidine DNA targets via the formation of Hoogsteen or reverse-Hoogsteen hydrogen bonds. If the complementary DNA target is within a gene or a regulatory region associated with a specific gene, triple-helix formation can interfere with the transcription of the gene. One focus of this research involves extending the range of recognition sequences. Specificity of the Hoogsteen base pairs and steric considerations strongly favor triple-helix formation along homopurine–homopyrimidine DNA tracts. Base modifications and novel oligonucleotides are being developed in an effort to make triplex formation viable for other DNA target sequences. However, despite good specificity, triple helices appear to be less stable and slower to form than double helices. Oligonucleotides have thus been modified by using substituents designed to enhance binding, eg, intercalators or cationic groups (71).

Although natural antisense oligonucleotides are short RNA sequences, synthetic antisense oligonucleotides, as well as other oligonucleotides designed for *in vivo* applications, have a wide variety of structural modifications. These modifications usually include chemical modification of the phosphate backbone and often include conjugation to other molecules. Oligonucleotides with sugar and base modifications have also been investigated. These chemical modifications are undertaken to address several practical considerations related to efficacy. In the design of an effective bioactive oligonucleotide, several factors must be considered. First, the oligonucleotide must be specific, binding with high affinity to a single sequence within the target RNA or DNA. Statistically, a 17-base pair sequence is expected to occur only once in the human genome; therefore, an oligonucleotide of this size range should have the required specificity. A second consideration is stability within the cellular environment. All cells produce nucleases, ie, enzymes that catalyze the rapid degradation of unmodified oligonucleotides via cleavage of the phosphodiester bonds. 3′-Exonucleases are particularly problematic. These enzymes degrade DNA and RNA nonspecifically in a stepwise manner that begins at the 3′-phosphodiester linkage. Thus all unmodi-

fied oligonucleotides are degraded too rapidly to be used effectively as therapeutic agents. A significant research effort has been directed toward discovering chemical modifications that can increase the nuclease resistance of the oligonucleotide backbone. An effective therapeutic agent must also have the ability to reach its target sequence *in vivo*. Bioavailability requires that the antisense oligonucleotide be able to pass through the cell membrane, and that it have a low affinity for nontarget cellular compartments and, in animal systems, non target organs. Cell membranes are lipophilic and designed to be barriers against large anionic molecules. In order to enhance membrane transport, bioactive oligonucleotides are frequently modified by covalent attachment of carrier molecules such as lipophilic groups.

5.2. Oligonucleotides with Modified Phosphodiester Backbones. Oligonucleotides having modified phosphate backbones have been studied extensively (72). Because altering the backbone makes derivatives more resistant to degradation by cellular nucleases, these materials have the potential to be more resilient antisense drugs. Some of the most frequently encountered backbone modifications are illustrated in Figure 14.

Changing the phosphodiester backbone usually reduces the anionic character of the oligonucleotide. This can have two consequences. First, the water solubility of the oligonucleotide is reduced, making the modified polymer less capable of forming a duplex structure that is appropriately hydrated. In this case, the modified oligonucleotide binds less well to its complementary sequence than the unmodified oligonucleotide. In contrast, an analogue having a less anionic, neutral, or cationic character is expected to experience less electrostatic repulsion by the phosphate groups on the target sequence. This effect enhances the nonspecific component of binding.

5.3. Phosphorothioates. Phosphorothioates (Fig. 4**a**) were among the first backbone modified oligonucleotides to be investigated and several have been developed as antisense pharmaceuticals. They are similar to naturally occurring phosphodiesters in that each linkage retains a formal negative charge (72). The magnitude of the charge is reduced, however, because of the nature of the sulfur atom. As a result, phosphorothioates are less soluble in water and have somewhat lower thermal stability than the phosphodiester homologues. The advantage of these sulfur-containing oligonucleotides as antisense agents is that they are $\sim$10 times more resistant to the nucleases than unmodified oligonucleotides. Moreover, these modified nucleotides are transported across the cell membrane, form duplex structures with their target RNA sequences, and activate RNase H. The major disadvantages of phosphorothioates is their increased affinity for cellular proteins, reduced affinity for their RNA target sequence, and toxicity. Despite these drawbacks several phosphorothiolates are currently being evaluated in clinical trials and the first antisense drug to be approved in 1998 was a phosphorothiolate called Vitravene. This compound, manufactured by ISIS pharmaceuticals, has been shown to be effective in the treatment of cytomegalvirus-induced retinitis in AIDS patients (66). Phosphorothioates are thus referred to as first generation antisense agents.

Phosphorothioates can be synthesized using the phosphoramidite method simply by changing the oxidant from I_2 to S_8 in the final step (72). The phosphoramidite method offers the flexibility of introducing a thiophosphate at various

positions within the oligonucleotide but chirality at the phosphorus is an unavoidable problem in all phosphorothioate syntheses. A mixture of both the R_p and the S_p diastereomers are produced. To avoid this problem, protocols involving enzymes have been used to generate chiral phosphorothioates that have found applications in studies of enzyme mechanism. The chiral center may be eliminated by replacing the other, nonbridging oxygen with sulfur generating a phosphorodithiolate backbone. Avoidance of the chirality problem is one reason for the interest in phosphorodithioates (73).

5.4. Phosphorodithioates. Phosphorodithioates (Fig. 14**b**) can be obtained by using a modified phosphoramidite method (74). Phosphoramidites are replaced by 3′-phosphorothioamidites that yield thiophosphite triesters in the condensation reaction. Using S_8 as the oxidizing agent, phosphorodithioates have been obtained in coupling yields ranging from 96 to 99%. No other changes in the standard protocol are necessary. The phosphorodithioates DNA derivatives have been shown to bind specifically to complementary DNA or RNA sequences to form stable adducts. Because they are also highly resistant to degradation by cellular exonucleases, these derivatives can be useful both for applications in research and as therapeutic drugs. Phosphorodithioate DNA has been shown to stimulate RNase H activity in nuclear cell extracts and is a potent inhibitor of HIV type-1 reverse transcriptase (75). Another important feature is the reactivity of the phosphorodithioates with alkylating agents. When one phosphorodithioate linkage is incorporated within an oligonucleotide, it can serve as a point of attachment for reporter groups bearing an alkylating function.

5.5. Methylphosphonates. Methylphosphonates (Fig. 14**c**) are derivatives were one of the nonbackbone phosphate oxygens is replaced by a methyl group. They are efficiently made by the phosphoramidite method using methylphosphonoamidites as coupling reagents with yields of ~96–97% (76). Again, the phosphoramidite method affords the flexibility of positioning a methylphosphonate linkage at any point in the oligonucleotide chain. The methylphosphonates differ from the phosphodiesters and phosphorothiolates in that the methyl derivatives are uncharged and are thus less water soluble. Moreover, compared to the naturally occurring phosphodiesters, the methylphosphonates form slightly less stable duplexes with complementary DNA and RNA sequences. This effect has been ascribed to the inevitable chirality problem; that is, if one isomer binds less well, the overall binding is decreased. Methylphosphonates can enter cell membranes by a passive mechanism and are completely resistant to nucleases.

5.6. N3′–P5′ Phosphoramidates. N3′–P5′ Phosphoramidates (Fig. 14**d**) have a nitrogen atom in place of the 3′ backbone oxygen. They are prepared using solid-phase synthesis based on *H*-phosphonate chemistry (77) and are of considerable interest as antisense pharmaceuticals. They have favorable properties including nuclease resistance and high affinity to their RNA targets. They act by preventing transcription of the target RNA and have been shown to be effective in the treatment of leukemia in mice (78).

5.7. Peptide Nucleic Acids. Peptide nucleic acids (Fig. 14**e**) are oligonucleotide analogues where the entire deoxyribose phosphate backbone is replaced by an achiral polyamide backbone (79). They were initially designed

based on computer-modeling studies that suggested an *N*-(2-aminoethyl)glycine backbone is structurally homomorphous to the deoxyribose phosphate backbone. Experiments have subsequently shown that PNAs form stable Watson-Crick duplexes with their complementary sequences as well triple helices. In addition, PNAs also interact with DNA via a duplex invasion mechanism where the PNA displaces one or both of the DNA strands. The oligomers are easily constructed by using a modified Merrifield solid-phase synthesis for peptides and are currently commercially available from a variety of sources. PNAs have several properties that make them good candidates for pharmaceuticals. They are nontoxic, have high affinity and specify for complementary nucleic acids and, being uncharged, have low affinity for proteins. Charge neutrality does pose a serious problem with respect to solubility and cellular uptake and these issues are being addressed by the development of several carriers that can be coupled to the PNA (80). These include negatively charged polymers, lipids and other peptides that are efficiently taken up by cells. While PNAs do not act by activating RNaseH, they may be effective as antisense agents by interfering with transcription but of greater potential are therapeutic approaches involving targeting of the DNA where the PNA would alter gene expression because of its strand invasion ability.

Oligonucleotides with Modified Sugars. Alteration of the sugar moiety within an oligonucleotide is typically undertaken for two reasons: to render the oligonucleotide more resistant to nuclease degradation i*n vivo* and to increase the binding affinity of the oligonucleotide to it's target sequence. These properties render the modified oligonucleotide more suitable for applications in biochemical studies and as antisense drugs. All of these derivatives are readily obtained by solid-phase synthesis using the phosphoramidite method, and many of the phosphoramidite precursors are commercially available (81).

The first and most common modification target is the 2′ hydroxyl of the ribose ring. Oligo(2′-*O*-alkylribonucleotides) having short, unbranched alkyl groups are resistant to nucleases, chemically stable, and capable of hybridizing with complementary RNA to form duplexes of greater stability than unmodified DNA or RNA. Alkylated oligonucleotides such as 2′-*O*-methyl and 2′-*O*-methoxy ethyl (Fig. 15**a**,**b**) have been found to offer all of the advantages as antisense agents outlined above as well as much lower toxicity when compared to their phosphorothiolate counterparts. However, these derivatives, as well as most oligonucleotides with altered sugars, do not activate RNaseH cleavage, limiting their effectiveness in antisense applications (66).

The 2′-hydroxyl modified oligonucelotides do have other uses in research and medicine. Replacement of a hydroxyl with alkoxy group has been used to determine catalytically essential 2′-hydroxyl groups in ribozyme mechanisms (72). Another important application of these oligonucleotides is in the development of aptamers able to recognize and bind strongly to specific cellular components including proteins, hormones, cofactors, and nucleoside triphosphates. Aptamers that are nuclease resistant may have uses in both therapeutics and diagnostics (82).

Building on the theme of 2′-hydroxyl modification, 2′-*O*,4′-C-ethylene-bridged nucleic acids or LNAs (Fig. 15**c**) were independently synthesized in the Imanishi and Wengel laboratories and are currently being evaluated as anti-

sense agents. These oligonucleotides are nuclease resistant and bind very strongly to their RNA and DNA target sequences (83,84). The structural consequence of the ethylene bridge is to lock the sugar ring in the C3′-endo conformation and, as a consequence, duplexes of LNA oligonucleotides favor an A-type helical conformation. In addition to these properties, low toxicity and good cellular uptake makes these compounds promising materials for antisense drugs. The locked backbone makes LNA a good framework for synthesizing new materials (see below) (85).

Several oligonucleotides incorporating non-sugar cyclic components have been synthesized and are currently being investigated as potential antisense drugs (84). These include cyclohexene nucleic acids (86) (Fig. 15**d**), tricyclo-DNA (87) (Fig. 15**e**) and morpholino analogues (88) (Fig. 15**f**). These derivatives have been used in several studies involving inhibition of gene expression. They form stable duplex structures with their DNA or RNA targets but do not activate RNase H.

In order to restore RNaseH cleavage, oligonucelotides called gapmers have been developed that incorporate a short central stretch of unmodified DNA within a modified sequence. The unmodified sequence acts as a substrate for the enzyme while the modified portions allow the molecule to retain desirable antisense characteristics (89).

Oligonucleotides with Modified Bases. Oligonucleotides incorporating modified bases are widely used in studies of biological mechanism and structural characterization. Several questions can be addressed by altering the natural purine and pyrimidine functional groups. For example, the requirements for sequence specific recognition by proteins can be studied using oligonucleotides with bases lacking the hydrogen-bonding heterocyclic nitrogen atoms (90). The importance of specific functional groups in stabilizing DNA or RNA conformation and base involvement in ribozyme mechanisms can also be investigated (91). Two examples of nonnatural bases used in these studies, 7-deazaadenine and 3-deazaadenine, are shown in Figure 16**a** and **b**.

Oligonucleotides having altered bases can also serve as fluorescent probes, cross-linkers and target -specific probes for DNA microarrays. Incorporating fluorescent bases such as the areneo[*g*]lumazine analogs of thymidine provides a means for studying dynamic processes involving structural change (92,93) (Fig. 16**c**). Chemical cross-linkers provide a means of trapping noncovalent complexes so that they may be characterized spectroscopically. Cross-links can be photo-induced when bases such as 5-iodouridine and 8-azidoadenosine are incorporated into the oligonucleotides (Fig. 16**d**,**e**). Covalent linkages based on disulfide bonds have also proved to be useful for this purpose (94,93) (Fig. 16**f**). Microarray technology is becoming increasingly important and the utility of method rests on the ability of the oligonucleotide probe to recognize a specific sequence in a biological sample. In samples where there are many similar sequences, undesirable cross-hybridization become a problem. The use of well-chosen unnatural bases has the potential to increase specificity ultimately leading to the ability to differentiate between two sequences with a single base pair difference (95).

Other molecular substitutions for bases have been carried out where the sugar-phosphate backbone may be thought of as a scaffolding or organizing fra-

mework for entirely new molecules. This category includes molecules where the bases have been replaced by metal binding ligands and simple aromatic hydrocarbons (96) (Fig. 16**g**,**h**). These variations may have applications in biology and nanomaterials science (see below).

Bioconjugated Oligonucleotides. Oligonucleotides modified by covalently attaching other molecular entities have many important applications. There has been a significant amount of research untaken to facilitate cellular uptake of antisense and RNAi oligonucleotides, ribozymes, and genetic elements by covalently attaching a molecular moiety that is easily transported through the cell membrane. Although many molecules, including proteins, intercalators and groove-binding ligands, bind to RNA and DNA, nucleic acids are most able to bind with the high specificity required to recognize a single sequence within the 3×10^9 base pairs of the human genome. The unique specificity of oligonucleotides can be exploited to direct a multitude of other chemical agents to a sequence of interest by attaching these agents to oligonucleotides through molecular linkers. Probes, where fluorescent groups, phosphors, radioactive tracers (qv) are attached to a sequence specific oligonucleotide, are widely used in molecular biology. They can be designed to recognize and RNA or DNA sequences characteristic of specific eukaryotic genes, viruses, or bacteria. Several analytical and diagnostic procedures have been developed based on the hybridization of the probe with its target sequence and the subsequent detection of the hybrid by the group attached to the oligonucleotide. Other modifications are undertaken to attach oligonucleotides to solid surfaces or particles as a way of immobilizing the oligonucleotide. Enzymes, drugs and cross-linking agents have been linked to oligonucleotides where the sequence specific recognition properties of the oligonucleotide serves to direct these reactive or therapeutically useful species to specific DNA or RNA sites. Bioconjugates have also been widely used in research because they enable scientists to learn more about the structure and function of nucleic acids and ligands that bind to them.

Several strategies have been devised to attach various chemical groups to oligonucleotides. Groups can be attached to the 3′- or 5′-terminus of the oligonucleotide, along the backbone through the phosphate or the 2′-hydroxy group of ribose, or to modified purines or pyrimidines. Most of these modifications are undertaken using a molecular linker to attach the oligonucleotide to the molecule of interest. Several linkers with reactive termini are commercially available as phosphoramidites so that bioconjugated oligonucleotides may be prepared easily. Additionally, phosphoramidites and oligonucleotides incorporating many of these groups are commercially available from a number of sources (93).

5.8. Cellular Uptake of Oligonucelotides. Antisense drugs, ribozymes, and gene therapy require transport of oligonucleotides through the cell membranes. Oligonucleotides are anionic species and therefore do not diffuse through the lipid bilayer and cells are not equipped with membrane proteins that act as active transporters of these molecular species. However, other molecules do traverse the cell membrane and attachment of these entities to oligonucleotides can facilitate their entry into the cell. Most of the successful delivery systems have involved conjugation of oligonucleotides to either liposomes or charged lipids (97). These lipids encapsulate and neutralize the oligonucleotide so that the complex may enter the cell via endocytosis. Oligonucleotides have

also been modified by attaching them to cationic, cell-penetrating, peptides. These peptides are probably also taken up by endocytosis although the mechanism remains controversial (98). Conjugation with polylysine, porphyrin derivatives, polymeric dendrimers, and nanoparticles has been shown to increase intracellular concentrations of conjugated oligonucleotides as well (99). Oligonucleotides have also been attached to antibodies that are recognized by transport receptors on the cell surface. This strategy was successful in allowing passage of a PNA through the blood brain barrier (100) and in targeting human ovarian carcinoma cells (101). In another example, attaching a glycopeptide ligand that was recognized by a cell surface receptor allowed transport by receptor-mediated endocytosis (102). Gene and oligonucleotide delivery using a bioconjugation strategy is a very active area of research.

5.9. Probes. Analytical methods require a strategy for detecting oligonucleotides. Historically, labeling with radioactive isotopes has served this purpose, but modification with other detectable groups is rapidly replacing this older method. Attachment of a fluorescent group is the most commonly used detection method and oligonucleotides functionalized with dyes that emit across the far–ultraviolet (UV) and visible spectrum are widely available (93). Oligonucleotides labeled with fluorescent groups are used as probes in analytical procedures such as Southern and Northern blotting, and in microarray and imaging techniques. When these probes are bound to target nucleic acid sequences, the fluorescent labels permit the visualization of the hybrid. For example, electron microscopy can be used to determine the location of nucleic acid sequences bound to the probe among various compartments in a cell (103) (see MICROSCOPY). In microarrays, labeled oligonucleotides derived from a biological sample allow identification of the positions on the plate where hybridization occurred (104). Fluorescence detection is also used in DNA-sequencing and in DNA fingerprinting involving multiplex PCR amplification. Fluorescence detection has facilitated automation of both of these procedures resulting in high throughput analyses. Recently, fluorescent labels have been employed in the development of nanoparticle probes. These probes consist of a metal (eg, gold) or carbon based nanoparticle with labeled oligonucleotides attached. In the unhybridized form the derivatized nanoparticle effectively quenches the fluorescence of the labeled oligonucleotide. The fluorescence is restored by hybridization of complementary DNA or RNA. The probes have many potential uses as emitters able to detect specific DNA sequences (105). Other detection systems are also widely used. For example, biotinylated conjugates form highly colored chromophores in the presence of the enzyme avidin (or streptavidin), and can be used in place of the fluorescent groups in the applications described above.

5.10. FRET. Fluorescent labels are frequently used in structural studies of nucleic acids using fluorescence resonance energy transfer (FRET). By studying the efficiency of fluorescence resonance energy transfer between the donor and acceptor groups, distance information can be extracted that defines the positions of the donor and acceptor groups in a macromolecule or complex. For example, fluorescent donor (5-carboxyfluorescein) and acceptor (5-carboxytetramethylrhodamine) groups were attached to the 3′- and 5′-ends of several model oligonucleotides representative of the hammerhead ribozyme, a catalytic oligoribonucleotide. These experiments provide information regarding the geometry of

the ribozyme, a Y-shaped substance made up of three double-stranded helical regions (106). FRET was also used to enhance detection of oligonucleotides in a DNA sequencing protocol (107) and in the design of molecular beacons that become visible when hybridized with complementary nucleic acids in a biological sample (108). Detection using FRET has great potential in diagnostics and genetic analyses.

5.11. Oligonucleotides Conjugated to Reactive Molecules. Reactive chemical groups attached to oligonucleotides can be used to affect sequence-specific modifications of DNA. For example, both copper–phenanthroline (109) and iron–EDTA (EDTA = ethylenediaminetetraacetic acid) (110) have been linked to triple-helix-forming oligonucleotides. These groups can be activated by using either thiol or hydrogen peroxide, respectively, to generate a hydroxyl radical *in situ*. The hydroxyl radical, a diffusible reactive species, is capable of cleaving the phosphodiester DNA backbone. Thus, when bound to a specific DNA or RNA target sequence, these reagents can act as artificial endonucleases. Ellipticine can also be activated photochemically to yield another reactive group capable of DNA cleavage.

Another synthetic approach to a site-specific DNA or RNA cleavage reagent is to chemically link a naturally occurring nonspecific nuclease to an oligonucleotide. The first example of this type of artificial endonuclease was an oligonucleotide conjugate with staphylococcal nuclease. This adduct has been shown to cleave target DNA 75% more effectively than nonspecific sites (111). RNase H has also been linked to oligonucleotides in an effort to obtain a site-specific RNA nuclease (112).

Modified oligonucleotides can be used to cross-link DNA sequences via a reactive group tethered to an oligonucleotide. When irradiated with UV light, psoralens reacts with thymine bases, and the reaction yields a cross-link if the thymine residues are adjacent to each other on opposite strands. Psoralens linked to oligonucleotides have been shown to induce site-specific cross-links *in vitro* (113). Alkylating agents tethered to oligonucleotides and other reactive linkers can link oligonucleotides and target DNA sequences (94).

Oligonucleotide conjugates have potential therapeutic uses as well. For example, a dye bound to an oligonucleotide acts as a photosensitizer able to generate singlet oxygen photochemically. Singlet oxygen causes damage to the DNA target by a mechanism that results in the cleavage of the phosphodiester backbone. Compounds of this type are used in photodynamic treatment of tumors and are being evaluated in clinical trials (114). Also, the binuclear platinum oligonucleotide conjugate can alkylate DNA through a reaction with guanine heterocyclic nitrogens (115). *cis*-Platinum, a similar platinum complex, is an important drug used in cancer treatment. Psoralens may also be useful in the treatment of cancer when conjugated to antisense drugs (116).

6. Oligonucleotides and Nanotechnology

Automated oligonucleotide synthesis and the availability of a plethora of modified oligonucleotides has inspired a new area in materials science. Researchers are beginning to use nucleic acids as a structural scaffolding to create nanoscale

materials with unique functional properties (85). Complementary base pairing serves as the basis for self-assembly of two- and three-dimensional macromolecular constructs that could potentially be modified to create molecular devices with applications in biology and materials science. This approach could be used to design biosensors, drug delivery devices, artificial enzymes, structures able to carry out multistep synthesis, molecular switches, and wires capable of transmitting an electrical signal. Nanoscale devices are on a molecular scale and represent a lower limit in miniaturization. They may be compared in scale to the macromolecules common to natural biochemical processes.

The first consideration in this field is development of technology to create structures that self-assemble in a predictable and reproducible manner. Some of the earliest examples of these structures involved oligonucleotides whose natural base pairing created Holliday junctions (117). Further assembly can be achieved by creating sticky ends at the termini of the holiday junction so that the individual units bind to one another through Watson-Crick base pairing to create a lattice structure (Fig. 17). Other approaches to structural self-assembly use oligonucleotides that have been modified with molecular pairs known to form strong bonds with one another. These pairs include metal complexes with chelating ligands and protein-receptor pairs (85). More complex 3D structures incorporating knots and rings, 2D arrays and other topological motifs found in nucleic acid structure have followed.

One of the goals in this area is to create materials that facilitate multistep reaction processes and simple systems capable of template-directed chemical reactions have recently been reported. The first example illustrates how oligonucleotides can be used to carry out two reactions in sequence (118). The strategy employed here is to use reactive pairs attached to self-complementary oligonucleotides with a reagent strand poised to covalently attach a chemical group to the complementary reactant strand. After it has reacted, reagent strand dissociates and the reactant strand continues in the process by reacting with another reagent oligonucleotide of slightly longer length. The added length facilitates a new alignment of the reacting groups. In another example, an oligonucleotide is used to direct the catalysis of an ester hydrolysis reaction (119). Here both the ester and catalyst are attached to the termini of oligonucleotides. A longer oligonucleotide that is complementary to both of the modified oligonucleotides functions to bring the two groups together so that they can react. These examples illustrate the feasibility of mediating reaction processes using nucleic acids. It is possible to envision more complex reactors incorporating the larger structural assemblies described above.

Oligonucleotide assemblies with bound metals are promising candidates for molecular electronic devices. One way to accomplish this is to use modified bases with chelating groups directed towards the center of the double helix. When copper ions are bound to these ligands, an oligonucleotide duplex with a core of metal ions results. Experiments have shown that these ions are ferromagnetically coupled to one another through unpaired *d*-electrons, suggesting that these structures could act as wires or magnets in molecular devices (120) (Fig. 18). In another approach, nucleic acid nanostructures are coated with metal in a process called metallization. It is possible to control where the metal is deposited by masking specific regions of the duplex using site-specific

DNA-binding proteins (121). This research could lead to the development of molecular circuits that could be employed in electronic diagnostic devices and computers (117).

Reversible changes in the conformation of nucleic acids in response to some environmental change may be used to design molecular switches and mechanical devices. Several simple devices of this type have been reported where conformational changes are induced by either altering redox potential, temperature, irradiation, or binding to another molecule. For example, a rotary device based on reversible oligonucleotide binding has been designed. The oligonucleotide alters the degree of helical twist in a structure formed from two intertwined duplexes (122). A molecular tweezers was constructed from three strands of DNA and operated through the interaction of a fourth oligonucleotide. When the fourth oligonucleotide is bound the tweezers is held together in the closed state. Dissociation of the oligonucleotide opens the tweezers (123). Future developments in this area could lead to nanorobotics and molecular scale machines.

The examples above suggest some of the potential applications of nucleic acid based materials and devices. This new area of research is certain to grow in the future and eventually lead to commercially important applications.

7. The Human Genome Project and The Future

The completion of the HPG was announced in 2001 ~50 years after the discovery of the double helix with the publication of the complete human DNA sequence (4,16). While this represents a milestone, what has been learned thus far is dwarfed by what is not known about the genome (124). There are ~30,000 genes that encode for proteins and many of these proteins have not been characterized. The protein-coding sequences make up only 1–2% of the genome and the function of the other 98% is largely unknown. Of this 98%, ~2% is under active selection suggesting that it plays a vital role. Another 50% consists of highly repetitive sequences of similarly unknown purpose. While 99.9% of the genome is invariant between individuals, the other 0.1% is of great interest to scientists. Within that 0.1% lie the keys to individual differences in the human population. Learning the function of each genetic element and the consequences of genetic differences will foster a better understanding of human disease. Scientists will spend decades characterizing the genome and the experiments required to do this work are highly dependent on oligonucleotide-based techniques. In addition, oligonucleotide-based drugs and gene therapies will likely follow from what is leaned about the genome. The role of nucleic acids in science, technology and medicine is sure to become more important in the future.

BIBLIOGRAPHY

"Nucleic Acids" in *ECT* 4th ed., Vol. 17, pp. 507–543, by J. Rehmann, Fordham University; "Nucleic Acids" in *ECT* (online), posting date: December 4, 2000, by J. Rehmann, Fordham University.

CITED PUBLICATIONS

1. S. Neidle ed., *Oxford Handbook of Nucleic Acid Structure*, Oxford University Press, Oxford, 1999.
2. R. F. Gesteland, T. R. Cech, and J. F. Atkins, *The RNA World*, 2nd ed., Cold Spring Harbor Laboratory Press, Cold Spring, New York, 1999.
3. P. Herdewijn, ed., *Oligonucleotide Synthesis: Methods and Applications*, Humana Press, Totowa, N.J., 2004.
4. J. Clayton and C. Dennis, eds., *50 Years of DNA*, Palgrave Macmillan, New York, 2003.
5. S. Neidle, *Nucleic Acid Structure and Recognition*, Oxford University Press, Oxford, 2002.
6. Dickerson, in Ref. 1, p. 145.
7. R. M. Wang, H. R. Drew, T. Takano, C. Broka, S. Takana, and R. E. Dickerson, *Nature (London)* **287**, 755 (1980).
8. http://ndbserver.rutgers.edu
9. M. C. Wahl and M. Sundaralingam, *Biopolymers* **44**, 45 (1997).
10. H.-L. Ng, M. L. Kopka, and R. E. Dickerson, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 2035 (2000).
11. J. N. S. Evans, *Biomolecular NMR Spectroscopy*, Oxford University Press, Oxford, 1995.
12. D. S. Goodsell, and co-workers, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 2930 (1993).
13. T. E. Haran, J. D. Kahn, and D. M. Crothers, *J. Mol. Biol.* **244**, 135 (1994).
14. D. MacDonald, K. Herbert, X. Zhang, T. Polgruto, P. Lu, *J. Mol. Biol.* **301**, 1081 (2001).
15. D. L. Beveridge, S. B. Dixit, G. Barreiro, and K. Thayer, *Biopolymers* **73**, 380 (2004).
16. International Human Genome Sequencing Consortium, *Nature (London)* **409**, 890 (2001); J. C. Venter and co-workers, *Science* **291**, 1304 (2001).
17. N. M. Luscombe and J. M. Thorton, *J. Mol. Biol.* **320**, 991 (2002).
18. S. C. Schultz, G. C. Sheilds and T. A. Steitz, *Science* **253**, 1001 (1991).
19. Wang and co-workers, *Nature (London)* **282**, 680 (1979).
20. B. H. Johnston, in D. M. J. Lilley and J. E. Dahlberg, eds., *Methods in Enzymology*, Vol. 211, Academic Press, Inc., San Diego, Calif., 1992, pp. 127–158.
21. A. R. Rahmouni and R. D. Wells, *J. Mol. Biol.* **223**, 131 (1992).
22. T. Swartz, J. Behike, K. Lowenhaupt, U. Heinemann, and A. Rich, *Nat. Struct. Biol.* **8**, 761 (2001).
23. V. N. Soyfer and V. N. Potaman, *Triple Helical Nucleic Acids*, Springer-Verlag, New York, 1996; M. D. Frank-Kamenetskii and S. M. Mirkin, *Annu. Rev. Biochem.* **64**, 65 (1995).
24. C. Giovannangeli and C. Helene, *Nat. Biotech.* **18**, 1245 (2000).
25. J. R. Goni, X. de la Cruz, and M. Orozco, *Nucleic Acids Res.* **32**, 354 (2004).
26. R. A. Cassidy, N. Puri, and P. S. Miller, *Nucleic Acids Res.* **31**, 4099 (2003).
27. S. Neidle and G. N. Parkinson, *Curr. Opin. Struct. Biol.* **13**, 275 (2003).
28. A. D. Bates and A. Maxwell, *DNA Topology*, Oxford University Press, Oxford, 1993; N. R. Cozarelli and J. C. Wang, eds., *DNA Topology and its Biological Effects*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, 1990; W. K. Olson, in Ref. , p. 499.
29. P. J. Horn and C. L. Peterson, *Science* **297**, 1824 (2002).
30. A. E. Ehrenhofer-Murray, *Eur. J. Biochem.* **271**, 1432 (2004).
31. J. C. Champoux, *Annu. Rev. Biochem.* **70**, 369 (2001).
32. J. A. Doudna and T. A. Cech, *Nature (London)* **418**, 222 (2002).

33. G. J. Hannon, *Nature (London)* **418**, 244 (2002).
34. W. W. Winkler, A. Nahvi, A. Roth, J. Collins, and R. Breaker, *Nature (London)* **428**, 281 (2004).
35. I. Tinoco, and C. Bustamante, *J. Mol. Biol.* **293**, 271 (1999); N. B. Leontis and E. Westhof, *Curr. Opin. Struct. Biol.* **13**, 300 (2003); T. R. Sosnick and T. Pan, *Curr. Opin. Struct. Biol.* **13**, 309 (2003).
36. H. H. Gan, S. Pasquali, T. Schlick, *Nucleic Acids Res.* **31**, 2926 (2003); H. M. Al-Hashimi and co-workers, *J. Mol. Biol.* **318**, 637 (2002).
37. S. Rocak and P. Linder, *Nature Rev. Mol. Cell Biol.* **5**, 232 (2004); P. Linder, *Science* **304**, 694 (2004); T. A. Steitz, in Ref. , p. 427.
38. P. L. Adams, M. R. Stahley, A. B. Kosek, J. Wang, and S. Strobel, *Nature (London)* **430**, 45 (2004).
39. H. W. Pley, K. M. Flaherty, and D. B. McKay, *Nature (London)* **372**, 68 (1994).
40. D. M. Dykxhoorn, C. D. Novina, and P. A. Sharp, *Nature Rev. Mol. Cell Biol.* **4**, 457 (2003).
41. G. Storz, *Science* **296**, 120 (2002).
42. M. Mandal and R. R. Breaker, *Nature Rev. Mol. Cell Biol.* **5**, 451 (2004).
43. T. Hermann and D. J. Patel, *Science* **287**, 820 (2000).
44. J. F. Lee, J. R. Hesselberth, L. A. Meyers, and A. D. Ellington, *Nucleic Acids Res.* **32**, D95 (2004).
45. M. J. McPerson, P. Quirke, and G. R. Taylor, *PCR, A Practical Approach*, Oxford University Press, Oxford, U.K., 1991.
46. J. Sambrook, E. F. Fritsch, and T. Maniatis, *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1989, p. 13.7.
47. B.-V. Chen and H. W. James, eds., *PCR Cloning Protocols*, Humana Press, Totowa, N.J., 2002.
48. T. Strachen and A. P. Read, *Human Molecular Genetics*, BIOS Scientific Pub. Ltd, Oxford, 1999, Chapt. 6.
49. J. Nowakowski and I. Tinoco, Jr., in Ref. 1, p. 567.
50. N. M. Luscombe, S. E. Austin, H. M. Berman, and J. M. Thorton, *Genome Biol.* **1**, 1 (2000).
51. M. Demeunynck, C. Bailly, and W. D. Wilson, eds., *DNA and RNA binders: from small molecules to drugs*, Vols. 1 and 2, Wiley-VCH, Weinheim, 2003.
52. http://www.pharmaceuticals.frost.com
53. R. T. Pons, in Y. E. Khudyakov and H. A. Fields, eds., *Artificial DNA: Methods and Applications*, CRC Press, Boca Raton, Fla., 2003, pp. 1–70.
54. X. Gao, E. Gulari, and X. Zhou, *Biopolymers* **73**, 579 (2004). C. Laustad and co-workers, *Genome Biology* **5**, R58 (2004).
55. Special issue in DNA Microarray Data Analysis, *Nat. Genetics*, 2002, December.
56. K. Itakura and co-workers, *Can. J. Chem.* **51**, 3649 (1973).
57. B. C. Froehler, P. G. Ng, and M. D. Matteucci, *Nucleic Acids Res.* **14**, 5399 (1986).
58. M. H. Caruthers and co-workers, *Methods in Enzymology*, Vol. 154, Academic Press, Inc., New York, 1987, pp. 287–313; M. H. Caruthers and co-workers, *Methods in Enzymology*, Vol. 211, Academic Press, Inc., New York, 1992, pp. 3–20.
59. T. Brown and D. J. S. Brown, *Methods in Enzymology*, Vol. 211, Academic Press, Inc., New York, 1992, pp. 20–35.
60. P. Dembek, K. Miyoshi, and K. Itakura, *J. Amer. Chem. Soc.* **103**, 706 (1981).
61. B. C. Froehler and M. D. Matteucci, *Tetrahedron Lett.* **27**, 469 (1986).
62. C. B. Reese and Q. L. Song, *Bioorg. Med. Chem. Lett.* **7**, 2787 (1997).
63. K. K. Ogilvie, A. L. Schifman, and C. L. Penney, *Can. J. Chem.* **57**, 2230 (1979).
64. F. Wincott and co-workers, *Nucleic Acids Res.* **23**, 2677 (1995).

65. S. A. Scaringe, *Methods* **23**, 206 (2001).
66. J. Kurreck, *Eur. J Biochem.* **270**, 1628 (2003); M. I. Phillips, ed., *Methods in Enzymology*, Vols. 313 and 314, Academic Press, Inc., New York, 2000.
67. R. W. Simmons and N. Klecker, *Annu. Rev. Genet.* **22**, 567 (1988).
68. M. A. Bonham and co-workers, *Nucleic Acids Res.* **23**, 1197 (1995).
69. M. J. Cairns, A. King, and L. Q. Sun, *Nucleic Acids, Res.* **31**, 2883 (2003); L. Q. Sun, M. J. Cairns, E. G. Saravolac, A. Baker, and W. L. Gerlach, *Pharm. Rev.* **52**, 325 (2000).
70. B. A. Sullenger and E. Gilboa, *Nature (London)* **418**, 252 (2002).
71. R. V. Guntaka, B. R. Varma, and K. T. Weber, *Int. J. Biochem. Cell Biol.* **35**, 22 (2003).
72. S. Verma and F. Eckstein, *Anu Rev. Biochem.* **67**, 99 (1998).
73. F. Eckstein, *Antisense Nucleic Acids Drug Dev.* **10**, 117 (2000).
74. W. T. Weisler and M. H. Caruthers, *J. Org. Chem* **61**, 4272 (1996).
75. W. S. Marshall and M. H. Caruthers, *Science* **259**, 1564 (1993).
76. P. S. Miller, C. D. Cushman, and J. T. Lewis, in F. Eckstein, ed., *Oligonucleotides and Analogues. A Practical Approach*, Oxford:IRL, 1991, p. 137.
77. S. Gryznov and J.-K. Chen, *J. Amer. Chem. Soc.* **116**, 3143 (1994).
78. M. Faira and co-workers, *Nat. Biotechnol.* **19**, 40 (2001).
79. P. N. Neilsen, ed., *Peptide Nucleic acids, Protocals and Applications*, 2nd ed., Horizon Bioscience, Norwalk, U.K., 2004.
80. D. A. Braasch and D. R. Corey, *Biochemistry* **41**, 4503 (2002).
81. N. Venkatesan, S. J. Kim, and B. H. Kim, *Curr. Med. Chem.* **10**, 1973 (2003).
82. S. E. Osbourne and A. D. Ellington, *Chem. Rev.* **97**, 349 (1997).
83. M. Koizumi, *Biol. Pharm. Bull.* **27**, 453 (2004).
84. C. J. Leumann, *Biorg. Med. Chem.* **10**, 841 (2002).
85. J. Wengel, *Org. Biomol. Chem.* **2**, 277 (2004).
86. J. Wang and co-workers, *J. Amer. Chem. Soc.* **122**, 8595 (2000).
87. D. Renneberg and C. J. Leumann, *J. Am. Chem. Soc.* **124**, 5993 (2002).
88. J. Heasman, *Dev. Biol.* **243**, 209 (2002).
89. M. N. M. Mangos and M. J. Damha, *Curr. Topics in Med. Chem.* **2**, 1147 (2002).
90. S. A. Smith, S. B. Rajur, and L. W. McLaughlin. *Nat. Struct. Biol.* **1**, 18 (1994).
91. S. Bevers, G. B. Xiang, and L. W. McLaughlin, *Biochemistry* **35**, 6483 (1996).
92. D. P. Millar, *Curr. Opin. Struct. Biol.* **6**, 322 (1996).
93. M. J. Martin, A. Shah, I. J. Bruce, *Chem. Soc. Rev.* **29**, 97 (2000).
94. G. L. Verdine and D. P. G. Norman, *Annu. Rev. Biochem.* **72**, 337 (2003).
95. I. Lee, A. A. Dombkowski, B. D. Athey, *Nucleic Acids. Res.* **32**, 681 (2004).
96. E. T. Kool, *Acc. Chem. Res.* **35**, 936 (2002).
97. L. Liang, D. P. Liu, and C. C. Liang, *Eur. J. Biochem.* **269**, 5753 (2002).
98. J. P. Ricard and co-workers, *J. Biol. Chem.* **278**, 585 (2003).
99. L. Q. Sun, M. J. Cairns, E. G. Saravolac, A. Baker, and W. L. Gerlach, *Pharmacol. Rev.* **52**, 325 (2000).
100. N. Shi, R. J. Brado, and W. M. Pardridge, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 14709 (2000).
101. T. Merden and co-workers, *Bioconjugate Chem.* **14**, 989 (2003).
102. R. J. Duff and co-workers, in M. I. Phillips, ed., *Methods in Enzymology*, Vol. 313, Academic Press, Inc., New York, 2000, p. 297.
103. J. L. McInnes and R. H. Symons, in R. H. Symons, ed., *Nucleic Acid Probes*, CRC Press, Inc., Boca Raton, Fla., 1989, pp. 33–80.
104. S. L. Beaucage, *Curr. Med. Chem.* **8**, 1213 (2001).
105. D. J. Maxwell, J. R. Taylor, and S. Nie, *J. Am. Chem. Soc.* **124**, 9606 (2002).
106. T. Tuschl and co-workers, *Science* **266**, 785 (1994).

107. J. Ju and co-workers, *Anal. Biochem.* **231**, 131 (1995).
108. S. Tyagi and F. R. Kramer, *Nat. Biotech.* **14**, 303 (1996).
109. J. C. Francois and co-workers, *J. Biol. Chem.* **264**, 5891 (1989).
110. P. A. Beal and P. B. Dervan, *Science* **251**, 1360 (1991); S. A. Strobel and P. B. Dervan, *Science* **249**, 73 (1991).
111. D. Pei, D. R. Corey, and P. G. Schultz, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 9858 (1990).
112. Y. Uchiyama and co-workers, *Bioconjugate Chem.* **5**, 327 (1994).
113. N. T. Thuong and C. Helene, *Angew. Chem. Int. Ed. Engl.* **32**, 666 (1993).
114. U. Moller, F. Schubert, and D. Cech, *Bioconjugate Chem.* **6**, 174 (1995).
115. E. S. Gruff and L. E. Orgel, *Nucleic Acids Res.* **19**, 6849 (1991).
116. A. Murakami, A. Yamayoshi, R. Iwase, J. Nishida, T. Yamaoka, N. Wake, *Eur. J. Pharm Sci.* **13**, 25 (2001).
117. N. C. Seeman, *Nature (London)* **421**, 427 (2003).
118. Z. J. Gartner, M. W. Kanan, and R. Lui, *J Am. Chem. Soc.* **124**, 10304 (2002).
119. J. Brunner, A. Mokhir, and R. Kraemer, *J Am. Chem. Soc.* **125**, 12410 (2003).
120. K. Tanaka and co-workers, *Science* **299**, 1212 (2003).
121. K. Keren and co-workers, *Science* **297**, 72 (2002); H. Yan, S. H. Park, G. Finkelstein, J. H. Reif, and T. H. LaBean, *Science* **301**, 1882 (2003).
122. H. Yan, X. Zhang, Z. Shen, and N. C. Seeman, *Nature (London)* **415**, 62 (2002).
123. B. Yurke, A. J. Tuberfield, A. P. Mills, Jr., F. C. Simmel, and J. L. Neumann, *Nature (London)* **406**, 605 (2000).
124. F. S. Collins, E. D. Green, A. E. Guttmacher, and M. S. Guyer, *Nature* **522**, 835 (2003).

JILL REHMANN
St. Joseph's College

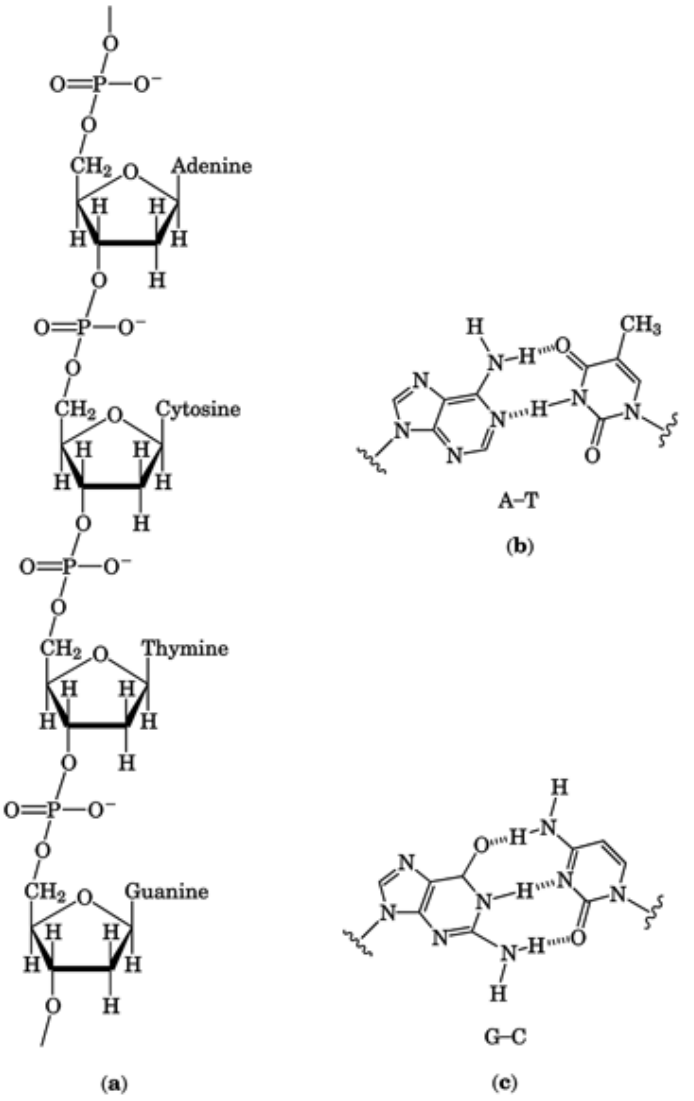


Fig. 1. Elements of DNA structure: (**a**) a deoxypolynucleotide chain, which reads d(ACTG) from $3' \rightarrow 5'$ or d(GTCA) from $3' \rightarrow 5'$ and (**b**) and (**c**) the Watson–Crick purine–pyrimidine base pairs, A–T and G–C, respectively, where — represents attachment to the deoxyribose.

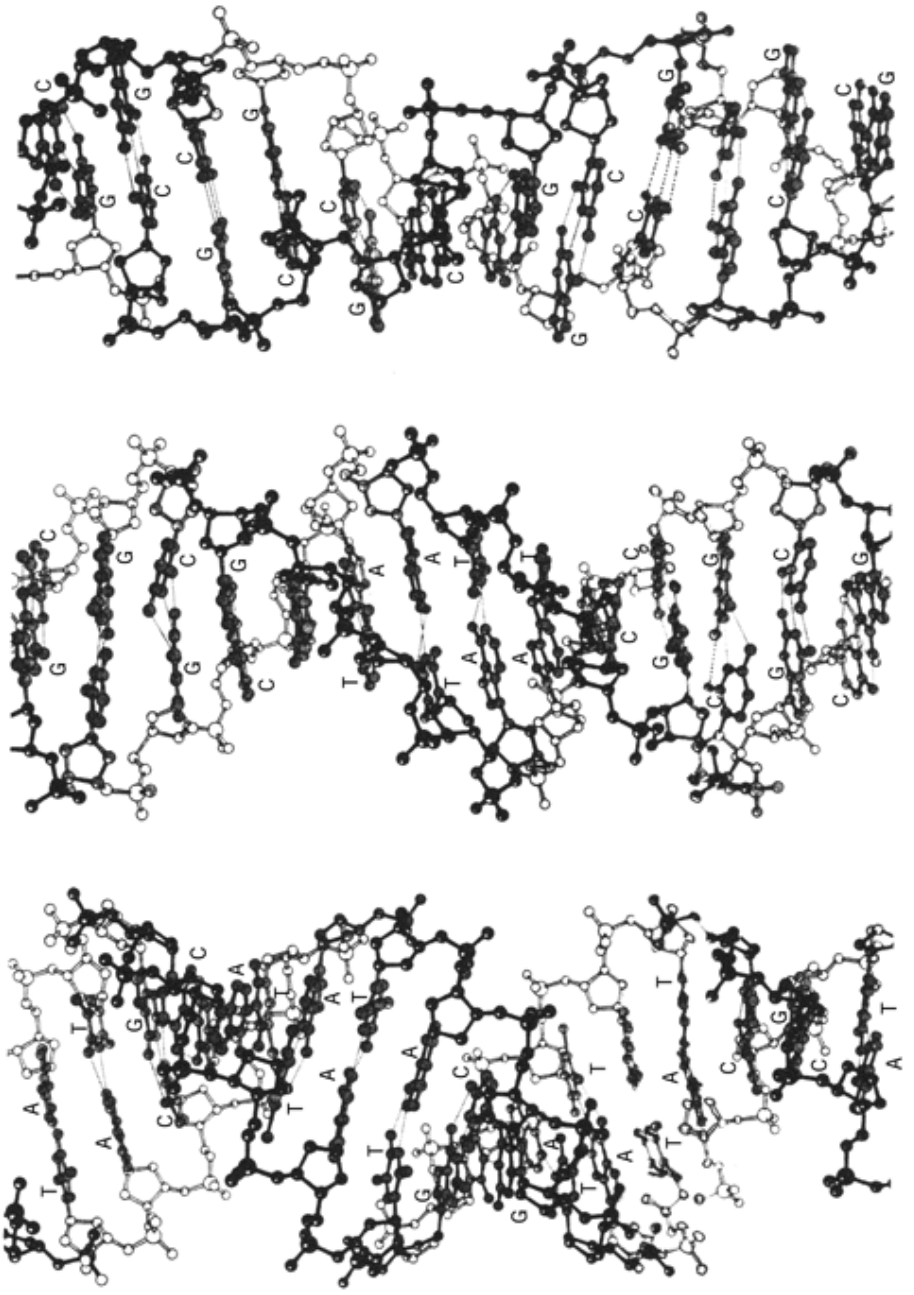


Fig. 2. Comparison of three forms of DNA: (**a**) A-DNA, (**b**) B-DNA, and (**c**) Z-DNA.

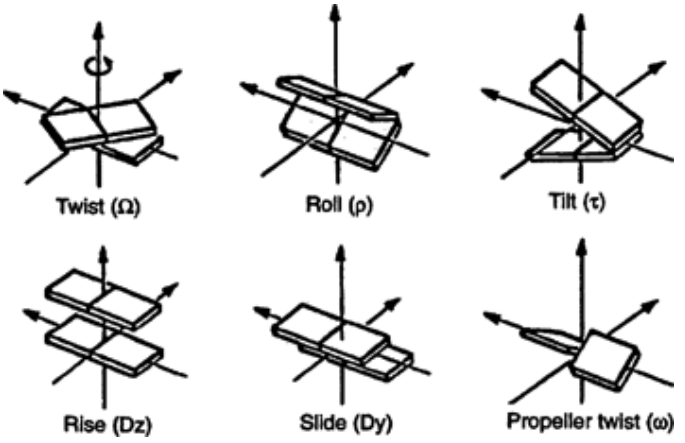


Fig. 3. Definitions of the most useful local variables in B-DNA helices; twist, roll, tilt, rise, and slide from one base pair to the next, and propeller within one base pair.

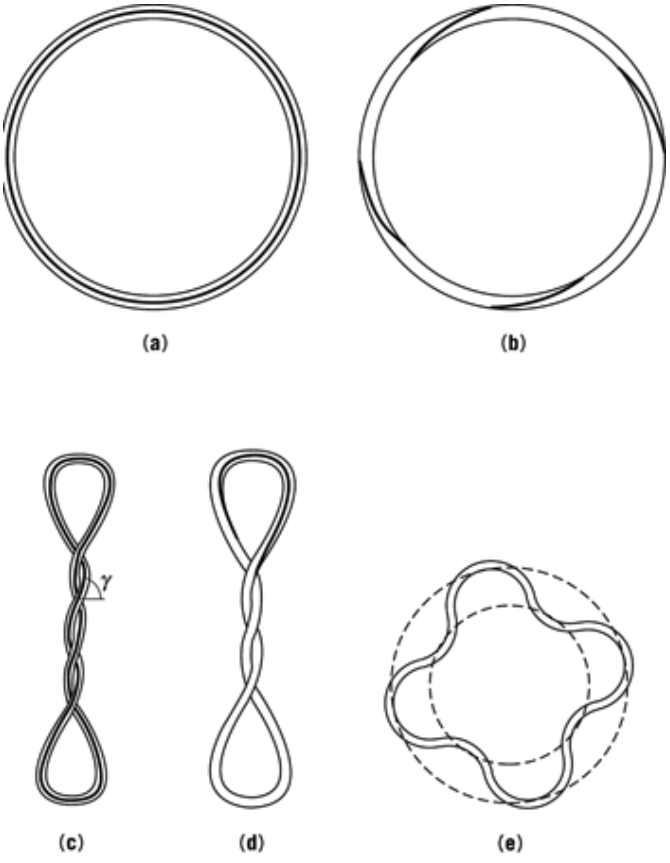


Fig. 4. Laboratory tubing model for supercoiling in closed-circular DNA: (**a**) relaxed DNA, $\Delta Lk = 0$; (**b**) $\Delta Lk = \Delta Tw = -4$; (**c**) $\Delta Lk = \Delta Wr = -4$; (**d**) plectonemic supercoil, $\Delta Lk = -4$; and (**e**) toroidal winding of DNA. (see text).

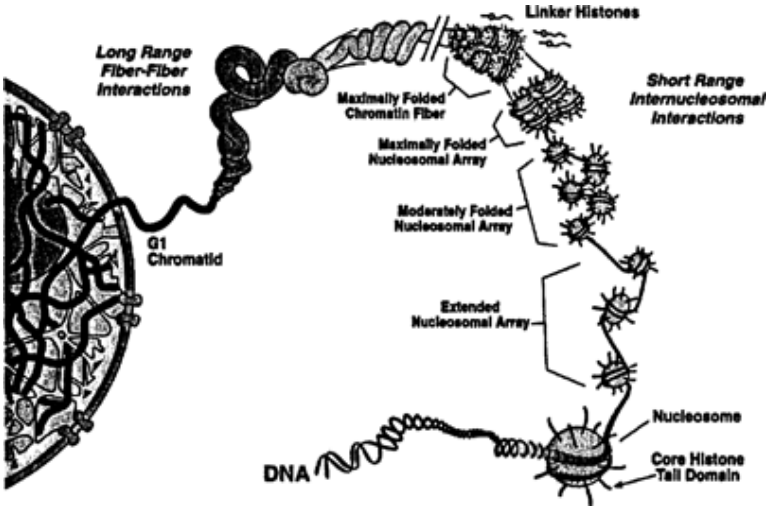


Fig. 5. Chromatin fiber condensation.

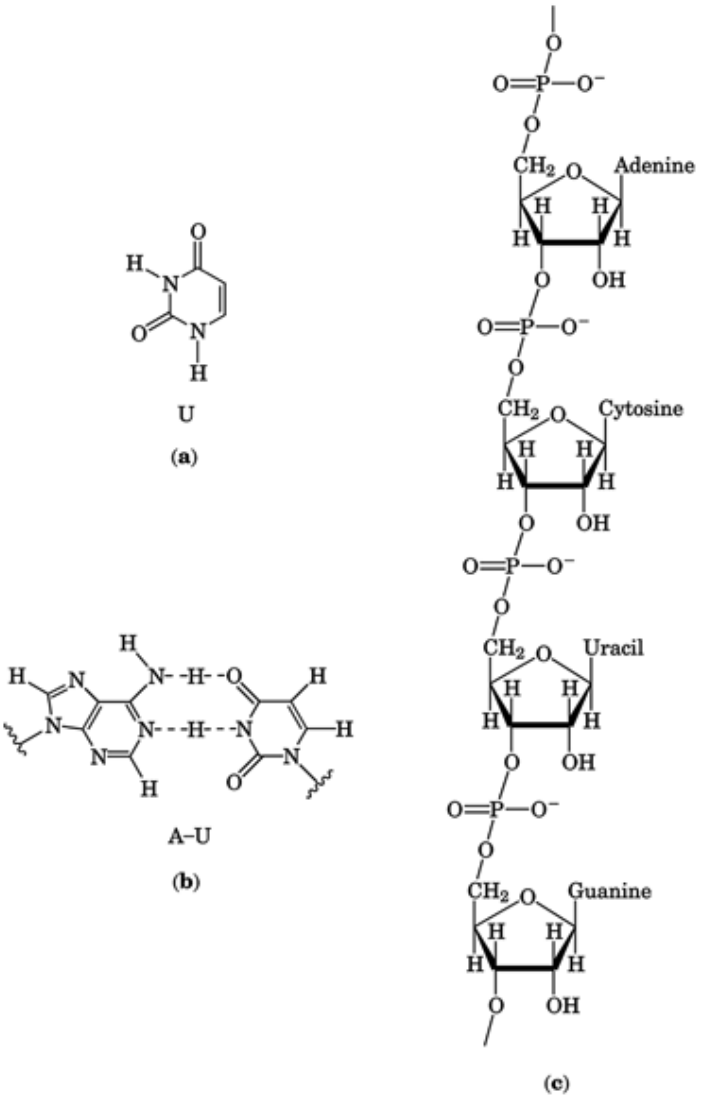


Fig. 6. Elements of RNA structure: (**a**) uracil; (**b**) a Watson-Crick A-U base pair; and (**c**) a polynucleotide chain that reads from $5' \rightarrow 3'$ ACUG, and from $3' \rightarrow 5'$ GUCA.

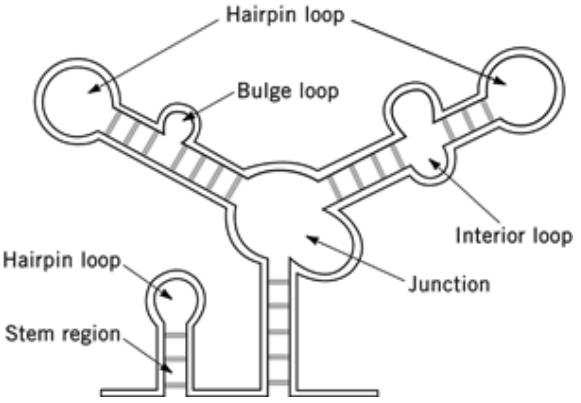


Fig. 7. Structural features of folded RNA where represents base pairing.

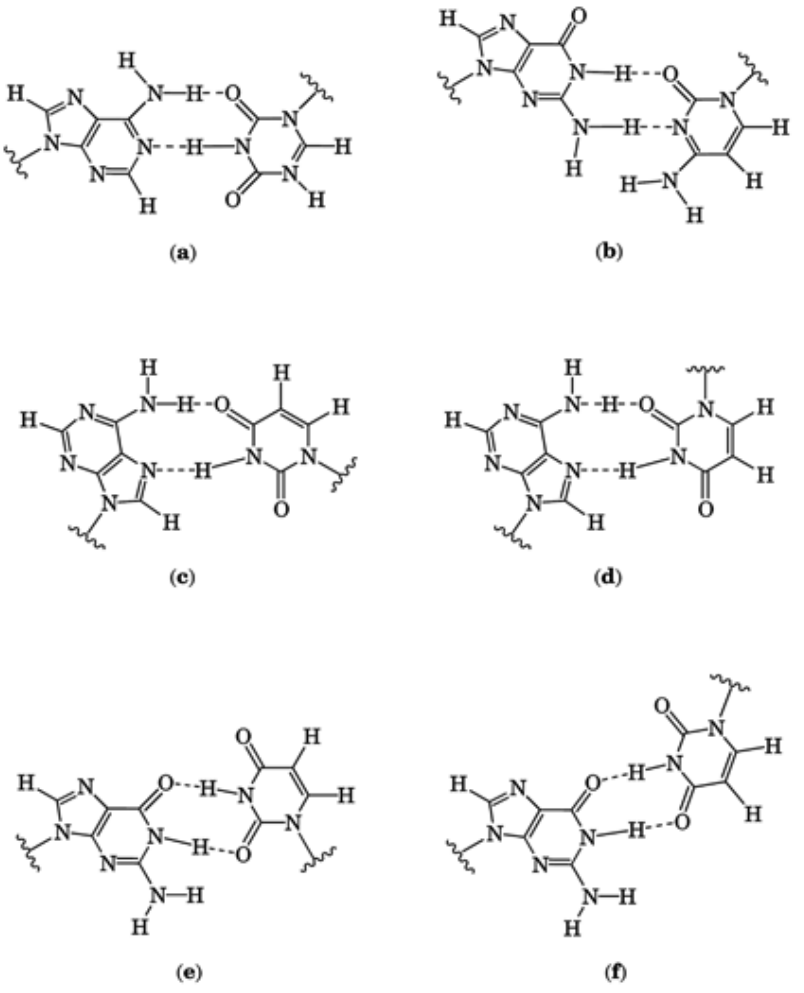


Fig. 8. Non-Watson-Crick base pairs occurring in double-stranded RNA where — represents the site of attachment to the sugar: (**a**) A-U reverse-Watson-Crick; (**b**) G-C reverse-Watson-Crick; (**c**) A-U Hoogsteen; (**d**) A-U reverse-Hoogsteen; (**e**) G-U wobble; and (**f**) G-U reverse-wobble.

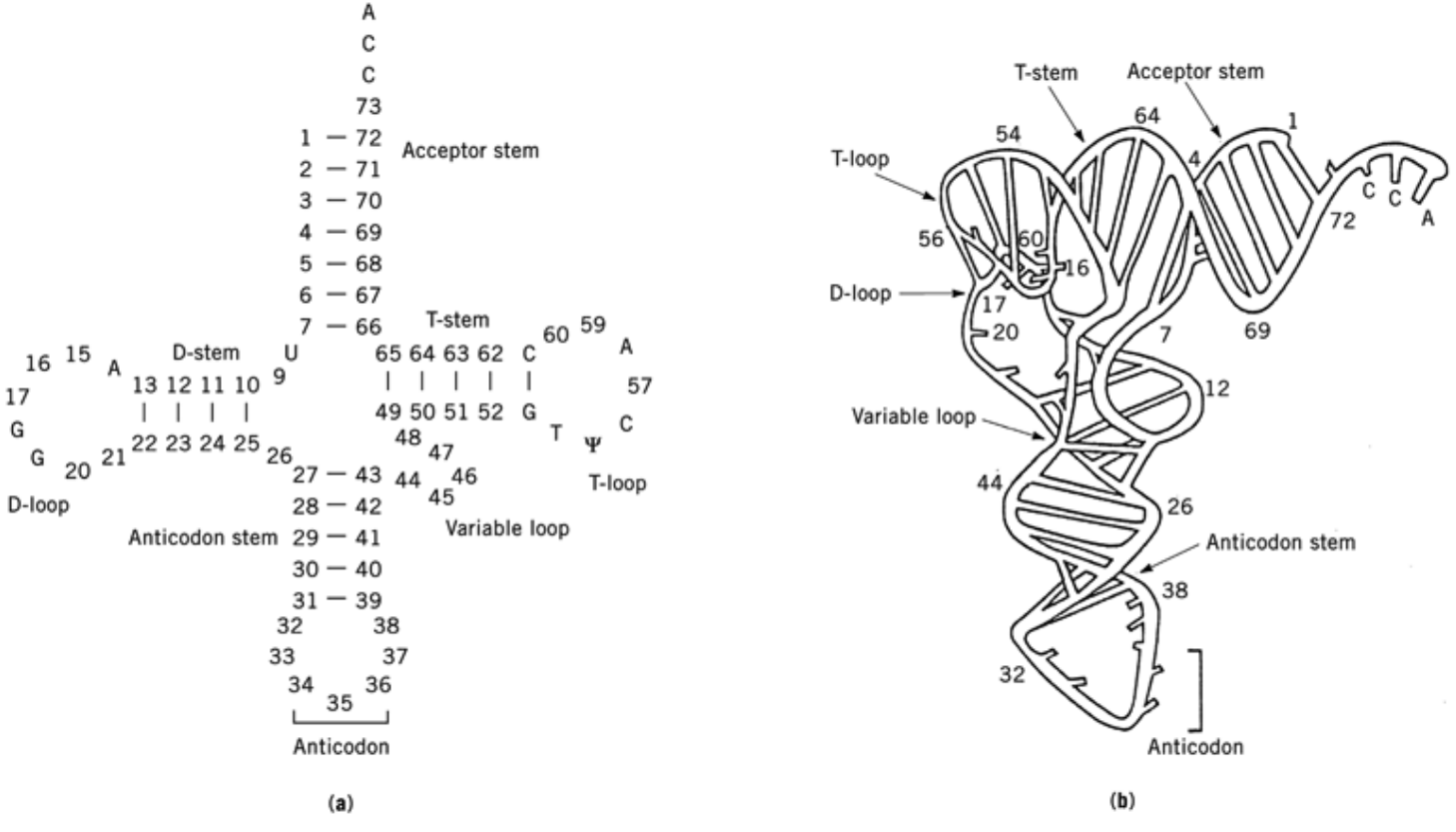


Fig. 9. Structures of yeast phenylalanine tRNA: (**a**) secondary structure of tRNA, where conserved base pairs are indicated by lines; and (**b**) tertiary structure.

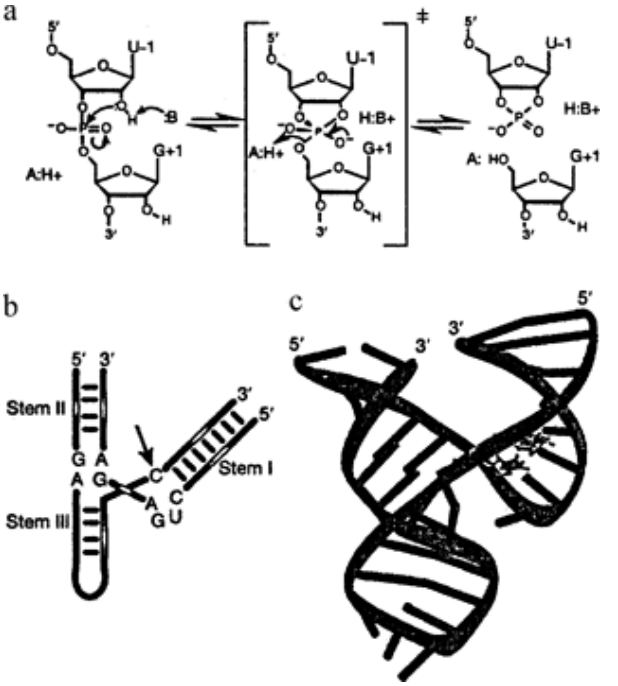


Fig. 10. (**a**) General mechanism for ribozyme cleavage reactions. (**b**) Secondary structure of the hammerhead ribozyme with illustrating the catalytically essential nucleotides. (**c**) Tertiary structure of the hammerhead ribozyme obtained by crystallography.

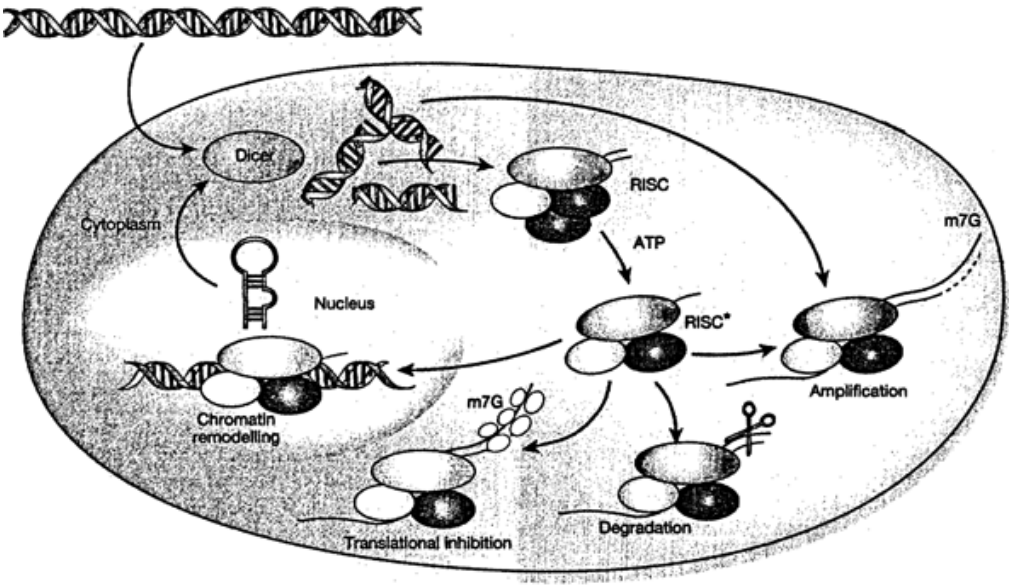


Fig. 11. The mechanism of RNAi.

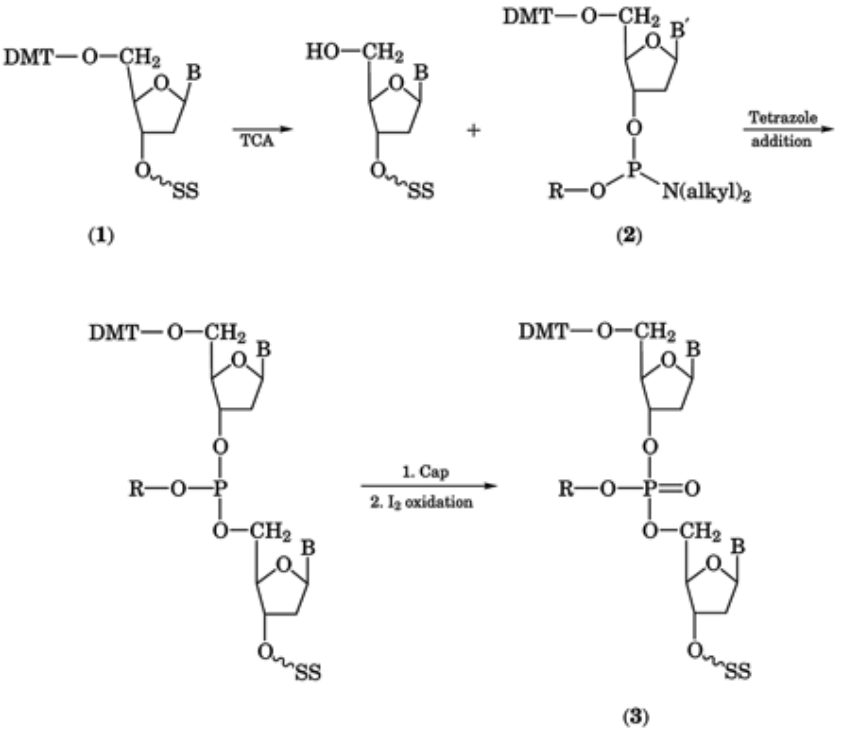


Fig. 12. The phosphoramidite method for synthesizing oligodeoxynucleotides, where B and B′ represent one of the protected bases thymine, N^4-benzoylcytosine, N^6-benzoyladenine, or N^2-isobutyrylguanine; $-N(alkyl)_2$ is $N(CH(Me_3)_2)_2$, DMT is *p*-(dimethoxytrityl); R is $NCCH_2CH_2$; SS is silica on a solid support; and TCA is trichloroacetic acid. (see text).

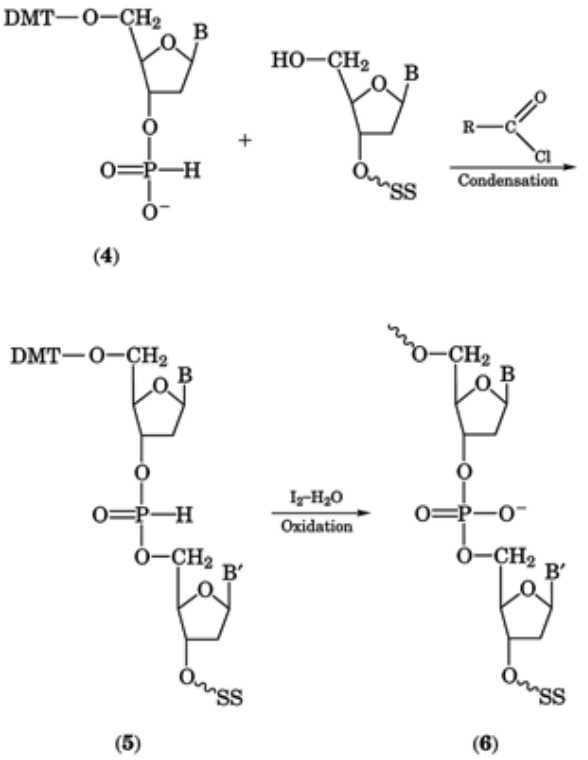


Fig. 13. *H*-Phosphonate method where is pivaloyl chloride or adamantoyl chloride; B, B′, DMT, and SS are as defined in Figure 11.

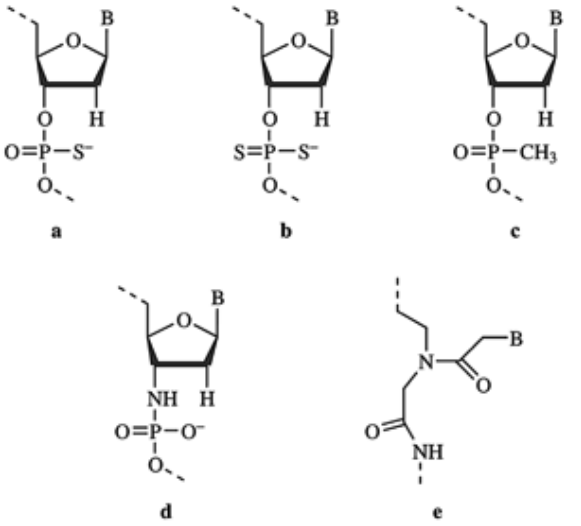


Fig. 14. Oligonucleotides with modified backbones (**a**) Phosphorothioate DNA. (**b**) Phosphorodithioate DNA. (**c**) Methylphosphonate DNA. (**d**) N3′-P5′ Phosphoroamidate DNA. (**e**) Peptide nucleic acid (PNA).

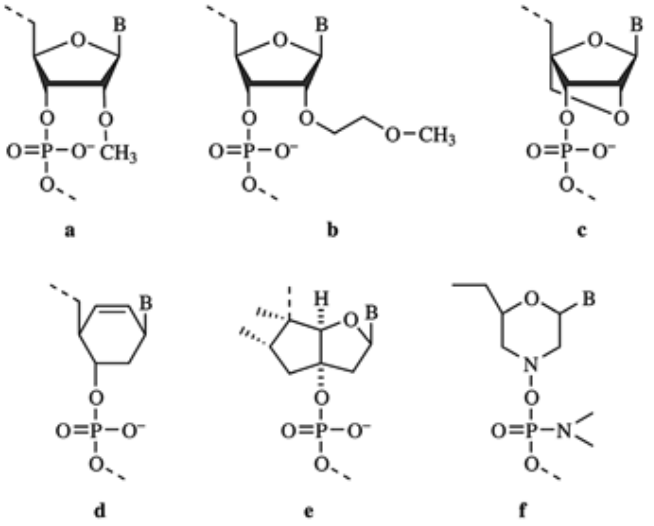


Fig. 15. Oligonucleotides with modification or replacement of the ribose unit. (**a**) 2′-*O*-Methyl. (**b**) 2′-*O*-Methoxy ethyl. (**c**) Locked nucleic acid (LNA). (**d**) Cyclohexene nucleic acid. (**e**) Tricyclo-DNA. (**f**) Morpholino phosphoramidate.

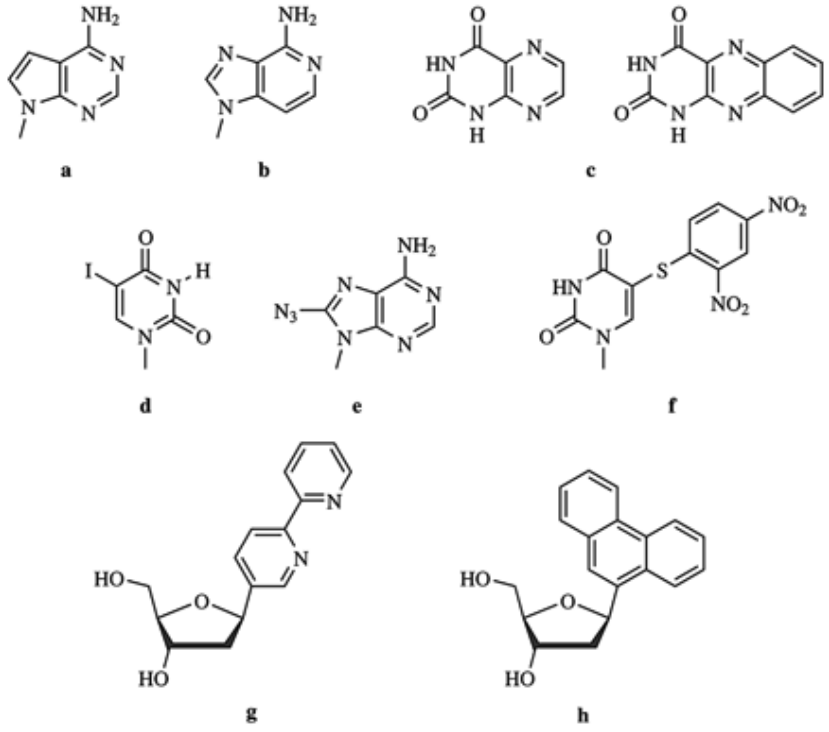


Fig. 16. Modified bases or base substitutions in oligonucleotides. (**a**) 7-Deazaadenine. (**b**) 3-Deazaadenine. (**c**) Areneo[*g*]lumazine analogues of thymidine. (**d**) 5-Iodouridine. (**e**) 8-Azidoadenosine. (**f**) 2,4-Dinitrophenyl modified uridine. (**g**) Bipyridyl. (**h**) Phenanthrene.

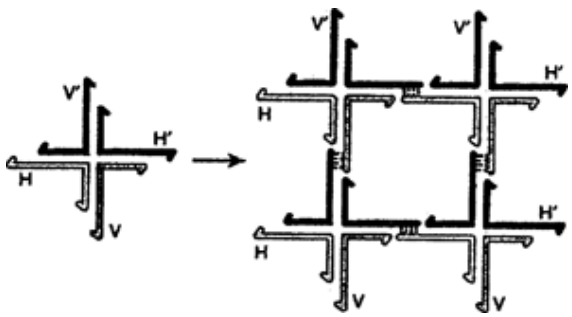


Fig. 17. Self-assembly of complementary oligonucleotides. The first structure represents a Holliday junction formed from four suitably complementary oligonucleotides with sticky end overhangs. Both V and V' and H and H' are complementary sticky ends. The sticky ends allow self-assembly of individual Holliday junctions to create a grid-like arrangement (117). Reprinted with permission from *Nature (London)* **421**, 427–431 (2003).

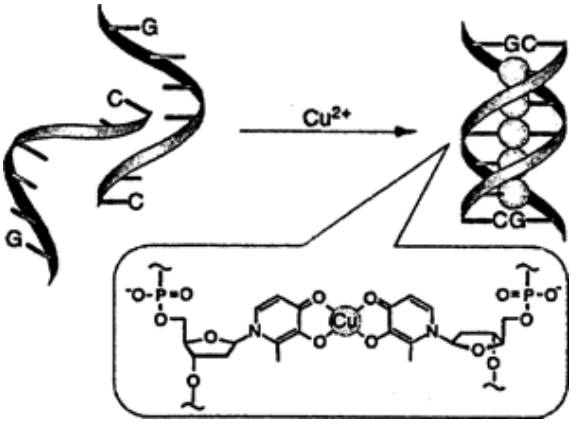


Fig. 18. Self-assembly of metal ions in the center of a helix formed by two modified oligonucleotides (120). Reprinted with permission from *Science* **299**, 1212–1213. Copyright 2003 AAAS.