Nucleic Acid Structural Energetics

Jeffrey R. Vieregg
California Institute of Technology, Pasadena, CA, USA

Nucleic acids perform many functions essential for life, and exhibit a correspondingly diverse array of structures. This article provides an overview of nucleic acid structure, as well as the forces that govern its formation. The current state of knowledge of nucleic acid thermodynamics is discussed, as well as techniques for predicting and designing structures of interest. Experimental methods used to determine the structure of nucleic acids and the thermodynamics of their reactions are also surveyed.

Update based on original article by Charles C. Hardin and Andrea F. Moon Encyclopedia of Analytical Chemistry, © 2000, John Wiley & Sons Ltd.

Encyclopedia of Analytical Chemistry R.A. Meyers (Ed.)

1 INTRODUCTION

Two related nucleic acids, DNA and RNA, are responsible for a host of processes that are essential to cellular life. DNA's role as the storehouse of genetic information has been known for decades, but details of the process by which that information is read out continue to emerge. RNA, long thought to be either a passive messenger or an inert structural adapter between DNA and protein, turns out to have many functions. RNA is involved in genetic regulation, immune responses, and even catalyzes (alone or with protein cofactors) many vital chemical reactions. Most of these functions require nucleic acids to fold into specific three-dimensional structures. Concomitant with our growing recognition of the diverse functions of natural nucleic acids has been an explosion of interest in designing new RNA and DNA molecules to carry out therapeutic and diagnostic roles. Understanding the structural capabilities of nucleic acids, as well as the energetics of their interactions, is essential as we seek to comprehend their diverse roles and capabilities.

2 STRUCTURAL ORGANIZATION

To begin, we examine the structural organization of nucleic acids in increasing order of scale and complexity from one-dimensional polymer chains to three-dimensional active biomolecules.

2.1 Primary Structure: Nucleic Acid Polymers

The monomer building blocks of RNA and DNA are called *nucleotides* (Figure 1), and consist of three parts: a five-carbon β-D-furanose sugar, a phosphate group, and a semiaromatic nitrogenous base, or nucleobase. Nucleotides are linked into polymers through phosphodiester bonds to the 3′ and 5′ carbon atoms of the sugars. In the cell, nucleic acid polymers are synthesized in the 5′ to 3′ direction; this process is carried out by polymerase enzymes and powered by hydrolysis of nucleotide triphosphates. Chemical synthesis of oligonucleotides typically proceeds in the opposite direction (3′ to 5′) utilizing nucleoside phosphoramidite monomers.

Ribose is the sugar for RNA, 2′-deoxyribose for DNA. Although often drawn as planar, the sugar ring is in fact puckered, with (typically) one carbon displaced ~50 pm out of the ring plane. The C2′-*endo* and C3′-*endo* configurations are in equilibrium for single-stranded nucleic acids, but the C2′-*endo* pucker is energetically preferred for (B-form) duplex DNA under physiological buffer conditions. The 2′ hydroxyl group present in RNA produces a steric clash in the C2′-*endo* configuration, so C3′-*endo* predominates for double-stranded RNA.

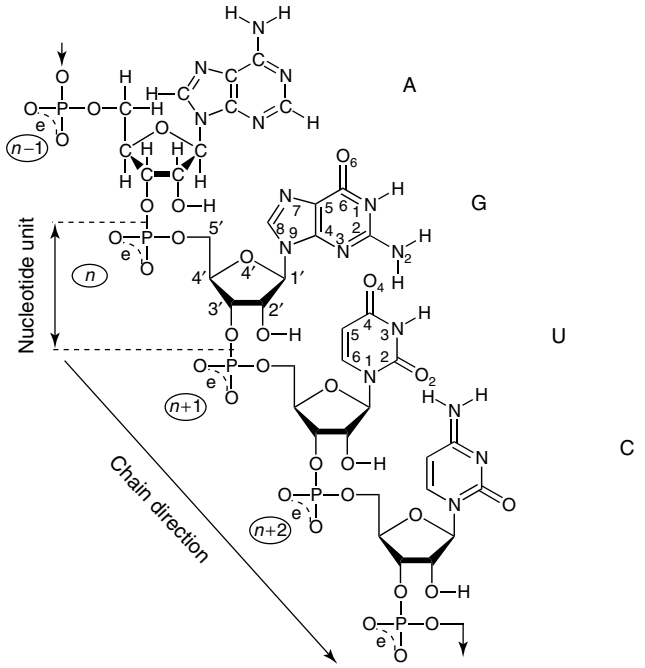

Figure 1 Components of a chain of ribonucleotides, with atom numbering. All hydrogens are drawn for the first nucleotide, only functional hydrogens for the others. By convention, nucleic acid polymers sequences are listed from 5′ to 3′: the notation AGUC represents this sequence containing the bases adenine, guanine, uracil, and cytosine. A DNA polynucleotide would have the sugars' 2′ OH groups replaced by hydrogens and uracil replaced by thymine (5-methyl-uracil). (Reproduced from Ref. 1. © Springer-Verlag, 1984.)

In addition to its conformational effect, the 2′ hydroxyl of RNA is chemically active as well. The most common degradation pathway for RNA is base- or metal-catalyzed hydrolytic backbone scission, wherein the 2′ oxygen attacks the adjacent 3′-linked phosphorus and breaks the phosphodiester linkage to the next nucleotide in the chain. Many ribozymes make use of the lability of the 2′ hyrdoxyl group to carry out their reactions as well. The 2′ position in DNA sugars is unreactive, which makes DNA much more chemically stable. The difference in reactivity between DNA and RNA makes sense given their different functions. DNA is the archive of genetic material, and should therefore be more stable than RNA, whose turnover allows the cell to respond to changing conditions.

The pK_a of the phosphates in nucleic acid polymers is near 1, so they are negatively charged at physiological pH. The nitrogenous bases are uncharged at these pH values. Nucleic acids are thus highly charged molecules and would not fold or form base pairs in the absence of neutralizing ions from the surrounding solution. The effect of cations on the thermodynamics of folding is discussed in Section 3.4.

In recent years, great interest has developed around the possibility of using nucleic acids as programmable therapeutics, either by taking advantage of natural genetic regulation pathways such as RNA interference[2] and antisense[3] or as targeting devices for other molecules. One of the primary challenges for these approaches is delivering nucleic acids into cells, as their large negative charge prevents them from crossing cell membranes. Nucleic acid analogs with uncharged backbones have been developed to address this, and promising results have been achieved with both methyl phosphonate and morpholine backbones,[4] as well with peptide nucleic acids (PNAs, *see* **PNA and Its Applications**).

In contrast to proteins, which are built of 20 amino acids with diverse chemical properties, nucleic acids perform their vital functions with only five nitrogenous bases. The bases, shown in Figure 2, are of two types. Adenine and guanine are members of the bicyclic class of molecules called *purines*. Cytosine, thymine, and uracil are pyrimidines; thymine is found in DNA and uracil in RNA. Both purines and pyrimidines are planar, feature partially π-delocalized electrons, and contain ionizable heteroatoms. In acidic solution, the imine nitrogens of adenine ($pK_a \sim 3.5$) and cytosine ($pK_a \sim 4.2$) can be protonated, giving the bases a positive charge. At high pH values, guanine, uracil, and thymine can be deprotonated (pK_a 9.2–9.7), resulting in a negative charge on the base. At neutral pH, which is the physiological case for most organisms, the bases are therefore largely uncharged. Each of the bases has the ability to both accept and

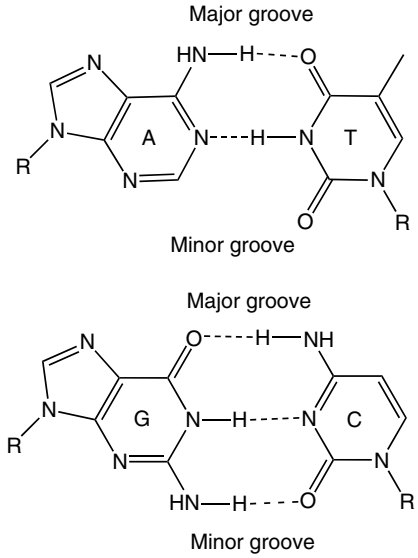

Figure 2 Watson–Crick pairing of DNA nucleobases. A·T pairs form two hydrogen bonds and G·C pairs form three; both pairs have the same distance between the C1′ sugar atoms (marked as R). When stacked with other pairs, the bases are exposed to solvent on two sides, the major and minor grooves.

donate hydrogen bonds, and this plays a key role in determining the equilibrium structure of a nucleic acid, as discussed below.

In addition to the canonical bases, a number of modified bases are employed by the cell for specific roles, including pseudouridine (found in tRNA), dihydrouridine (tRNA and rRNA), and methylated bases such as 5-methylcytidine and 7-methylguanosine, as well as many others. Chemists have also developed numerous synthetic base analogs in order to elucidate biological processes or alter them in various ways.[5]

The sequence of bases in a DNA or RNA molecule is referred to as its *primary structure*, and contains the information needed for the molecule to perform its biological function, whether that is coding for a protein, binding to a target, or catalyzing a chemical reaction. Many of these functions also require nucleic acids to adopt a specific three-dimensional structure. In the following sections, we describe common nucleic acid structures and the forces that give rise to them.

2.2 Secondary Structure: Base Pairing

The most common structure adopted by nucleic acids is the base-paired duplex. In this structure, pairs of nucleotides align such that their nucleobases are coplanar and complementary hydrogen bond donors and acceptors are aligned (Figure 2). This process is called *hybridization*. Guanine and cytosine form three hydrogen bonds when paired, while adenine forms only two with either thymine (in DNA) or uracil (in RNA). These are the canonical, or Watson–Crick, base pairs. Stacking adjacent base pairs upon each other produces the familiar double helix structure (Figure 3), in which the two strands (or two regions of the same strand) spiral around one another in an antiparallel configuration with the charged phosphates on the outside and the bases in the interior. As each contains one purine and one pyrimidine nucleobase, G·C and A·T/U pairs are roughly the same size, which allows the duplex to accommodate arbitrary sequences provided that they are complementary.

In physiological solutions, double-stranded DNA adopts the configuration shown in Figure 3a, which is known as *B-form*. In this configuration, the bases are stacked with their faces perpendicular to the helix axis and an average spacing of 0.33 nm per nucleotide. Each pair is offset by an average of 36°, resulting in a helix with a pitch of 10 base pairs (bps) and a width of 2.0 nm. As discussed above, the sugars in a B-form helix largely adopt the 2′-*endo* configuration. The bases are exposed to the outside environment on two sides (Figure 2), which appear as grooves in space-filling structural representations. One groove (the minor groove) is largely occluded by the sugar–phosphate backbone, while the other (the major groove) is much wider and thus more accessible for binding by enzymes and other recognition molecules. The major groove also has higher information content than the minor groove: reversal of a pair (A·T vs T·A or G·C vs C·G) produces the opposite pattern of hydrogen bond donors and acceptors in the major groove while leaving the minor groove pattern unchanged. As a result, most DNA-binding molecules that recognize specific sequences bind in the major groove.

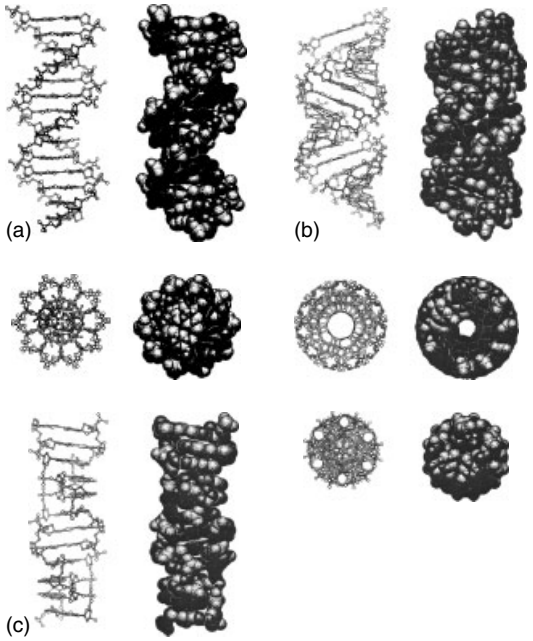

Figure 3 Common nucleic acid double helices: (a) B-form DNA: bases perpendicular to helical axis, wide major groove, narrow minor groove. (b) A-form DNA: bases inclined at 19° to the helical axis, narrow major groove, wide minor groove. (c) Z-DNA: left-handed spiral, two base nucleotide repeating units inclined opposite to the A-form. (Reproduced from Ref. 6. © Springer-Verlag, 1984.)

The 3′ hydroxyl of RNA is too bulky to comfortably fit into a B-form duplex, which forces double-stranded RNA and DNA–RNA hybrid duplexes into an alternate geometry called *A-form* (Figure 3b). The A-form helix is wider (2.6 nm) and shorter (0.23 nm bp^{-1}) than the B-form, and the pitch differs as well, at 10.7 bp per helical turn. In contrast to the perpendicular stacking of the B-form duplex, A-form bases are tilted 19° with respect to the axis of the helix. The relative widths of the two grooves are reversed in A-form helices: the major groove is narrow and deep and the minor groove is wide and shallow.

In addition to A- and B-form helices, a wide variety of other geometries are also possible and can occur under special circumstances. One example is Z-DNA (Figure 3c), which forms in regions of torsional stress or at very high salt concentrations. In contrast to the right-handed A- and B-form helices, Z-DNA is left-handed

and consists of pairs of base pairs that zigzag about the helical axis. The helix is 1.8 nm wide with a pitch of 12 bp over 4.56 nm. The bases are tilted 10° in the opposite direction relative to the A-form and the grooves are roughly equal in width.

Except for short sections undergoing replication or transcription, natural DNA molecules are almost entirely double-stranded helical duplexes, in keeping with their role of information storage. RNA, by contrast, adopts a diverse and complex array of structures. Synthetic DNA molecules are also often engineered to adopt specific structures. The secondary structure of a nucleic acid is defined by its base pairing. Mathematically, for a molecule with N nucleotides, the secondary structure is defined as the ordered list of base pairs $(i \cdot j)$ between nucleotides i and j such that $i < j$. This divides a structure into a collection of nonoverlapping base-paired helices and unpaired loops, bulges, and junctions, as illustrated in Figure 4. As is discussed in Section 3, secondary structure dominates the thermodynamics of RNA folding and is also the most amenable to prediction and systematization.

The most common secondary structure motif is the stem loop (also called a *hairpin loop*), in which a nucleic acid strand folds back on itself to produce a helical duplex with an unpaired loop at one end. If an unpaired region is flanked by two paired regions, it is referred to as either an *internal loop* or a *bulge*, depending on whether unpaired bases are present in both strands or just one. Regions where more than two helices meet are called *junctions*. Depending on the number and identity of unpaired bases between helices, junctions can either be minimally structured or highly constrained; ribozyme active sites (*see* **Structural Analysis of Ribozymes**) and ligand binding pockets (*see* **Aptamers**) are often located in junction regions as their flexibility allows for a diverse variety of chemical microenvironments.

2.3 Tertiary Structure

The tertiary structure of a nucleic acid is formed by the interaction of secondary structure elements, including unpaired bases, and is responsible for its final three-dimensional form. RNA tertiary structure is discussed in detail in its own article (*see* **RNA Tertiary Structure**), but a few general principles and common structures are listed here.

Many tertiary structures form as a result of pairing between single-stranded regions (such as loops and bulges) of secondary structure. These interactions, such as the kissing loops shown in Figure 5, can bring together sections of a molecule that are quite distant in terms of sequence and thus contribute to formation of compact three-dimensional folds. Another important tertiary motif is the tetraloop–receptor interaction (Figure 6), in which the unpaired bases of a stem loop bind

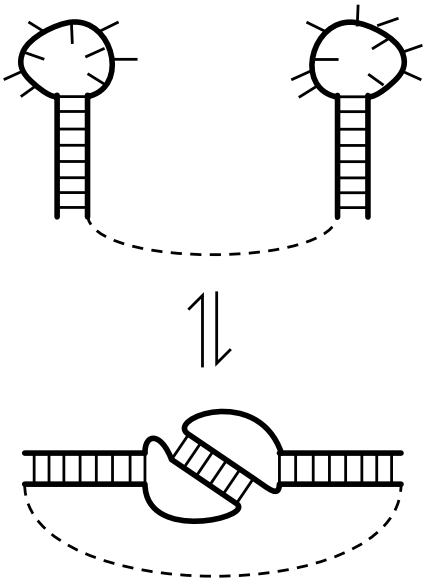

Figure 5 Kissing loops form when the unpaired bases of two hairpin stem-loops are complementary. The helix formed by pairing of these bases is often stacked with one or both of the stems. Factors determining kissing-loop stability include the number and identity of the pairing bases as well as the flexibility of the loops and the sequence (if there is one) that links the two hairpins.

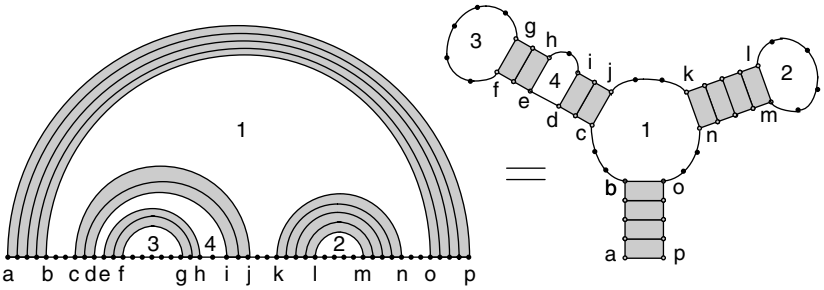

Figure 4 Nonoverlapping base pairs define the secondary structure of a nucleic acid. Secondary structure motifs include paired helices (gray) and unpaired hairpin loops (2 and 3), bulges (4), and junctions (1). Internal loops (not shown) resemble bulges but have unpaired bases on both strands.

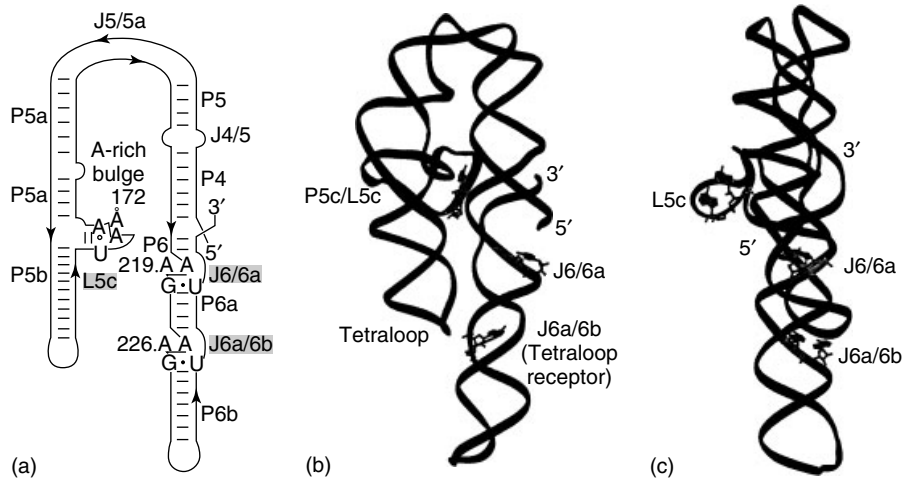

Figure 6 (a) Secondary structure of a tetraloop–receptor interaction from the *Tetrahymena thermophila* ribozyme. Hairpin loop L5c docks to a receptor (J6a/6b) within a duplex. A second receptor (J6/6a) forms a tertiary contact with an A-rich bulge. (b), (c) Together, the two tertiary interactions fold the sequence into a compact shape. (Reproduced from Ref. 7. © American Association for the Advancement of Science, 1996.)

to specific sequences within a paired duplex. Two helices can also interact either by stacking coaxially to maximize hydrophobic interactions (Section 3.1) or through alignment of the wide, shallow minor grooves present in the A-form duplex.

For the most part, the interactions that govern tertiary structure formation involve small numbers of bases (due to the rigidity of secondary structure helices), and require the presence of multivalent metal cations in specific locations to neutralize the electrostatic repulsion of the phosphate backbone. One tertiary motif that breaks both these rules is the pseudoknot motif (Figure 7), in which an unpaired region pairs to the terminal loop of a stem-loop structure.[8] If the loop is large and the unpaired region is relatively unconstrained (it must wrap around the loop bases, which are topologically fixed[9]), large numbers of base pairs can potentially be formed, resulting in a structure that is quite stable. Pseudoknots have been implicated in a variety of biological roles, including both catalysis and gene regulation.[10]

2.4 Other Structural Forms

Nucleic acids can form structures that do not easily fit into the secondary/tertiary classification scheme. These structures often feature non-canonical base pairing and usually consist of specific sequences that fulfill specific biological functions, as discussed in **DNA Structures of Biological Relevance, Studies of Unusual Sequences**. One example of this is the DNA triple helix formation, in which purine bases form hydrogen bonds not only on the Watson–Crick face but also on the side that normally faces the major groove. This is known as *Hoogsteen pairing*, and is illustrated in Figure 8. Note that one of the cytosine bases in the $C^{+}{\cdot}G{\cdot}C$ triplet is protonated – the high density of negatively charged phosphates in triplex DNA raises the pK_a sufficiently to allow protonation at neutral pH. Methylation of dC also stabilizes the protonated form. Formation of triplex structures requires long runs of purines and pyrimidines, enabling their prediction via sequence analysis. Individual RNA base triples also sometimes form as part of loop–helix tertiary contacts.

DNA and RNA can also form four-stranded structures. One common motif is the G-quadruplex (or G-quartet), which is pictured in Figure 9. This occurs when four guanine bases form a closed square that is stabilized by both Watson–Crick and Hoogsteen hydrogen bonding.[13] Multiple squares can stack upon each other to create an extremely stable structure. Quadruplexes require G-rich sequences, and are particularly common in the telomere region of chromosomes, as well as transcription promoters and regulatory RNA sequences.[14] Four-stranded complexes consisting of two intercalated DNA duplexes also exist, and are known as *i-motif* DNA. RNA rarely forms these structures, as the 2′ hydroxyl group is too bulky to fit comfortably within their compact geometry. Just as in the triplex case, quadruplex-forming sequences contain characteristic sequence patterns that can often be identified using software searches.

3 ENERGETICS

The division between secondary and tertiary structure may seem arbitrary, but it, in fact, reflects a deep truth about nucleic acid folding. Unlike proteins, nucleic acids fold in a largely hierarchical manner.[16] Most of the free

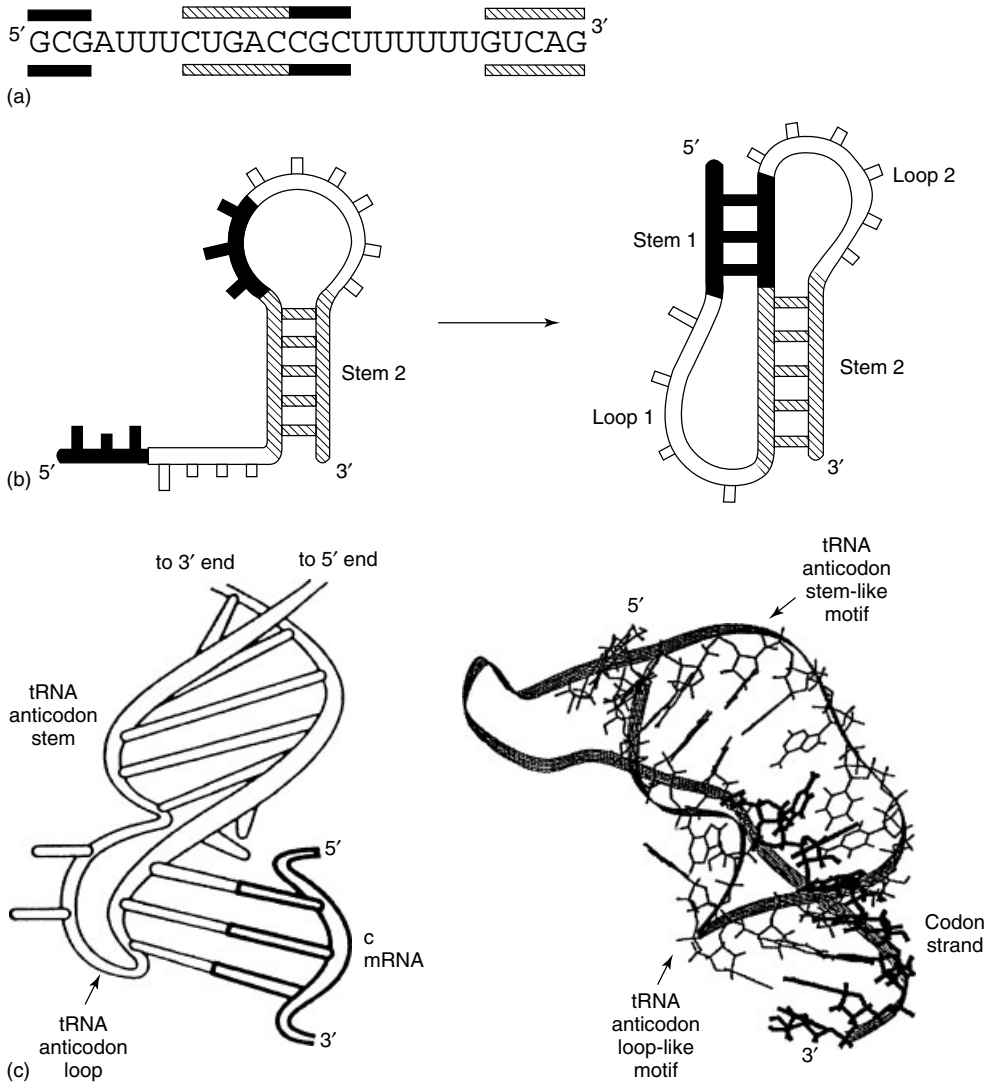

Figure 7 (a) Sequences with alternating complementary regions are likely to form pseudoknots. (b) Pseudoknot formation typically begins with formation of a hairpin loop, followed by pairing of bases within the loop to an outside complementary region. (c) The helices of many pseudoknots are at least partially stacked, an energetically favorable tertiary interaction also seen in tRNA-mRNA codon-anticodon binding. (Reproduced from Ref. 8. Copyright © 1989 John Wiley & Sons, Reprinted by permission of Wiley-Liss, Inc., a division of John Wiley & Sons, Inc., and Reproduced from Ref. 11 © Wiley Periodicals, Inc., 1995.)

energy gained by folding comes from forming secondary structures, with tertiary interactions providing (despite their biological importance) a much smaller contribution. Secondary structure elements from large molecules are usually stable as isolated sequences, implying that their energies are largely context independent, again in contrast to protein structures. In other words, one can measure or compute the folding thermodynamics of a secondary structure motif and reasonably expect that the result will still be relevant in the context of a larger molecule. Experimental and theoretical evidence also indicates that the nucleic acid secondary/tertiary structural hierarchy extends beyond thermodynamics to folding kinetics, with secondary structure forming first and helping to dictate which tertiary contacts occur.[17] Finally, secondary structure thermodynamics has been extensively measured and semiempirical models developed that allow prediction of the energy and likely structures formed by an arbitrary sequence. This section discusses what is known about the thermodynamics of nucleic acid folding at both levels of organization, as well as the forces governing them.

3.1 Secondary Structure

As discussed above, the secondary structure of a nucleic acid is defined by its base pairing. A common misconception is that the main driving force for formation

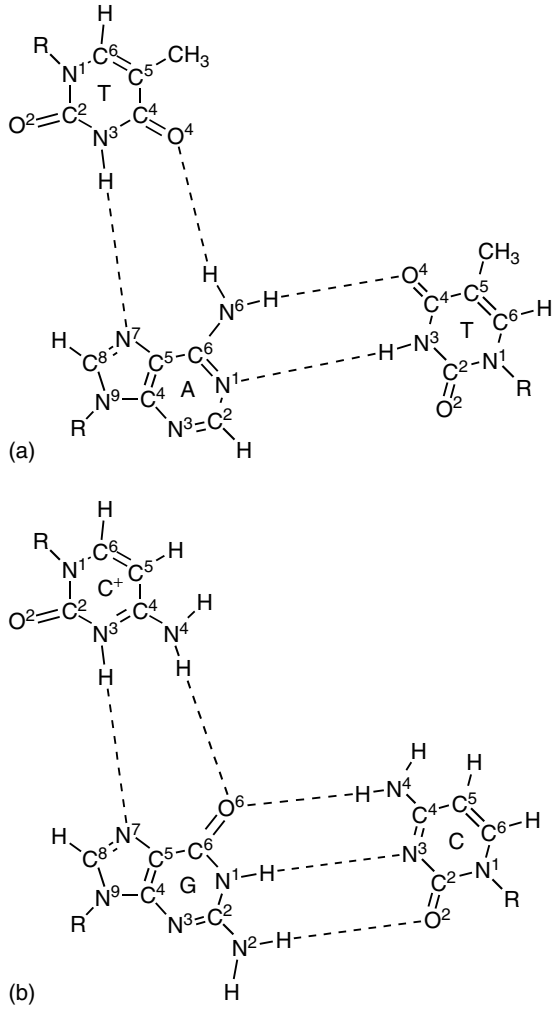

Figure 8 T·A·T and C^+·G·C base triplets. (Reproduced with permission from Ref. 12. © Annual Review, 1994.)

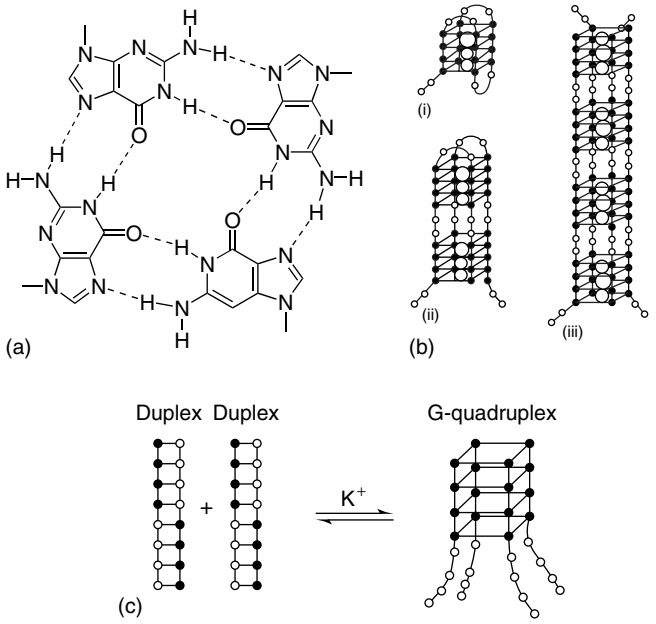

Figure 9 (a) G-quadruplex hydrogen bonding scheme. (b) Quadruplexes formed from (i) one, (ii) two, and (iii) four strands. Closed circles refer to guanines, small open circles indicate pyrimidines, and large open circles represent octahedral coordination complexes involving monovalent cations caged within two stacked G-quartets. (c) K^+-induced formation of G-quadruplex structure from the self-complementary duplex d(G4C4). (Reproduced from Ref. 13. Copyright © 1995 John Wiley & Sons, Inc., reprinted by permission of John Wiley & Sons, Inc., and reproduced from Ref. 15. © American Association for the Advancement of Science, 1990.)

of base-paired helices is the formation of hydrogen bonds between complementary bases. The pattern of H-bond donors and acceptors on the bases is a major factor in controlling the specificity of pairing, but it is only one (and not the dominant one) of several forces that determine the equilibrium between the folded and unfolded states.

Just as for their better-studied protein cousins, nucleic acid folding reflects a struggle between powerful stabilizing and destabilizing forces. These forces have been studied extensively, and several good reviews are available.[18,19] The chief contributors are listed in Table 1, along with whether they tend to favor the folded or unfolded state.

The most important interaction favoring formation of paired helices is the stacking interaction between the nucleobases. Each of the nucleobases contains a planar six-membered ring with at least partial aromatic character. In the paired helix, the bases are in face-to-face contact with their neighbors, which allows the π-delocalized electrons to interact via attractive van der Waals forces. Each of the nucleobases contains electric dipoles due to the exocyclic amine and carbonyl oxygens. These permanent dipoles induce complementary dipole moments in the π-systems of the neighboring

Table 1 Forces governing nucleic acid hybridization

Interaction	Enthalpy	Entropy
Base stacking	Strongly stabilizing	
Conformational rigidity		Strongly destabilizing
Electrostatic repulsion	Strongly destabilizing	
Watson–Crick H-bonding	Stabilizing	Weakly destabilizing
Solvation	Destabilizing	Stabilizing

bases that results in an attractive force. Additional attraction comes from the London dispersion forces (induced dipole–induced dipole interactions) produced by coupling of electronic fluctuations in the π-systems of neighboring bases. Both of these effects are distinct from the classical hydrophobic effect associated with protein folding; nucleobases, while aromatic, are solvated by water much more readily than hydrophobic amino acid residues. Solvent-associated hydrophobic folding is entropically driven and enthalpically opposed; by contrast, nucleic acid base stacking is strongly enthalpically favored, with only minor changes in entropy.

Two forces strongly oppose hybridization: the electrostatic repulsion between the phosphate backbones of two strands and the loss of conformational freedom upon folding. Nucleic acids are extremely negatively charged. Hybridization into a duplex brings the backbone phosphates close together (1.8–2.6 nm, Section 2.2) and thus must overcome the electrostatic repulsion between the strands. In physiological solutions, dissolved cations partially neutralize the negative charges and allow two strands to get close enough to each other for shorter range attractive interactions to take effect. Folding thermodynamics is therefore sensitive to salt concentration (and species); this is discussed in Section 3.4.2.

Hybridization also results in significant loss of conformational freedom, particularly for the sugar–phosphate backbone, which must contort into a rigid helix from a flexible chain. The nucleobases are also constrained to adopt the syn conformation relative to the sugars. NMR (nuclear magnetic resonance) and X-ray data suggest that constraining the angles of the six backbone and one glycosidic bonds for each of two nucleotides should result in an entropy loss of 22 J $(K \cdot mol)^{-1}$ per base pair added to an existing duplex. This accounts for a major part of the experimentally measured unfavorable entropy of hybridization for nucleic acids (Table 2). Formation of hairpin and internal loops is entropically disfavored for the same reason. When two nucleic acids hybridize into a single duplex, a small amount of entropy is also lost because of the decrease in the translational entropy of the system.

Formation of Watson–Crick hydrogen bonds between the paired bases provides an enthalpic boost for hybridization, as well as a small entropic penalty from fixing the rotational angle of the exocyclic amines. Several methods have been used to dissect the contribution of H-bonding from the overall thermodynamics of hybridization. Substituting guanine with inosine, which is chemically identical but for one fewer H-bond donor group less, in a G·C pair suggests a marginal enthalpy gain of up to 1.5 kcal mol^{-1} per hydrogen bond formed. This is likely an overestimate, as H-bond formation is a cooperative process. Comparing the stability gain from adding a single base pair to a helix to that of stacking the bases separately (with no partner) results in much smaller, sequence-dependent enthalpy gains from H-bonding, on the order of 0.2–0.5 kcal mol^{-1} per bond. In either case, base stacking provides significantly more stabilization than hydrogen bonding, as can be seen from Table 2.

One of the reasons that hydrogen bonding plays a relatively small role in stabilizing hybridization is that water readily solvates both single-stranded and paired nucleic acids. Forming Watson–Crick hydrogen bonds between bases requires breaking H-bonds between the bases and water. This is enthalpically unfavorable, but entropically positive in that the water molecules are freed to circulate in solution. The major and minor grooves of the duplex are also unique environments for hydration. Understanding the role of solvation is one of the main challenges for first-principles calculation of nucleic acid energetics.

Despite this, a remarkably successful semiempirical framework has been developed for predicting the folding thermodynamics of nucleic acid secondary structure. The importance of base stacking for stability requires that any accurate model cannot consider individual bases in isolation but must also take into account the identity and folding state of nearby bases. (Well-known melting temperature formulas based on GC content are only accurate for long molecules with large numbers of every base combination.) Remarkably, it is possible to construct an accurate model of both RNA and DNA secondary structure in which the free energy of pairing for a pair of bases depends only on the identity of the pair and the pairs on either side.[20] Unpaired regions (loops,

Table 2 Nearest-neighbor parameters for base pairing of RNA duplexes, including free energy change at 37 °C. Sequences are presented as 5′-XY-3′ for the top row paired to 3′-ZW-5′ on the bottom. For comparison, the free energy penalty for forming a hairpin, internal, or bulge loop at 37 °C is 4–6 kcal mol^{-1} minus the free energy of forming the closing base pair. Data from Xia et al.[22]

Sequence	GC CG	GG CC	CG GC	GA CU	GU CA	CA GU	CU GA	UA AU	AU UA	AA UU
ΔH° (kcal mol^{-1})	−14.9	−13.4	−10.6	−12.4	−11.4	−10.4	−10.5	−7.7	−9.4	−6.8
ΔS° (cal $(mol \cdot K)^{-1}$)	−36.9	−32.7	−26.7	−32.5	−29.5	−26.9	−27.1	−20.5	−26.7	−19.0
ΔG°_{37} (kcal mol^{-1})	−3.42	−3.26	−2.36	−2.35	−2.24	−2.11	−2.08	−1.33	−1.10	−0.93

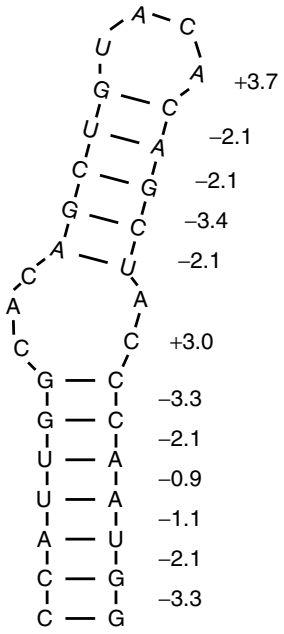

Figure 10 Folding free energy calculation for a 33 nt RNA hairpin at 37 °C. Favorable free energies for forming each nearest-neighbor pair stabilize hairpin formation, whereas unfavorable entropy losses from closing internal and hairpin loops destabilize folding. Thermodynamic parameters from MFOLD.[27]

bulges, and junctions) provide size-dependent additive entropic penalties, as does association of a duplex. It is thus possible to calculate the free energy of any secondary structure using a manageably small set of parameters: 10 nearest-neighbor pairs plus loop and bulge terms. To illustrate this, the pairing parameters for RNA at 37 °C are listed in Table 2 and a sample calculation is illustrated in Figure 10. The importance of considering nearest neighbors is clearly illustrated by comparing the first and third columns: GC/CG and CG/GC dyads have exactly the same base composition, but the difference in stacking order is responsible for a folding free energy difference of more than 30%. The free energy parameters have been extensively measured for RNA and DNA, and partial sets are available for other nucleic acid combinations as well.[21–25] Including common noncanonical pairings such as G·U wobble pairs in RNA increases the accuracy of the model further.[26]

Typically, nearest-neighbor models are capable of predicting the folding free energy of a specified structure with an accuracy of better than 10%. A number of web-based tools have been developed for this task, and these are discussed further in Section 4.

3.2 Tertiary Structure

Nucleic acid tertiary structure is far more diverse than secondary structure, which makes the task of systematizing its energetics correspondingly more difficult. An additional complication is that the additive free energy model, which works well for secondary structure cannot, in principle, be strictly accurate for tertiary interactions. This is due to the fact that tertiary contact between two secondary structures produces a correlation in their energies as a result of both specific interactions and longer range electrostatic forces. The magnitude of these correlations is largely unknown, however, and awaits a campaign of systematic thermodynamic measurement similar to those that produced the nearest-neighbor parameters.

Pseudoknots and kissing loops are the best-characterized tertiary motifs, which is not surprising given that they are produced by pairing of single-stranded regions and are thus somewhat analogous to secondary structures. In both cases, the nearest-neighbor model approximates the folding enthalpy fairly well, particularly the differential effect of mutations within the newly base-paired regions.[28–31] Accurately modeling the entropy change of folding requires knowledge of the conformational space available to the bases that are unpaired before and after formation of the tertiary contacts, data that are difficult to measure. Statistical mechanical models have been developed to account for the entropy change, with several approaches represented in the literature.[32,33] One interesting approach described recently is fitting thermodynamic parameters on the basis of their ability to predict pseudoknot formation in a large set of sequences with known structures.[34] One may be reasonably optimistic that, with continued measurement and modeling, the folding thermodynamics of at least the more common forms of these two tertiary structures will soon be accurately understood.

Another tertiary interaction that has been measured in some detail is coaxial stacking of two helices. This interaction is associated with a large favorable free energy in model systems, and is commonly observed in natural RNA molecules. Thermodynamic parameters for coaxial stacking of helices terminated by Watson–Crick base pairs, as well as with intervening mismatches, have been measured[35,36] and are incorporated into many of the popular calculation tools described in Section 4. Folding thermodynamics of other helix–helix interactions, as well as loop–helix interactions, have not been characterized at the same level of detail and await further investigations.

3.3 Other Structures

Thermodynamic parameters for formation of DNA triple helices (T·A·T and C^+·G·C, Section 2.4) have been measured.[37,38] Only a few sequence patterns form

triple helices, so they are measured individually rather than parameterized as functions of base composition or nearest neighbors. Just as for duplex formation, triplex formation is driven forward by enthalpy and opposed by loss of conformational entropy. At neutral or acidic pH, protonated C^+·G·C triplets are substantially more stable than T·A·T triplets. Interestingly, the most stable triple helix sequence is one that alternates between the two. At higher pH values, C^+·G·C triple helices are unstable, likely because of deprotonation.

Folding thermodynamics for G-quadruplex formation have also been measured for a number of sequences.[18,39] At 25 °C, the energetics of quadruplex formation are comparable to those for base-pair formation, with a free energy change of -2 to -3 kcal mol^{-1} per quadruplex in a buffer containing potassium cations. Interestingly, G-quadruplex stability is very sensitive to cation size: they are much less stable when K^+ is replaced by Na^+ or, particularly, Li^+. The sequence of the loops in intramolecular quadruplexes also plays a large role in determining stability.[40]

3.4 Environmental Effects

3.4.1 Temperature

As discussed above, the entropy change for nucleic acid folding is nearly always negative. As a result, the stability of nucleic acid structures decreases with increasing temperature. Each sequence and structure has a temperature above which folding is disfavored; this temperature is called the *melting temperature*, T_m (Figure 12). Measurement of T_m and the shape of the melting transition vs temperature plot forms the basis for many of the techniques used to measure folding thermodynamics. These are discussed in Section 5.2. One interesting consequence of the negative entropy of nucleic acid folding is the observation that bacteria and archaeans that live at very high temperatures (extreme thermophiles) have evolved transfer RNAs with higher GC content and larger T_m values than those of other organisms.[41] Similarly, psychrophiles (organisms adapted to very cold temperatures) produce more of the helicase proteins needed to unwind DNA for replication and transcription.[42]

3.4.2 Salt Concentration

Interaction of cations with nucleic acids can be divided into two classes: specific binding, in which a cation is localized at a particular site in the folded structure, and nonspecific binding, in which the cations are free to move around the nucleic acid and exchange with others in the bulk solution. In the context of folding thermodynamics, specific binding is important mainly for stabilizing tertiary structures, which can have particularly high negative charge densities. Nonspecific interactions with cations are important for all aspects of nucleic acid energetics, and these are discussed first.

Each nucleotide in a nucleic acid chain has one negatively charged phosphate group. In B-form duplex DNA, phosphate groups are separated by only 0.3 nm along the helical axis, which is an enormous charge density. Cations from the surrounding solution are drawn to the resulting potential well and arrange themselves to neutralize the electric field. The physics of this process is described by Manning condensation theory,[18] but the relevant consequence for the thermodynamics of folding is that a number of cations are confined near the nucleic acid and unable to circulate freely in solution. This reduces the entropy of the system by an amount

$$\Delta S = R \ln \frac{c_{loc}}{c_{sol}} \tag{1}$$

per mole of confined cations, where R is the gas constant, c_{sol} is the bulk concentration of cations, and c_{loc} is the average concentration of the confined cations.

The number of cations confined in this way is determined by the charge density of the nucleic acid structure; when the structure changes, so will the number of confined cations. Folded structures have higher charge density than unfolded ones, so cations must be extracted from solution during folding. From Equation (1), the cost of doing so decreases inversely with the logarithm of the bulk cation concentration; thus, addition of salt stabilizes nucleic acid folding. Numerous experiments have found that folding free energies do exhibit the predicted logarithmic behavior, though ab initio calculation of the number of bound cations is challenging even for simple structures.

Another prediction of condensation theory is that the increase in stability of the folded state should saturate once the bulk cation concentration exceeds a particular value. For sodium, saturation effects start becoming apparent for $[Na^+] > 0.1$ M. Above 1 M, further increases in sodium concentration actually reduce the stability of folding. Multivalent cations should be much more effective at neutralizing the phosphate backbone charge and therefore saturate at a lower concentration. These predictions are borne out by experiment, with saturation occurring at about 0.01 M Mg^{2+}.[18]

In addition to the diffuse binding described above, cations can also bind to specific sites in a nucleic acid structure. Binding affinity of the cations varies with background salt concentration in the same logarithmic manner as folding stability, as other cations are displaced by the specific binding event in order to maintain a

constant overall charge. The effect of specific binding on the stability of the structure, on the other hand, is more complex, and depends on a variety of factors, including the local charge density of the structure (larger charge densities require more cation charge to stabilize), the degree to which the water molecules hydrating the cation are displaced (displacement requires a tight fit, but produces an entropic benefit from the freed water), and long-range correlations between cations that are very difficult to model.

Site-specific binding appears to be minimal for most secondary structures. This can be determined either from atomic-scale methods such as NMR and crystallography, or by measuring the effect of different cations on stabilization. In condensation theory, which describes nonspecific binding, the only relevant quality of a cation is its charge. All monovalent cations should therefore be equally effective at stabilizing folding. As discussed in Section 3.3, this is not the case for G-quadruplexes, and specific binding of potassium-sized cations in the center of each quadruplex square is supported by a variety of experiments. In contrast, recent measurements[43,44] found that RNA secondary structure folding shows only a slight differential stability between sodium and potassium. An RNA kissing loop, by contrast, exhibits substantial differential stabilization, indicating that site-specific binding is likely to occur in these structures. Nearly all tertiary structures that have been measured at the atomic scale show specific binding of multivalent cations. A number of common motifs for specific binding have been identified, including G·A and G·U mismatches, structured junctions, and tight bends. Multivalent cations are particularly likely to bind in the latter two motifs, as local negative charge densities can be very high in these areas and are more effectively neutralized by cations with correspondingly high positive charges.

No general theory describing salt effects on nucleic acid energetic currently exists. The theoretical framework for interpreting measurements of salt effects on stability is well developed[45] and computational techniques for predicting thermodynamic effects on folding are advancing.[44,46] The published nearest neighbor parameters for DNA secondary structure thermodynamics include approximate salt dependence terms, but no similar parameters have been published for RNA secondary structure. As with tertiary structure thermodynamics, a lack of systematic experimental data has been a problem.

3.4.3 *pH*

Secondary structure stability is relatively insensitive to pH changes from 5 to 9.[18] At higher and lower values, stability decreases due to changes in protonation of the nucleobases (Section 2.1). Deprotonation of guanine, thymine, and uracil occurs at high pH, destabilizing pairing both through loss of hydrogen bonding and increased electrostatic repulsion. At low pH, the nitrogen atoms on the Watson–Crick faces of the bases are protonated, disfavoring pairing.

Tertiary interactions can be more pH sensitive, since their diverse conformations can produce local shifts in pK_a values sufficient to induce protonation or deprotonation at physiological pH values. pH-dependent conformational shifts have been observed for both bacterial 5S ribosomal RNA[47] and kissing loops at the 5′ end of HIV's RNA genome.[48] Triple-helix and *i*-motif formation in DNA is also pH sensitive, as discussed in Sections 2.4 and 3.3.

3.4 *Solvent Effects*

At moderate and high ionic strengths, cosolvents tend to destabilize folded structures, either by decreasing the activity of water, which is released upon folding (Section 3.1), or through specific interaction with unpaired strands. The latter case is exemplified by urea, which forms hydrogen bonds with the bases that compete with Watson–Crick pairing.[49] At low ionic strengths, the osmolyte effect is reversed: cosolvents (which are typically less polar than water) decrease the dielectric constant of the solution, which strengthens the hydrogen bonds of base pairing.

3.5 *Ligands and Proteins*

Many protein molecules bind nucleic acids. In some cases, secondary structural elements such as helices or loops are bound in isolation, with little change to the structure of the rest of the nucleic acid. In other cases, RNA molecules and proteins form intimately connected complexes with structures substantially different from that adopted by the components in isolation. One obvious example is the ribosome. The important subject of protein–nucleic acid binding is explored in an accompanying article (*see* **Protein–Oligonucleotide Interactions**). Section 5.1 describes methods that can be useful for determining what parts of a nucleic acid are bound by proteins.

Riboswitches are RNA molecules that regulate gene expression by changing conformation in response to the presence or absence of small molecules.[50] They are an important part of the genetic regulation networks for many bacteria, and their sequences have been found in eukaryotes as well. The energetics of binding and conformational change upon ligand binding are beginning to be investigated,[51] and will likely lead to interesting results with riboswitches, as allosteric control is a function previously thought exclusive to proteins.

4 STRUCTURE PREDICTION AND DESIGN

Anfinsen's influential ''thermodynamic hypothesis'' holds that the active structure of a biomolecule is the one that is most stable in its native environment.[52] As described above, nucleic acid energetics are fairly well understood, at least at the secondary structure level. This suggests the possibility of predicting the functional structure of nucleic acids from their sequence alone. This section discusses the computational tools that have been developed to address this challenge, as well as the inverse problem of designing sequences that fold into a desired structure.

4.1 Predicting Structure from Sequence

The hierarchical nature of nucleic acid folding, together with the success of the nearest-neighbor model at predicting secondary structure thermodynamics, suggests a strategy for predicting the equilibrium structure of a sequence.[16,53] First, the most likely (i.e. lowest energy) secondary structure for the sequence can be computed using the nearest-neighbor model. Likely tertiary structures can then be computed on the basis of interactions of the secondary structural elements. This approach is computationally tractable, whereas an unguided exploration of the conformational space for even small nucleic acids is far beyond the capabilities of today's computers and will likely remain so for the forseeable future.

Even with the nearest-neighbor model providing an efficient and reasonably accurate method for computing the energy of any secondary structure, determining the minimum-energy secondary structure for a given sequence is still a nontrivial problem. The number of possible structures grows exponentially with the length of the sequence, making exhaustive enumeration unfeasible for all but the smallest sequences. Instead, dynamic programming algorithms are used to recursively search structure space. This allows the minimum free energy (MFE) structure of a sequence to be computed in a time that grows only cubically with sequence length; a fast computer can compute the MFE structure for a several hundred nucleotide sequence in a few minutes or less. A number of software packages have been developed for this purpose, of which the best-known are MFOLD[27] (mfold.bioinfo.rpi.edu) and the Vienna package[54] (rna.tbi.univie.ac.at). Both of these have web interfaces and are capable of calculating MFE structures for single RNA or DNA sequences at arbitrary temperatures. The predicted structures agree well with experimental data for many small sequences and the MFE technique is used widely to understand the behavior of nucleic acids.

MFE predictions of secondary structure are not perfect, for several reasons. The first is that the nearest-neighbor parameters are not known with infinite precision. Typical uncertainties are a few percent for the free energy of the base stack parameters and small loops near 37 °C. The uncertainties for other motifs, as well as the enthalpy and entropy parameters needed for computation at arbitrary temperatures, can be substantially larger. A second weakness of the MFE technique comes from weaknesses in the nearest-neighbor model itself. In particular, the thermodynamics of noncanonical base pairs and internal loops display context dependence that cannot be accounted for within the nearest-neighbor model. For larger sequences and more complex structures, these inaccuracies compound to an extent that determination of the MFE structure is not accurate. One method for dealing with this is to compute not only the lowest energy structure but also a collection of ''suboptimal'' structures with similar free energies. Efficient algorithms for suboptimal structure prediction have been developed and have been shown to increase the accuracy of structure prediction.

A more powerful approach than computing a selection of structures is to compute the thermodynamic partition function for a sequence using the secondary structure model energy model. To the extent that the model is accurate, the partition function contains all available information about the sequence's folding, including the pairing probabilities of every base to every other base. At first glance, it would seem that computation of the partition function would require enumeration of all possible structures, but a dynamic programming algorithm similar to that used for MFE structure prediction can also compute the partition function in polynomial time. Both the MFOLD (as Dinamelt: dinamelt.bioinfo.rpi.edu) and Vienna package sites have been extended to include partition function calculations. The Nupack software package (nupack.org) computes the secondary structure partition function for both individual and multiple interacting nucleic acid strands at arbitrary concentrations[55] and also includes the capability of predicting structures containing certain types of pseudoknots. Experimental validation of the partition function approach is incomplete, but it seems likely that this will become the most widely used method for predicting nucleic acid secondary structure folding in the future.

The methods discussed above are useful for predicting the folding tendencies of molecules about which nothing is known other than the sequence. If information is available from other sources, such as experimentally determined pairings of some bases or homology to other sequences of known structure or function, the problem of structure prediction becomes much easier. One powerful method is identifying subsequences whose pairing is

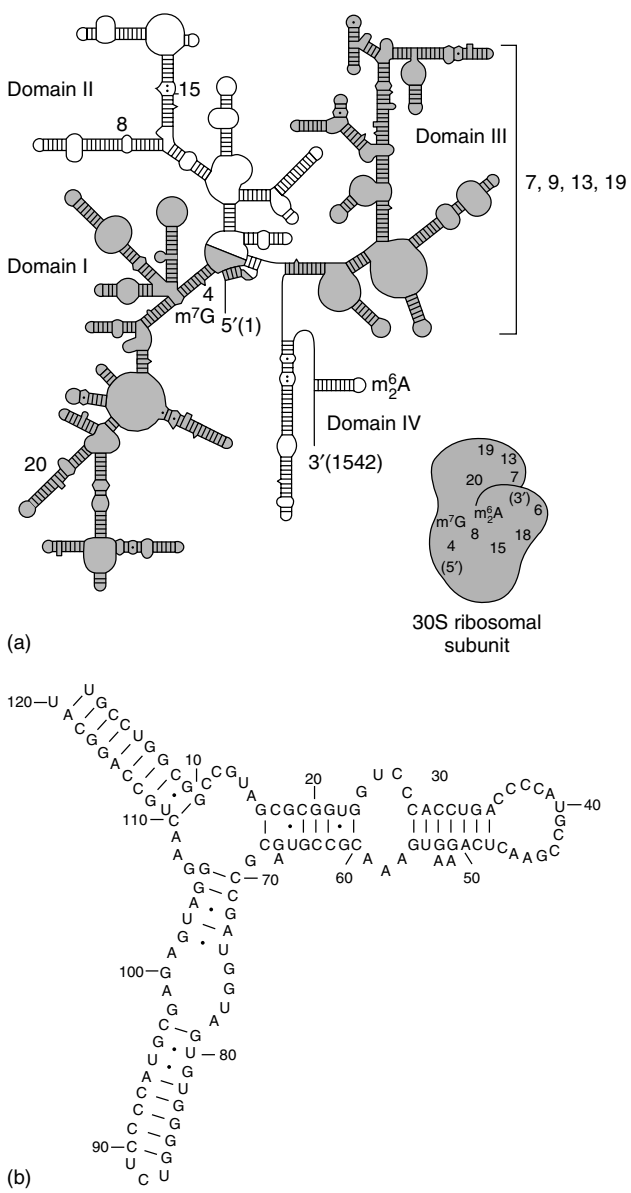

Figure 11 (a) Predicted secondary structure of the 1542 nucleotide *Escherichia coli* 16S rRNA from comparative sequence analysis. Agreement with X-ray crystal structures is excellent. The placement of certain features with respect to specific ribosomal proteins and the entire 30S ribosomal subunit are indicated. (b) Sequence and predicted secondary structure of *E. coli* 5S rRNA. (Reproduced from Ref. 57. © John Wiley & Sons, Inc, 1990.)

conserved across multiple species, either by strict sequence conservation or compensating mutations (G·C to A·U, for instance). When the number of homologous sequences is large, comparative methods can be very powerful, predicting 97% of ribosomal RNA base pairings (Figure 11) correctly.[56] As with partition function calculations, most of the popular software packages now include comparative analysis features as well.

Coarse-grained molecular dynamics simulations, which are becoming widely used in the protein folding field, are starting to be applied to RNA folding as well. In principle, this approach offers the potential to predict folded structures that are not described by the conventional secondary structure model and account for tertiary interactions whose thermodynamics have not been systematized. At present, only short molecules (less than 50 nt or so) can be treated in this way owing to the scaling problem discussed above for secondary structure prediction. Computer power is also growing exponentially, however, and molecular dynamics–based methods will likely grow in importance over the near future.

4.2 Designing Sequence for Structure

The inverse problem of structure prediction is to design a sequence that folds into a desired structure. Recent years have seen an explosion of interest in nucleic acid–based nanotechnology, in which the programmability of nucleic acid sequences is exploited to construct sensors, therapeutics, or nanoscale machines.[58–62] At present, sequence design software has been integrated into the Vienna and Nupack software packages, with the Vienna package offering MFE and partition function–based design of single sequences and Nupack capable of designing sequences for multiple interacting molecules using the partition function approach. Both programs use a similar methodology: a sequence is generated at random, the most likely structure is predicted, and mutations are made until the most likely structure is as close as possible to the desired one. As interest in nucleic acid applications grows, sequence design will likely become as standard a technique as structure prediction is today.

5 EXPERIMENTAL METHODS

We close this article with a brief discussion of experimental techniques that are commonly used to determine the structure and folding thermodynamics of nucleic acids. Many of these techniques are discussed in greater detail in other articles, and links to these are provided.

5.1 Structure Determination

Nucleic acid structure can be measured at several levels of detail. Atomic-scale structures are typically measured using either X-ray crystallography (*see* **X-ray Structures of Nucleic Acids**) or NMR techniques (*see* **Nuclear Magnetic Resonance and Nucleic Acid Structures**). NMR studies are particularly useful in that they can determine not only the structure of the nucleic acid but details of its interaction with cations in solution.

In many cases, however, atomic coordinates are not necessary to understand the functional structure of a nucleic acid. Instead, what is needed is knowledge of which bases are paired and which are exposed to solution. Owing to the stability of nucleic acid hybridization, only the latter are typically available for binding to other molecules. Pairing data define the secondary structure of the molecule, and give significant information about many tertiary interactions as well (Section 2.3). One way to determine the pairing state of the bases of a molecule is to introduce short oligonucleotide probes complementary to various parts of the sequence; the probes will only bind to bases that are not already paired. This technique is called *footprinting*, and is described in (**Sequencing and Fingerprinting DNA by Hybridization with Oligonucleotide Probes**). Chemical probes can also be used: dimethyl sulfate (DMS) reacts with the Watson–Crick faces of exposed bases and hydroxyl radicals preferentially cleave the sugar–phosphate backbone in flexible (i.e. unpaired) regions,[63] as do certain nuclease enzymes. The pattern of chemical or enzymatic cleavage can be examined by gel electrophoresis and analyzed to reveal the pairing state of the nucleic acid structure. Sites of protein binding can be revealed using the same techniques. Comparative sequence analysis also reveals structural information, as discussed in Section 4.

5.2 Thermodynamic Characterization

The vast majority of the thermodynamic data discussed in this article was obtained by melting experiments, in which the temperature of a solution containing a folded nucleic acid is gradually increased until entropy succeeds in disordering the molecule (Section 3.4.1). This type of experiment is illustrated in Figure 12. The molecule is assumed to inhabit only two conformational states, folded and unfolded (multiple states can be treated as well, provided the melting temperatures are sufficiently different that they can be determined independently). In this case, at constant pressure, the standard Gibbs free energy ΔG^0 of folding is related to the standard enthalpy and entropy changes ΔH^0 and ΔS^0 and the equilibrium constant K by the well-known expression

$$\Delta G^0 = \Delta H^0 - T\Delta S^0 = RT \ln K \qquad (2)$$

The temperature where the concentrations of folded and unfolded species are equal is defined as the melting temperature T_m. For a unimolecular reaction, $T_m = \Delta H^0/\Delta S^0$. Differentiating Equation (2) with respect to temperature gives the van't Hoff relation:

$$\frac{\partial \ln K}{\partial T} = \frac{\Delta H^0}{RT^2} \qquad (3)$$

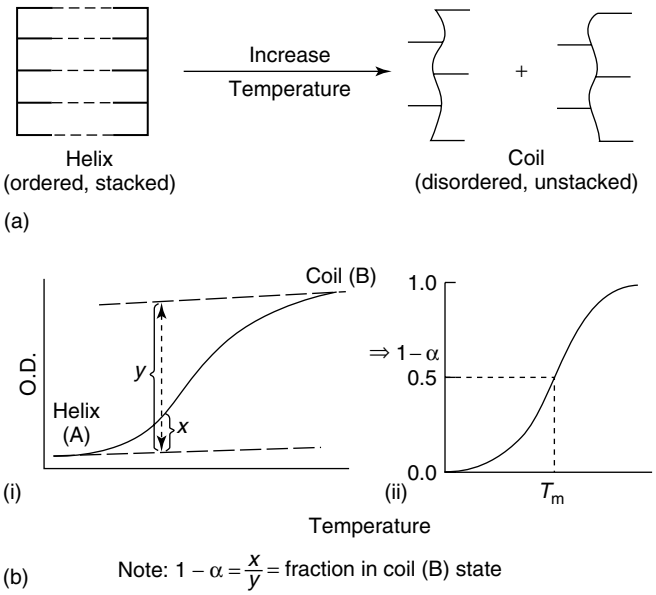

Figure 12 (a) Schematic representation of melting a folded nucleic acid. Increasing temperature favors the less-organized unfolded state. (b) UV melting experiments measure absorbance (represented here as optical depth) vs temperature. Fitting the extinction coefficients of the folded and unfolded states (i) allows for conversion to population fraction vs temperature (ii). (Reproduced from Ref. 64. © Springer-Verlag, 1986.)

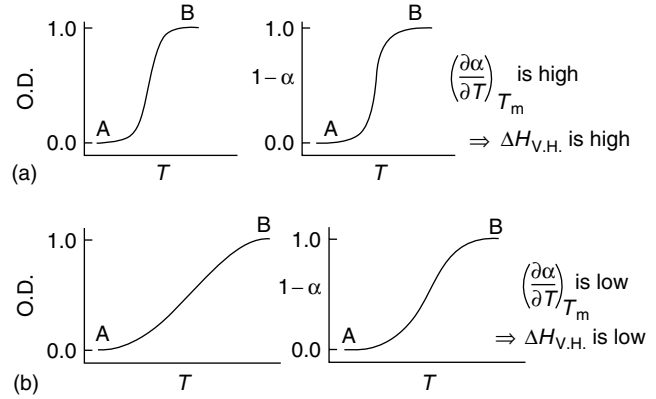

Figure 13 Determining enthalpy changes from melting curves. (a) Transition width is narrow, folding enthalpy is high. (b) Transition is gradual, folding enthalpy is low. (Reproduced from Ref. 64. © Springer-Verlag, 1986.)

Reactions with large enthalpy changes (e.g. base pairing) thus display steeper transitions around T_m, than those with small enthalpy changes (e.g. some tertiary interactions). This is illustrated in Figure 13.

In principle, any method that can determine the relative concentration of folded and unfolded species as a function of temperature can be used to measure folding thermodynamics. In practice, the dominant technique is to measure the molecule's absorbance in the ultraviolet vs temperature. UV (ultraviolet) melting relies on the

observation that nucleic acids absorb significantly more light at 260 nm when single-stranded than when base-paired. This phenomenon is called *hypochroism* and is related to the electronic interactions that stabilize base stacking (Section 3.1). As illustrated in Figure 12, absorbance data can be fit using four parameters: ΔH^0, ΔS^0 (or K), and the extinction coefficients of the folded and unfolded species. The latter two are obtained by fitting the absorbance baselines far from the transition, and all four parameters are assumed not to vary with temperature (i.e. constant specific heat, $\Delta C_p = 0$). This imposes several limitations on the UV melting technique in addition to the two-state condition: the folding reaction in question must take place near the temperature at which the thermodynamic parameters will be used (to minimize errors from ΔC_p) and well away from the freezing or boiling point of the solution (in order to accurately fit the extinction coefficients). Despite these limitations, UV melting is a workhorse technique in nucleic acid thermodynamics. A more detailed discussion of experimental considerations, including molecularity effects, is available in the review literature.[65]

Nucleic acid thermodynamic parameters can also be measured calorimetrically. In the most common method, differential scanning calorimetry (DSC, *see* **Differential Scanning Calorimetry and Differential Thermal Analysis**), the temperature of two matched cells is slowly ramped over a broad range using electrical heating. One of the cells contains the nucleic acid of interest and the other only a buffer solution. The differential amount of power needed to ramp the temperatures of the cells provides a measure of ΔC_p, the heat absorbed during the unfolding reaction. If other reactions such as hydration of the buffer solutes are assumed to be constant over the transition region (t_1, t_2), enthalpy and entropy changes can be obtained by integration of the differential heat capacity:

$$\Delta H^0 = \int_{t_1}^{t_2} \Delta C_p \, \mathrm{d}T \tag{4}$$

$$\Delta S^0 = \int_{t_1}^{t_2} \frac{\Delta C_p}{T} \, \mathrm{d}T \tag{5}$$

In contrast to the melting technique, DSC does not require that the reaction need be two-state, which can be a significant advantage for larger structures. One disadvantage is that larger amounts of material are typically required in order to produce a measurable signal. Both techniques will likely continue to be widely used as the quest to more fully understand nucleic acid energetics continues.

ABBREVIATIONS AND ACRONYMS

bp	Base Pair
DSC	Differential Scanning Calorimetry
DMS	Dimethyl Sulfate
MFE	Minimum Free Energy
NMR	Nuclear Magnetic Resonance
PNA	Peptide Nucleic Acid

RELATED ARTICLES

Electronic Absorption and Luminescence
Circular Dichroism and Linear Dichroism

Nuclear Magnetic Resonance and Electron Spin Resonance Spectroscopy
Multidimensional Nuclear Magnetic Resonance of Biomolecules

Nucleic Acids Structure and Mapping
Aptamers • DNA Molecules, Properties and Detection of Single • DNA Probes • DNA Structures of Biological Relevance, Studies of Unusual Sequences • Fluorescence In Situ Hybridization • Mass Spectrometry of Nucleic Acids • Nuclear Magnetic Resonance and Nucleic Acid Structures • X-ray Structures of Nucleic Acids • Nucleic Acids Structure and Mapping: Introduction • PNA and Its Applications • Polymerase Chain Reaction and Other Amplification Systems • RNA Tertiary Structure • Sequencing and Fingerprinting DNA by Hybridization with Oligonucleotide Probes • Structural Analysis of Ribozymes

Peptides and Proteins
Protein–Oligonucleotide Interactions

Thermal Analysis
Differential Scanning Calorimetry and Differential Thermal Analysis

REFERENCES

1. W. Saenger, *Principles of Nucleic Acid Structure*, Springer, Berlin, 10, 14, 1984.
2. D.H. Kim, J.J. Rossi, 'Strategies for Silencing Human Disease Using RNA Interference', *Nat. Rev. Genet.*, **8**(3), 173–184 (2007).
3. N. Dias, C.A. Stein, 'Antisense Oligonucleotides: Basic Concepts and Mechanisms', *Mol. Cancer Ther.*, **1**(5), 347–355 (2002).
4. M.A. Behlke, 'Chemical Modification of siRNAs for In vivo Use', *Oligonucleotides*, **18**(4), 305–319 (2008).

5. S. Sasaki, F. Nagatsugi, 'Application of Unnatural Oligonucleotides to Chemical Modification of Gene Expression', *Curr. Opin. Chem. Biol.*, **10**(6), 615–621 (2006).
6. W. Saenger, *Principles of Nucleic Acid Structure*, Springer, Berlin, 257, 262, 286, 1984.
7. J.H. Cate, A.R. Gooding, E. Podell, K. Zhou, B.L. Golden, A.A. Szewczak, C.E. Kundrot, T.R. Cech, J.A. Doudna, 'Crystal Structure of a Group I Ribozyme Domain: Principles of RNA Packing', *Science*, **273**, 1697 (1996).
8. J.R. Wyatt, J.D. Puglisi, I. Tinoco, Jr, *Bioessays*, **11**(4), 104 (1989).
9. J.S. Bois, S. Venkataraman, H.M. Choi, A.J. Spakowitz, Z.G. Wang, N.A. Pierce, 'Topological Constraints in Nucleic Acid Hybridization Kinetics', *Nucleic Acids Res.*, **33**(13), 4090–4095 (2005).
10. D.W. Staple, S.E. Butcher, 'Pseudoknots: RNA Structures with Diverse Functions', *PLoS Biol.*, **3**(6), 956–959 (2005).
11. L.X. Shen, I. Tinoco, Jr, 'The Structure of an RNA Pseudoknot that Causes Efficient Frameshifting in Mouse Mammary Tumor Virus', *J. Mol. Biol.*, **247**, 974 (1995).
12. S.M. Mirkin, M.D. Frank-Kamenetskii, 'Triplex DNA STRUCTURES', *Annu. Rev. Biophys. Biomol. Struct.*, **23**, 543 (1994), www.AnnualReviews.org.
13. H. Deng, W.H. Branulin, *Biopolymers*, **35**, 680 (1995).
14. J.L. Huppert, 'Four-stranded Nucleic Acids: Structure, Function and Targeting of G-quadruplexes', *Chem. Soc. Rev.*, **3**(7), 1375–1384 (2008).
15. R. Jin, K.J. Breslauer, R.A. Jones, B.L. Gaffney, 'Tetraplex Formation of a Guanine-containing Nonameric DNA Fragment', *Science*, **250**, 544 (1990).
16. I. Tinoco, C. Bustamante, 'How RNA Folds', *J. Mol. Biol.*, **293**(2), 271–281 (1999).
17. P. Rangan, B. Masquida, E. Westhof, S.A. Woodson, 'Assembly of Core Helices and Rapid Tertiary Folding of a Small Bacterial Group I Ribozyme', *Proc. Natl. Acad. Sci. U.S.A.*, **100**(4), 1574–1579 (2003).
18. V.A. Bloomfield, D.C. Crothers, I. Tinoco, *Nucleic Acids: Structures, Properties and Functions*, University Science Books, Sausalito, CA, 2000.
19. E.T. Kool, 'Hydrogen Bonding, Base Stacking, and Steric Effects in DNA Replication', *Annu. Rev. Biophys. Biomol. Struct.*, **30**, 1–22 (2001).
20. I. Tinoco, O. Uhlenbeck, M.D. Levine, 'Estimation of Secondary Structure in Ribonucleic Acids', *Nature*, **230**(5293), 362–367 (1971).
21. J. SantaLucia, 'A Unified View of Polymer, Dumbbell, and Oligonucleotide DNA Nearest-neighbor Thermodynamics', *Proc. Natl. Acad. Sci. U.S.A.*, **95**(4), 1460–1465 (1998).
22. T.B. Xia, J. SantaLucia, M.E. Burkard, R. Kierzek, S.J. Schroeder, X.Q. Jiao, C. Cox, D.H. Turner, 'Thermodynamic Parameters for an Expanded Nearest-neighbor Model for Formation of RNA Duplexes with Watson-Crick Base Pairs', *Biochemistry*, **37**(42), 14719–14735 (1998).
23. D.H. Turner, D.H. Mathews, 'NNDB: The Nearest Neighbor Parameter Database for Predicting Stability of Nucleic Acid Secondary Structure', *Nucleic Acids Res.*, **38**(Database), D280–D282 (2010).
24. N. Sugimoto, S. Nakano, M. Katoh, A. Matsumura, H. Nakamuta, T. Ohmichi, M. Yoneyama, M. Sasaki, 'Thermodynamic Parameters to Predict Stability of RNA/DNA Hybrid Duplexes', *Biochemistry*, **34**(35), 11211–11216 (1995).
25. E. Kierzek, D.H. Mathews, A. Ciesielska, D.H. Turner, R. Kierzek, 'Nearest Neighbor Parameters for Watson-Crick Complementary Heteroduplexes Formed Between 2′-O-methyl RNA and RNA Oligonucleotides', *Nucleic Acids Res.*, **34**(13), 3609–3614 (2006).
26. D.H. Mathews, J. Sabina, M. Zuker, D.H. Turner, 'Expanded Sequence Dependence of Thermodynamic Parameters Improves Prediction of RNA Secondary Structure', *J. Mol. Biol.*, **288**(5), 911–940 (1999).
27. M. Zuker, 'Mfold Web Server for Nucleic Acid Folding and Hybridization Prediction', *Nucleic Acids Res.*, **31**(13), 3406–3415 (2003).
28. S. Cao, S.-J. Chen, 'Predicting RNA Pseudoknot Folding Thermodynamics', *Nucleic Acids Res.*, **34**(9), 2634–2652 (2006).
29. R.S. Gregorian, D.M. Crothers, 'Determinants of RNA Hairpin Loop-loop Complex Stability', *J. Mol. Biol.*, **248**(5), 968–984 (1995).
30. A. Weixlbaumer, A. Werner, C. Flamm, E. Westhof, R. Schroeder, 'Determination of Thermodynamic Parameters for HIV DIS Type Loop-loop Kissing Complexes', *Nucleic Acids Res.*, **32**(17), 5126–5133 (2004).
31. C. Lorenz, N. Piganeau, R. Schroeder, 'Stabilities of HIV-1 DIS Type RNA Loop-loop Interactions In vitro and In vivo', *Nucleic Acids Res.*, **34**(1), 334–342 (2006).
32. S. Cao, S.-J. Chen, 'Predicting Structures and Stabilities for H-type Pseudoknots with Interhelix Loops', *RNA*, **15**(4), 696–706 (2009).
33. J. Sperschneider, A. Datta, 'Dotknot: Pseudoknot Prediction Using the Probability Dot Plot Under a Refined Energy Model', *Nucleic Acids Res.*, **38**(7), e103–e103 (2010).
34. M.S. Andronescu, C. Pop, A.E. Condon, 'Improved Free Energy Parameters for RNA Pseudoknotted Secondary Structure Prediction', *RNA*, **16**(1), 26–42 (2010).
35. A.E. Walter, D.H. Turner, 'Sequence Dependence of Stability for Coaxial Stacking of RNA Helixes with

Watson-Crick Base Paired Interfaces', *Biochemistry*, **33**(42), 12715–12719 (1994).

36. J. Kim, A.E. Walter, D.H. Turner, 'Thermodynamics of Coaxially Stacked Helixes with GA and CC Mismatches', *Biochemistry*, **35**(43), 13753–13761 (1996).
37. R.W. Roberts, D.M. Crothers, 'Prediction of the Stability of DNA Triplexes', *Proc. Natl. Acad. Sci. U.S.A.*, **93**(9), 4320–4325 (1996).
38. P.L. James, T. Brown, K.R. Fox, 'Thermodynamic and Kinetic Stability of Intermolecular Triple Helices Containing Different Proportions of C·GC and T·AT Triplets', *Nucleic Acids Res.*, **31**(19), 5598–5606 (2003).
39. C.C. Hardin, A.G. Perry, K. White, 'Thermodynamic and Kinetic Characterization of the Dissociation and Assembly of Quadruplex Nucleic Acids', *Biopolymers*, **56**(3), 147–194 (2000).
40. A. Guedin, P. Alberti, J.-L. Mergny, 'Stability of Intramolecular Quadruplexes: Sequence Effects in the Central Loop', *Nucleic Acids Res.*, **37**(16), 5559–5567 (2009).
41. A. Dutta, K. Chaudhuri, 'Analysis of tRNA Composition and Folding in Psychrophilic, Mesophilic and Thermophilic Genomes: Indications for Thermal Adaptation', *FEMS Microbiol. Lett.*, **305**(2), 100–108 (2010).
42. S. D'Amico, T. Collins, J.C. Marx, G. Feller, C. Gerday, 'Psychrophilic Microorganisms: Challenges for Life', *EMBO Rep.*, **7**(4), 385–389 (2006).
43. J. Vieregg, W. Cheng, C. Bustamante, I. Tinoco, 'Measurement of the Effect of Monovalent Cations on RNA Hairpin Stability', *J. Am. Chem. Soc.*, **129**(48), 14966–14973 (2007).
44. A.A. Chen, M. Marucho, N.A. Baker, R.V. Pappu, 'Simulations of RNA Interactions with Monovalent Ions', *Methods Enzymol.*, **469**, 411–432 (2009).
45. D. Leipply, D. Lambert, D.E. Draper, 'Ion-RNA Interactions: Thermodynamic Analysis of the Effects of Mono- and Divalent Ions on RNA Conformational Equilibria', *Methods Enzymol.*, **469**, 433–463 (2009).
46. S.-J. Chen, 'RNA Folding: Conformational Statistics, Folding Kinetics, and Ion Electrostatics', *Ann. Rev. Biophys.*, **37**, 197–214 (2008).
47. T.H. Kao, D.M. Crothers, 'A Proton-coupled Conformational Switch of Escherichia-coli 5 S Ribosomal-RNA', *Proc. Natl. Acad. Sci. U.S.A.*, **77**(6), 3360–3364 (1980).
48. M.R. Mihailescu, J.P. Marino, 'A Proton-coupled Dynamic Conformational Switch in the HIV-1 Dimerization Initiation Site Kissing Complex', *Proc. Natl. Acad. Sci. U.S.A.*, **101**(5), 1189–1194 (2004).
49. U.D. Priyakumar, C. Hyeon, D. Thirumalai, A.D. Mackerell, 'Urea Destabilizes RNA by Forming Stacking Interactions and Multiple Hydrogen Bonds with Nucleic Acid Bases', *J. Am. Chem. Soc.*, **131**(49), 17759–17761 (2009).
50. A. Roth, R.R. Breaker, 'The Structural and Functional Diversity of Metabolite-binding Riboswitches', *Annu. Rev. Biochem.*, **78**, 305–334 (2009).
51. I. Tinoco, D. Collin, P.T.X. Li, 'The Effect of Force on Thermodynamics and Kinetics: Unfolding Single RNA Molecules', *Biochem. Soc. Trans.*, **32**, 757–760 (2004).
52. C.B. Anfinsen, 'Principles that Govern the Folding of Protein Chains', *Science*, **181**(96), 223–230 (1973).
53. D.H. Mathews, 'Revolutions in RNA Secondary Structure Prediction', *J. Mol. Biol.*, **359**(3), 526–532 (2006).
54. A.R. Gruber, R. Lorenz, S.H. Bernhart, R. Neubock, I.L. Hofacker, 'The Vienna RNA Websuite', *Nucleic Acids Res.*, **36**(Web Server issue), W70–W74 (2008).
55. R.M. Dirks, J.S. Bois, J.M. Schaeffer, E. Winfree, N.A. Pierce, 'Thermodynamic Analysis of Interacting Nucleic Acid Strands', *SIAM Rev*, **49**(1), 65–88 (2007).
56. R.R. Gutell, J.C. Lee, J.J. Cannone, 'The Accuracy of Ribosomal RNA Comparative Structure Models', *Curr. Opin. Struct. Biol.*, **12**(3), 301–310 (2002).
57. D. Voet, J.G. Voet, *Biochemistry*, John Wiley & Sons, New York, 916, 1990.
58. B. Yurke, A.P. Mills, 'Using DNA to Power Nanostructures', *Genet. Program. Evol. Mach.*, **4**(2), 111–122 (2003).
59. Y. He, T. Ye, M. Su, C. Zhang, A.E. Ribbe, W. Jiang, C. Mao, 'Hierarchical Self-assembly of DNA into Symmetric Supramolecular Polyhedra', *Nature*, **452**(7184), 198–201 (2008).
60. P. Yin, H.M.T. Choi, C.R. Calvert, N.A. Pierce, 'Programming Biomolecular Self-assembly Pathways', *Nature*, **451**(7176), 318–322 (2008).
61. T. Omabegho, R. Sha, N.C. Seeman, 'A Bipedal DNA Brownian Motor with Coordinated Legs', *Science*, **324**(5923), 67–71 (2009).
62. S.M. Douglas, H. Dietz, T. Liedl, B. Högberg, F. Graf, W.M. Shih, 'Self-assembly of DNA into Nanoscale Three-dimensional Shapes', *Nature*, **459**(7245), 414–418 (2009).
63. Z. Xu, G.M. Culver, 'Chemical Probing of RNA and RNA/Protein Complexes', *Methods Enzymol.*, **468**, 147–165 (2009).
64. K.J. Breslauer, in *Thermodynamic Data for Biochemistry and Biotechnology*, ed. H.-J. Hinz, Springer, New York, 408, 1986.
65. S.J. Schroeder, D.H. Turner, 'Optical Melting Measurements of Nucleic Acid Thermodynamics', *Methods Enzymol.*, **468**, 371–387 (2009).