A Structural Basis for RNA–Ligand Interactions

Christine S. Chow* and Felicia M. Bogdan

Department of Chemistry, Wayne State University, Detroit, Michigan 48202

Received December 10, 1996 (Revised Manuscript Received April 8, 1997)

Contents

I. Introduction 1489
II. General Considerations 1490
A. RNA Structure 1490
B. Types of Noncovalent Interactions 1492
C. Early Studies of RNA–Ligand Binding 1493
III. Nonspecific RNA–Ligand Interactions 1494
A. Organic Ligands 1494
B. Inorganic Ligands 1495
IV. Specific RNA–Ligand Interactions 1496
A. Organic Ligands 1496
1. Dyes and Intercalators 1496
2. Cations 1497
3. Guanosine 1497
4. Amino Acids 1498
5. Cofactors 1500
6. Aminoglycoside Antibiotics 1503
7. Antitumor Antibiotics 1506
8. Alkaloids 1507
B. Inorganic Ligands 1507
1. Rhodium and Ruthenium Polypyridals 1507
2. Nickel Complexes 1509
3. Iron Complexes 1509
4. Porphyrin–Metal Complexes 1510
C. Peptides 1510
1. Tat 1510
2. Rev 1511
V. Conclusions and Future Challenges 1511
VI. Acknowledgments 1512
VII. References 1512

Christine S. Chow received her A.B. degree in Chemistry from Bowdoin College in 1987. She obtained her M.S. at Columbia University in 1988 and her Ph.D. in Inorganic Chemistry (1992) at Caltech under the direction of Professor Jacqueline K. Barton. Her graduate studies involved the development of transition-metal complexes as probes for local variations in RNA tertiary structure. During postdoctoral studies from 1992 to 1994 with Professor Stephen J. Lippard at MIT, she examined the interactions of the HMG1 protein with cisplatin-modified DNA. She is currently an Assistant Professor at Wayne State University where her work focuses on new methods for site-specific incorporation of modified nucleosides into RNA, small molecule–RNA interactions, and the development of new RNA structural probes.

Felicia M. Bogdan is a native of Romania where she received her B.S. in Chemistry from the University "Babes-Bolyai" in 1991. She spent one year teaching high school chemistry and physics in Sibiu before moving to the United States. She joined the Chemistry Department at Wayne State University in 1993 and worked on dioxetane chemistry with Dr. Paul Schaap. Felicia is currently working toward her Ph.D. with Dr. Chow on the development of novel synthetic methods for the incorporation of modified nucleosides into RNA and examining RNA–ligand interactions. She is an active participant in the Great Lakes RNA Club.

I. Introduction

A considerable amount of attention has focused recently on new RNA-binding molecules. The interactions between RNA and biological macromolecules are clearly essential for many vital processes in molecular biology. Therefore, it seems reasonable to ask whether small molecular effectors might play a role in mediating these processes. In addition, the excitement over RNA-based viruses has fueled an interest in the development of potential RNA inhibitors. RNA offers several selective advantages over DNA as a therapeutic agent. First, chromosomal DNA is packaged extensively, significantly limiting its accessibility to small molecule reagents. On the other hand, RNA is highly structured, but generally considered to be an accessible target. Second, DNA repair systems are available in the cell, whereas analogous enzymes for RNA repair are virtually unknown. Finally, RNA exhibits a high level of diversity in terms of tertiary folding, and therefore will likely have a greater potential for selective targeting based on *structure*, rather than sequence.

The small RNA-binding molecules, or ligands, that are discussed in this review have typical molecular masses of less than 1000 Da. Molecules in this group represent an array of classes, including organic dyes,

S0009-2665(96)00415-3 CCC: $28.00

organic cations, inorganic metal complexes, and antibiotics. Molecular recognition between RNA and these types of ligands has become a central area of research in recent years. The development of high-resolution RNA structural analysis, using both X-ray crystallography[1−4] and nuclear magnetic resonance (NMR) spectroscopy,[5−7] has clearly made a profound impact in this area. Furthermore, *in vitro* selection experiments have proven to be a powerful approach for the development of RNA−ligand model systems.[8,9] Taken together, these experimental developments have provided valuable insights into RNA structure and its binding abilities. However, an even deeper understanding of RNA−ligand interactions is still needed for the rational design of a new generation of RNA-binding antiviral drugs, or molecules that can bind selectively to mRNA and inhibit translation.

The primary goal of this review is to provide the reader with an overview of RNA−ligand interactions with an emphasis on the role of RNA tertiary structure and specific binding modes. Several peptide examples will be included with the small molecules. There will only be a brief mention of proteins as RNA ligands, which has been the subject of recent reviews.[10,11] Similarly, the large RNA-binding toxins such as α-sarcin and ricin will not be included in this discussion.[12,13] The literature of the past 10 years will be summarized here, with some earlier works being included for an historical perspective. A complete understanding of the molecular mechanism of small molecule−RNA recognition is somewhat limited at this time, mainly because only several high-resolution structures have been reported to date. Those structures, as well as indirect biochemical data on RNA−ligand interactions, will be discussed.

This review is divided into five parts. The next section will summarize structural features of RNA, discuss potential binding sites, and give an overview of ligand-binding modes. The importance of understanding high-resolution RNA structures and folding properties of RNA will be stressed. A visualization of RNA−ligand complexes at the molecular level will ultimately provide a clearer understanding of the possible types of molecular interactions. The number of known RNA tertiary structures and the methods for studying them are increasing rapidly.[6,7,14,15] Here, we will review the early studies that provide a basis for the works discussed in later sections.

In the third section, we present examples of nonspecific binding interactions between ligands and RNA. In some cases, the specificity for RNA is greater than for DNA, which is an important criterion in the design of RNA-targeting therapeutics. Both organic and inorganic molecules can act as ligands for RNA recognition, and they each have their own unique features that affect binding. Inorganic complexes generally have the advantage of inducing RNA strand scission, which allows direct mapping of their binding sites. The fourth section reviews the specific RNA−ligand interactions which provide the strongest basis for RNA−drug design. Again, both organic and inorganic examples will be presented. In addition, two specific examples of peptide−RNA complexes will be presented here and their relationship to the small molecule−RNA interactions will be discussed. Finally, some conclusions and future challenges in the study of RNA−ligand interactions are discussed in section V.

A considerable amount of new information regarding RNA−ligand interactions has emerged over the last 10 years. This knowledge provides a basis for understanding drug action and will facilitate the design of new and improved drugs. Such agents may be developed as antiviral drugs effective against specific RNA viruses, or antibiotics with increased efficacy and lower toxicity. Clearly, more detailed investigations are needed for a complete understanding of the potential target RNA structures, ligand binding affinities, RNA or ligand conformational changes that are induced upon binding, and specific molecular contacts between the target RNA and small molecule.

II. General Considerations

A. RNA Structure

The versatility of RNA in binding to various ligands stems from its ability to assume a variety of tertiary conformations. Therefore, it is worth making some comments about RNA tertiary structure and describing some of the early experiments that established the groundwork for understanding RNA−ligand interactions. RNA structure has been studied extensively by a variety of spectroscopic methods and by chemical-probing experiments. Over the past two decades, several key RNA secondary and tertiary structures have been elucidated and some major ligand-binding modes have been defined.

Naturally occurring RNAs are either completely double helical with A-form conformations or globular with short double-helical domains connected by single-stranded regions. The double-helical regions that contain the Watson−Crick base pairs and purely single-stranded regions are considered as secondary structure. In the A-form duplex conformation, the nucleic-acid bases are pushed outward from the helix axis in the minor groove direction and tilted substantially with respect to the helix axis. The resulting helix has a shallow and wide minor groove and a major groove that is narrow and pulled deeply into the interior of the molecule. The basis for the A-form conformation is a $C_{3'}$-*endo* sugar pucker which leads to a short phosphate−phosphate distance of ∼5.9 Å. In general, these properties make the major groove of double-helical RNA inaccessible to many ligands.[16,17]

Evidence for tertiary, or higher order, structures in RNA was originally available from X-ray crystal data and NMR studies on tRNA.[1,2,5,18] Tertiary interactions as defined in tRNA by Kim *et al.*[1] are taken to mean the hydrogen bonds between the bases (non-Watson−Crick), between bases and the phosphate backbone, and between backbone residues. Figure 1 highlights the general tertiary structure features of tRNA. Mismatched base pairs that occur within double-helical regions and stacking interactions between bases will also be considered as tertiary interactions. Overall, the noncanonical intra- and interstrand base-stacking and hydrogen-bonding interactions serve to stabilize the RNA molecule beyond

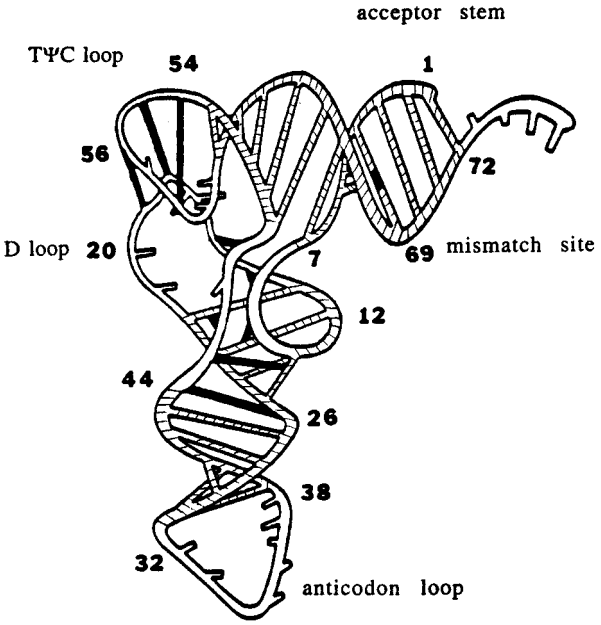


Figure 1. A three-dimensional representation of tRNA based on the crystal structure of yeast $tRNA^{Phe}$.[1] All bases involved in Watson–Crick base pairing are shown with hatched lines. The non-Watson–Crick base pairs and tertiary interactions are shown by black lines. The single-stranded regions are shown in white.

the secondary hydrogen-bonding interactions such as those generally found in DNA.

Yeast $tRNA^{Phe}$ and $tRNA^{Asp}$ contain a number of nonstandard base pairs in the outer corner of the classic L configuration and triple-base interactions in the center of the molecule.[1,2] In addition, tRNA also exhibits unusual changes in sugar pucker, phosphate–oxygen torsion angles, and turns or chain reversals in the loops that are stabilized by base–phosphate interactions. In conjunction with NMR spectroscopy, the X-ray analyses have revolutionized current opinions about RNA tertiary structure. A general theme that has emerged from the tRNA structure studies is the ability of RNA to display a wide variety of tertiary interactions. It was later discovered that these structures also constitute important binding sites for both natural and unnatural ligands. It should be noted that UV spectroscopy, thermal melting (T_m) studies, fluorescence spectroscopy, viscometry, and chemical probing methods have also been used to examine RNA tertiary structure, as well as the interactions with small molecules.

Recently, evidence for novel RNA tertiary interactions has come from both NMR[19] and X-ray crystal studies on the hammerhead ribozyme and group I intron fragment.[3,4,20] Some RNA structures that have been defined include RNA pseudoknots, RNA hairpins, bulge loops, mismatches, and triple-strand interactions in which specific hydrogen-bonding interactions define the folded structures. Figure 2 compares schematically some of the secondary structures to double-helical (or duplex) RNA. In most cases, the single-stranded regions of these structures can form unique tertiary interactions. Large hairpin loops (Figure 2B) are components of tRNA and smaller hairpins with four loop residues (tetraloops) are commonly found in ribosomal RNAs. Some tetraloops exhibit unusual base-pairing and hydrogen-

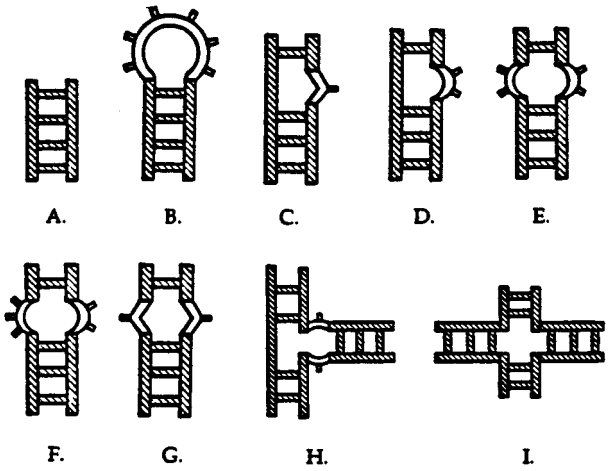


Figure 2. Representations of RNA secondary structures: (A) duplex, (B) hairpin loop, (C) single-base bulge, (D) multiple-base bulge, (E) symmetric internal loop (2:2), (F) asymmetric internal loop (3:2), (G) mismatch loop, (H) three-stem junction, and (I) four-stem junction. All bases involved in Watson–Crick base pairing are shown with hatched lines. The single-stranded regions are shown in white.

bonding interactions, for example the GAAA loop has a G–A base pair[21] and the UUYG loop (where Y is a pyrimidine residue) contains a U–G pair.[22,23] The different tetraloops also differ in structure with variable stacking arrangements and sugar conformations, thus providing unique binding sites for small molecules.

Bulges are formed when there is an unequal number of bases on the duplex strands. For single-base bulges (Figure 2C), the unpaired nucleotide can either stack into the duplex or loop out into solution, depending on the base composition. Multiple-base bulges (Figure 2D) may have different effects on the RNA structure such as a distortion of base stacking in the RNA duplex, bending of the RNA helix, or reduced stability of the duplex. Weeks and Crothers have suggested that certain base bulges will lead to an opening of the major groove accessibility, thus creating sites for binding by proteins or small molecules.[24]

Internal loops (Figure 2E–G) can involve symmetric or asymmetric loops (designated X:Y, where X and Y correspond to the number of nucleotides on each side of the loop) within the duplex. These loops may contain one (mismatch loop) or more unpaired or mispaired bases on each strand of the duplex. These sites have also been implicated as protein and small-molecule binding sites. More specifically, these sites generally serve to make the major groove of the RNA more accessible or they can undergo conformational changes when bound to ligands.[25] RNA junctions are regions that connect three or more stems (Figure 2H,I). A common four-stem junction occurs in tRNA and a three-stem junction is found in 5S rRNA. Continuity in the helicity at the junction site is possible, which is observed between the acceptor and T-stem or between the anticodon and D stems of tRNA. These sites may also serve as recognition sites for small planar molecules.

An example of a more complex RNA tertiary structure that has been characterized by NMR spec-

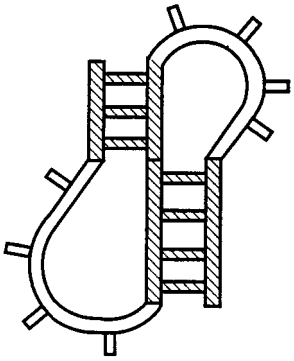


Figure 3. A secondary-structure representation of an RNA pseudoknot. All bases involved in Watson−Crick base pairing are shown with hatched lines. The single-stranded regions are shown in white.

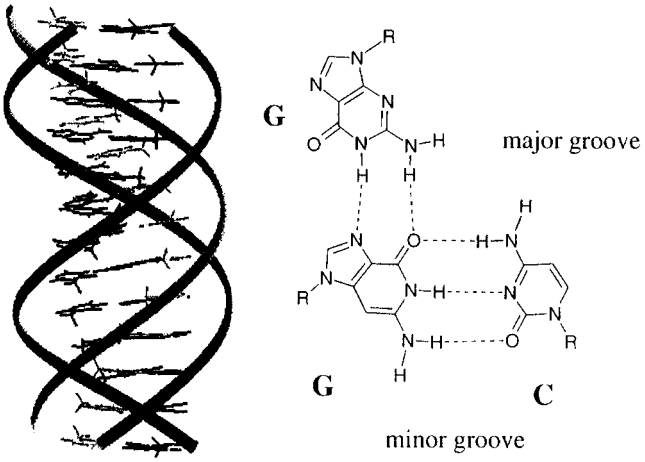


Figure 4. A model of an RNA triple helix. The Watson−Crick strands are shown in black. The non-Watson−Crick strand is shown in gray. A typical triple-base interaction (G−G−C), such as one found in tRNA, is shown to the right.

troscopy is a pseudoknot.[26] Schematically, this structure involves base pairing between one strand of an internal loop and a distinct single-stranded region, or between single-stranded regions of two separate hairpin loops (Figure 3). Although this structure has not been characterized in detail with respect to small molecule interactions, it has been implicated in the binding of proteins such as the human nerve growth factor and HIV-1 reverse transcriptase by *in vitro* selection experiments.[27,28] A less complicated tertiary interaction is the triple-stranded RNA, which uses Hoogsteen base pairs to add a polypyrimidine third strand to a polypurine−polypyrimidine duplex (Figure 4). In one type of triple strand, the two pyrimidine strands are antiparallel relative to one another. Triple-base interactions also occur in tRNA in which a third base interacts with base pairs in the major groove of a duplex region.[1] Other examples are known and will be discussed in section IV with respect to peptide binding and small molecule recognition.

At this time it is not known whether the structures mentioned here are a good representation of the total range of RNA tertiary interactions. It is clear, however, that RNA has the potential to form a vast array of complex structures and, like proteins, may contain binding clefts and pockets that can serve as unique recognition sites for substrates. The range of three-dimensional RNA structures that can be assumed by single-stranded RNA may even equal or exceed that of proteins. Although some structures appear to have common features (*e.g.*, tetraloops), it is likely that many other kinds of tertiary structures remain to be identified.

Individual RNA tertiary structures may serve as novel binding or target sites for specific RNA-binding ligands. It should be noted, however, that the RNA domains might also assume altered structures when bound to ligands. In some cases, the ligand might stabilize an otherwise floppy, or unstable, RNA structure. Similarly, the RNA target site may exist in multiple conformations whose equilibria are influenced by the ligand. In this manner, a small molecule may act as a conformational switch. Thus, an understanding of RNA target stability in the presence or absence of ligand is critical for the development of effective RNA-binding drugs. Overall, it appears that there is great potential for the development of small molecules as therapeutic agents that target unique structures of RNA.

B. Types of Noncovalent Interactions

The known (and likely many unknown) RNA structures are likely determinants of the types of interactions which form between the nucleic acid and ligand. An understanding of the specific associations between small molecules and RNA first requires a general knowledge of the types of binding modes. Not surprising, there are several different ways for a molecule to interact with RNA. Noncovalent, or reversible, interactions play substantial roles in binding specificity. The binding of free metal ions will not be discussed here (direct coordination or water-mediated binding), nor will covalent binding of inorganic complexes such as cisplatin [*cis*-diamminedichloroplatinum(II)].[29] In general, specific or nonspecific binding interactions are broadly defined. For our purposes, we will refer to nonspecific interactions as being governed largely by single or multiple binding modes at many sites along the RNA. Specific interactions will be considered when only one or a few binding sites are known. Site-specific recognition implies binding to a single, unique RNA motif. It could also mean binding to an RNA domain or structure, such as a nucleotide bulge site, that might occur in more than one RNA species. Another consideration for specific binding is the ability of the molecule to distinguish between RNA, DNA, and other biological macromolecules. For example, a ligand with high specificity for RNA (*i.e.*, a single or few binding sites) might also have high affinity at comparable DNA sites. In order for a ligand to be an effective RNA-targeting drug, low binding affinity for DNA is desirable.

Electrostatic effects between cationic species and the negatively charged RNA phosphate backbone are ideal nonspecific interactions that might be important for enhancing the binding of small molecules. These binding interactions generally occur along the exterior of the helix (Figure 5A). A second general binding mode is a groove-bound association. This interaction generally involves direct hydrogen-bonding or van der Waals interactions with the nucleic

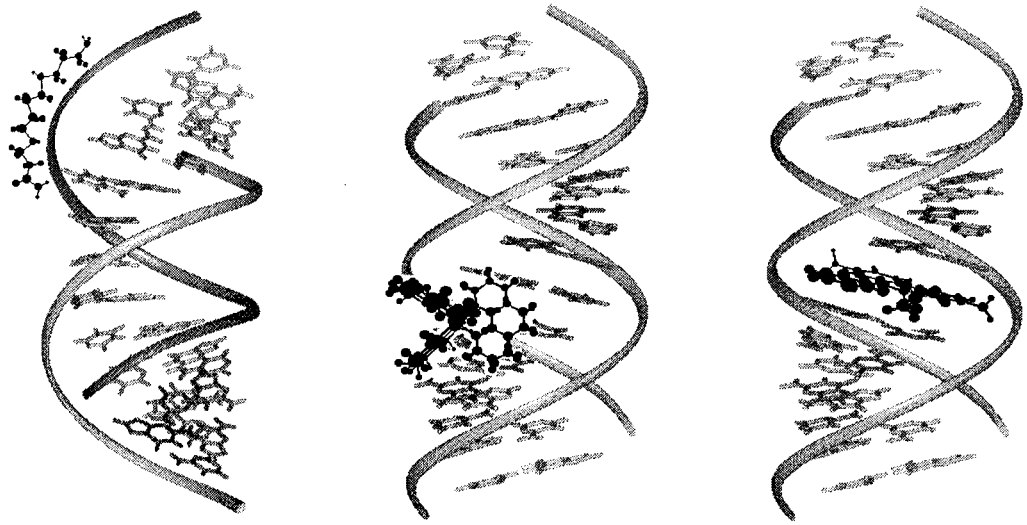


Figure 5. Representations of three ligand-binding modes with nucleic acids: (A) electrostatic (B) surface (or groove) binding, and (C) intercalation.

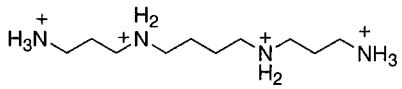
acid bases in the deep major groove or the wide shallow minor groove of the RNA helix (Figure 5B). In some cases, the structure of the ligand may preclude binding at all sites along the RNA groove, thus specific binding to RNA by this mode is possible. In addition, hydrogen bonding between the various acceptor and donor groups on the ligand in the groove may also govern specific RNA−ligand interactions. The electrostatic and groove-binding modes do not require a change in the RNA conformation, but an alteration in the RNA structure upon binding is possible.

Stacking interactions between RNA bases and aromatic ligands are important in defining a third type of binding mode known as an intercalation (Figure 5C). Intercalation is classically defined in DNA when a planar, heteroaromatic moiety slides between the DNA base pairs and binds perpendicular to the helix axis.[30] In contrast to the previous examples, this binding mode requires a distortion of the RNA helix in order to accommodate the binding ligand. Specifically, the adjacent base pairs need to separate to allow insertion of the planar aromatic molecule, which can lead to an unwinding of the RNA helix and to changes in the sugar conformations.

These types of noncovalent interactions will be used to classify the binding interactions between an array of ligands and their RNA targets. First, we will consider the nature of the binding interactions, then the effects of RNA conformation and ligand structure on the molecular associations, and finally use structure-binding relationships to analyze the role of functional groups in mediating the specific type of interaction.

C. Early Studies of RNA–Ligand Binding

The studies discussed in this section have provided a basis for understanding modes of small-molecule binding to RNA. A spermine−tRNA^{Phe} complex has been crystallized and the structure solved at 2.5 Å resolution.[31] The polyamine binds at two major sites, the first in the major groove at the end of the anticodon stem and the second near the variable loop near phosphate-10 at a turn in the RNA backbone. Although the spermine molecule (Figure 6A) has the potential for an extended structure (~15 Å long), it binds to RNA with a bent shape in the crystal. Extensive hydrogen-bonding contacts with the RNA are apparent, and electrostatic effects of this polycation are also important. Binding at the anticodon site of RNA leads to a shortening of the phosphate−phosphate distance by ~3 Å, presumably because of the concentration of positive charges in the major groove. In the second site, the spermine wraps around a phosphate residue and helps to stabilize the negative charge of the RNA backbone at the center of the tRNA molecule.



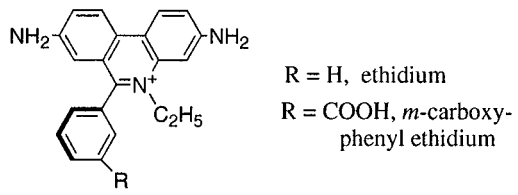


Figure 6. Chemical structures of spermine (A) and ethidium intercalators (B).

More recent NMR studies have shown that spermine and related polyamines bind to additional sites of tRNA, in particular the TΨC loop, that were not seen in the crystal structure.[32−34] In contrast to the static crystal data, these studies are reflective of a more dynamic interaction between spermine and its RNA target. Overall, the binding of spermine helps to maintain the tertiary folding of the tRNA and its binding assists in the recognition by proteins such as aminoacyl-tRNA synthetase. Nature uses this flexible polyamine, which can be extended to bind to

DNA or bent to recognize specific RNA tertiary structures either by electrostatic binding on the exterior of the helix or by major groove surface binding and hydrogen-bonding interactions. In addition, specific interactions with RNA may provide a basis for the antiproliferative action of this class of molecules.[35]

Another early example of RNA recognition by a ligand involved the well-studied yeast $tRNA^{Phe}$. Solution studies indicated a principal binding site for an ethidium molecule (Figure 6B) on tRNA, presumably by a classical intercalative binding mode typical for this planar aromatic dye binding to double-stranded DNA.[36] In contrast, X-ray diffraction studies on crystals of the dye and yeast $tRNA^{Phe}$ revealed a nonintercalative mode of binding.[37] It was clear, however, that the specific mode of binding was dependent upon the RNA tertiary structure and its ability to fold into nonstandard conformations. The binding site in the crystal was near the variable loop in a cavity created by the RNA tertiary structure, specifically at a hairpin turn. The dye stacks with the U8 residue which is involved in a reverse−Hoogsteen base pair. On the basis of NMR data, Jones *et al.*[38,39] later proposed that ethidium binds by intercalating between the base pairs of U6−A67 and U7−A66, close to the site found by X-ray studies. Fluorescence-detected circular dichroism studies supported the intercalation binding mode.[40] Presumably, the differences between the solution and solid states influence the binding of the ligand to its RNA target site. Overall, the results indicate the possibility of more than one binding mode for this class of molecules, specifically an intercalative or surface-binding mode. By using spermine and ethidium, all three binding modes were identified on tRNA. In addition, these early results indicated the possibility of targeting RNA in a site-selective manner.

III. Nonspecific RNA−Ligand Interactions

The first goal in designing an RNA-binding ligand is to determine its binding affinity and level of specificity for RNA. Nonspecific binding interactions are not likely to be important for drug candidates, however, an understanding of these types of interactions can lead to the design of better RNA-binding reagents. Therefore, the second goal is to determine the nature of the ligand−RNA interaction and elucidate the binding modes for a specific class of molecules. The factors that give rise to affinity and specificity will be discussed and for some ligands a comparison between DNA and RNA binding will be made.

A. Organic Ligands

Berenil (1,3-bis(4′-amidinophenyl)triazene) (Figure 7A), an organic cation, exhibits a mixture of intercalative and minor-groove binding modes with RNA duplexes (poly(rA)·poly(rU)), but prefers to bind duplex DNA.[41] When the drug binds to duplex RNA, the thermal stability of the RNA is enhanced. Furthermore, salt-dependent melting data suggest that the positively charged amidino groups of berenil participate in complexation of the drug to RNA. Similarly, DAPI (4′,6-diamidino-2-phenylindole) (Figure 7B) is proposed to bind by an intercalative mode to duplex RNA (poly(rA)·poly(rU)).[42] As with berenil, this drug binds to DNA by a different mode (groove-binding) and with higher affinity. In contrast, propidium (Figure 7C) interacts with both DNA and RNA by an intercalative binding mode and with similar binding affinities. Overall, one can conclude from these studies that the potential exists for the design of molecules with high affinity for RNA and with binding modes that are different than for DNA. Further optimization is necessary, however, in order to develop molecules with higher selectivity for RNA over DNA and with relatively few binding sites.

The interaction of a series of 2-phenylquinoline derivatives (Figure 7D) with RNA has also been investigated.[43] Substitutions at specific sites result in molecules with different physical properties, such as barrier to rotation, which ultimately affect the binding interactions with RNA. Piperazyl substituents with positive charges provide molecules with stronger affinity for RNA. These analogues are proposed to interact by intercalative (R_2 = piperazyl, $R_1 = R_3 = H$) or threading intercalation modes (R_3 = piperazyl, $R_1 = R_2 = H$), with the later compound binding with the higher affinity. Further studies by Zhao and co-workers indicated that small changes in cationic substituents can strongly affect RNA binding affinity as well as the binding mode.[44] In particular, the interactions of diphenylfuran analogues of furamidine (Figure 7E) with RNA were examined. Most of the derivatives bind to duplex RNA by an intercalative mode, in contrast to a minor-groove-binding mode to A−T-rich DNA. One substituent, an imidazoline cation (compound 2), gives a high-affinity RNA-binding species, and can also enhance a nonintercalative binding mode.

Several important conclusions can be drawn from the above-mentioned studies. First, several of the drugs examined were originally designed as DNA-binding drugs. The fact that they also bind to RNA is significant because RNA interactions could be a source of drug loss for those designed to target DNA. Second, selective binding modes to RNA could be further enhanced for the design of new RNA-targeting drugs, in which DNA binding is minimized. In one such example, McConnaughie *et al.*[45] have demonstrated that a series of polycationic ligands (Figure 7F,G) has the ability to selectively interact with RNA by a groove-binding mode. As with the diphenylfuran derivatives, those molecules with basic residues (*e.g.*, imidazole) show the greatest degree of selectivity. For example, imidazoline-substituted compounds (compounds 9 and 18) show enhanced binding in a surface-bound mode compared to other substituted molecules. The most important feature of these compounds is the presence of basic residues and hydrogen-bonding moieties that can enhance or regulate binding to RNA. In addition, the steric bulk of these nonaromatic compounds can prevent binding in the narrow minor grooves of DNA, which is generally the target site for cationic ligands.

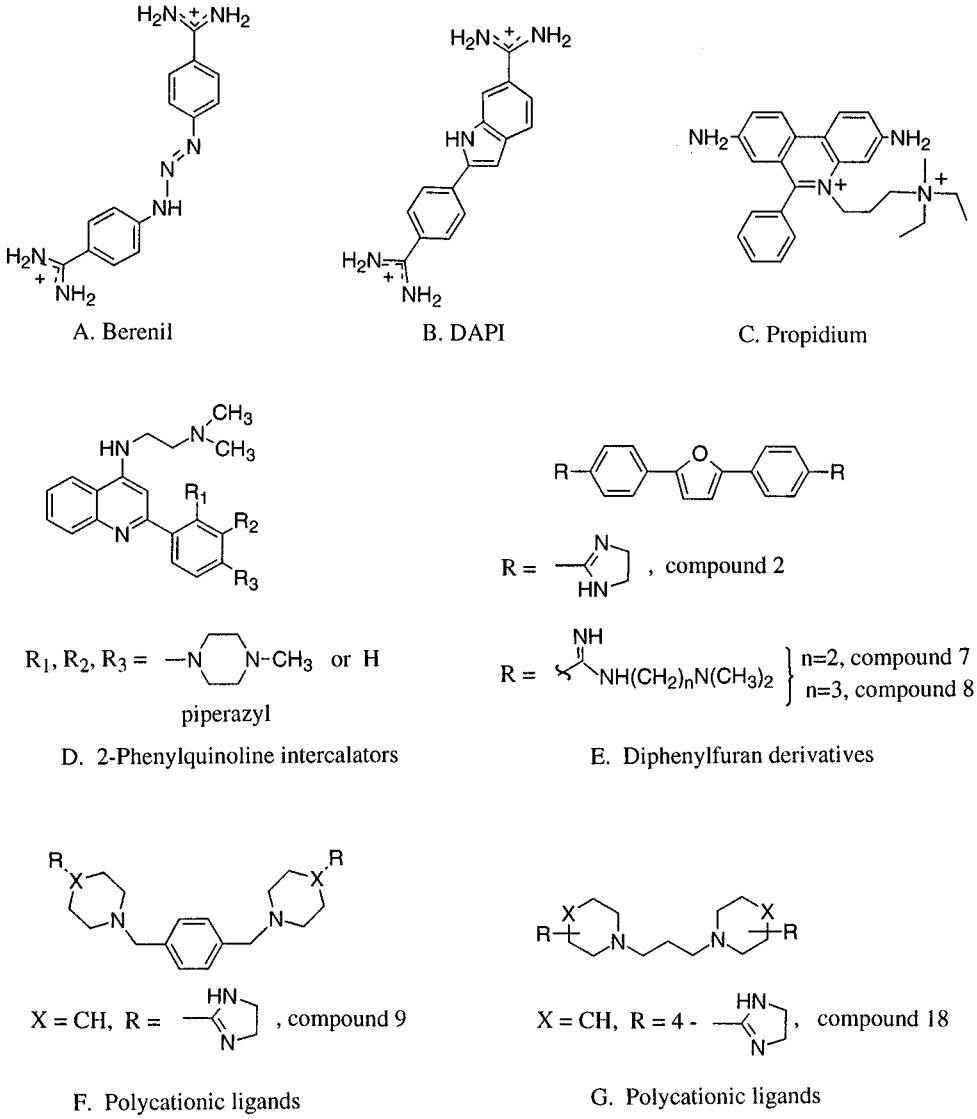


Figure 7. Chemical structures of organic cations and intercalators.

B. Inorganic Ligands

There is a battery of inorganic complexes that can lead to RNA strand scission and can be used to investigate the secondary and tertiary structure features of RNA. These features may be important for binding to related small molecules or ligands. Several complexes that exhibit relatively nonspecific binding and cleavage of RNA will be highlighted in this section.

$Fe(EDTA)^{2-}$ (Figure 8A) is a nonspecific-binding ligand that catalyzes RNA strand scission in the presence of sodium ascorbate and H_2O_2. This metal complex exhibits similar reactivity towards single- and double-stranded RNA. The metal complex can produce neutral radicals that react indiscriminately with the RNA backbone where it is exposed to solvent. In this manner, the complex can be used to probe the interior and exterior regions of RNA.[46,47]

Similarly, bis(1,10-phenanthroline)copper(I) [$Cu(phen)_2^+$] (Figure 8B) promotes RNA strand scission, but preferentially recognizes single-stranded regions of the RNA.[48] The specificity and efficiency of the cleavage reaction are in part dependent upon the presence of functional groups on the phenanthroline ligand, such as 4,7-dimethyl, 5,6-dimethyl, and 3,4,7,8-tetramethyl, but the interactions of these complexes with RNA at the single-stranded loops, bulges, or mismatches have not yet been characterized in detail.[48–50] Nonetheless, this complex served as an early example of an RNA binding and cleavage agent with modest specificity. Some nickel, rhodium, ruthenium, and related complexes also display low specificity in binding to RNA, but have been developed further into structure-specific reagents. This topic will be covered in section IV of this review.

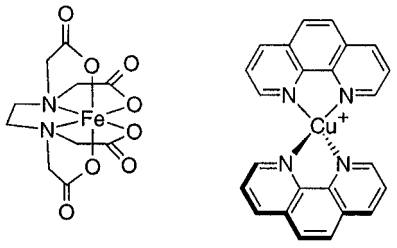


Figure 8. Chemical structures of inorganic complexes.

IV. Specific RNA–Ligand Interactions

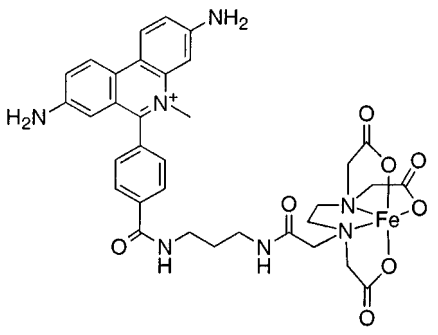
In comparison to DNA, there is a relative paucity of examples of small molecules that specifically recognize RNA. The ability of RNA to recognize small molecules has been addressed in part by *in vitro* selection-amplification experiments starting with pools of random RNA sequences.[51,52] This section will cover some of the basic findings such as specific binding of selected RNAs to amino acids, organic dyes, aminoglycoside antibiotics, and biological cofactors, as well as discuss some related binding studies with inorganic and peptide ligands. We will draw on specific examples to make some general conclusions about the binding modes and molecular interactions that govern the recognition of RNA by specific classes of small molecules.

A. Organic Ligands

1. Dyes and Intercalators

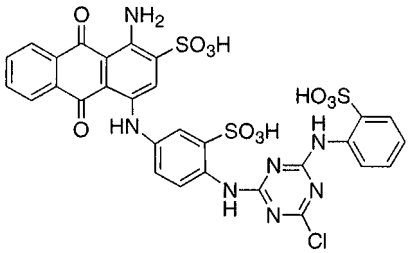
As mentioned in section II, some early studies concerned the binding of dyes and intercalators to tRNA. Further studies suggested an increased binding of ethidium to RNA duplexes in the presence of mismatched bases.[53] Later work by Kean, White, and Draper[54−56] showed that RNA helices containing single-bulged residues have unusually high affinities for the dye. Binding studies with the related molecule MPE-Fe^{II} (methidiumpropyl−EDTA−Fe^{II}) (Figure 9) revealed that CpG sequences are the preferred binding sites and that A, G, or U bulges to the 3′ side of the C further enhance the binding. The binding site was identified by MPE-Fe^{II}-induced RNA cleavage. In addition, interactions at the bulge site result in a conformational change of the RNA. Noncanonical interactions involving the bulge residues are likely to affect RNA stability. These experiments set precedence for other small molecule−RNA interactions in which the RNA tertiary structure and its ability to undergo a change in structure is critical for ligand binding. In addition, the effects on the model RNA hairpins in White and Draper's studies are likely relevant to larger RNAs in the context of ribosomal RNA and more extensive tertiary structures that are involved in protein recognition or RNA−RNA interactions.

Similarly, Tanner and Cech demonstrated that the *Tetrahymena* pre-rRNA self-splicing reaction can be inhibited by intercalating molecules.[57,58] In particular, binding of the dyes ethidium, acridine, and proflavin affect reactivity of the ribozyme in a reversible manner with binding constants (K's) $\sim 10^5$ M^{-1}. Two possible mechanisms involving dye binding could explain the inhibitory effects. First, the dyes could bind and disrupt the RNA secondary or tertiary interactions that are necessary for the splicing reaction. Conversely, the dyes could bind and stabilize a specific structure in the RNA and prevent a conformational change that is crucial for the reaction to occur. The fact that ethidium inhibition is strongly dependent on $MgCl_2$ concentration suggests that RNA structure is important for binding of the dyes. The group I intron splicing reaction is also inhibited by MPE-Fe^{II}.[57] In this case, the actual site of binding can be localized by MPE-Fe^{II}-induced RNA cleavage. Perhaps not surprising, one strong binding site lies between two A−U base pairs at a bulged A residue. Overall, these dyes have not become useful drugs because of their high affinities and nonspecific binding to DNA. However, these dyes can be utilized in RNA structure mapping because of their moderate level of specificity for RNA tertiary structure.

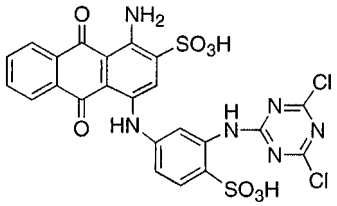


MPE-Fe(II)

Figure 9. The structure of methidiumpropyl−EDTA−Fe^{II} (MPE-Fe^{II}).



A. Cibacron Blue 3GA (CB)



B. Reactive Blue 4 (B4)

Figure 10. Structures of organic dyes Cibracon Blue (CB) and Reactive Blue (B4).

In 1990, Ellington and Szostak reported an *in vitro* selection method that was used to isolate subpopulations of RNA molecules that bind specifically to organic dyes.[52] The selected RNAs are able to fold so as to create specific binding sites for Cibacron Blue 3GA (CB) (Figure 10A), Reactive Red 120 (R), Reactive Yellow 86 (Y), Reactive Brown 10 (BR), Reactive Green 19 (GR), and Reactive Blue 4 (B4) (Figure 10B). The CB- and B4-binding RNAs were characterized in more detail and the dissociation constants (K_d's) were determined to be $\sim$100 and 600 μM, respectively. These experiments were important because they demonstrated the ability to find RNA sequences within a random pool that are capable of specific binding to a variety of small-molecule ligands, although the specific binding modes were not identified.

2. Cations

For intercalative cations, the structural basis for RNA affinity depends on how well the cation fits into the intercalation site and stacks with the RNA base pairs. Thus, the preference for intercalative binding can be reduced by using specific substituents that favor groove binding. One obvious extension of this work would be to design molecules that bind selectively at accessible sites along the RNA major groove and exclude binding to DNA. As mentioned, White and Draper found that ethidium binds with a 4−5-fold increase in affinity at CpG sites when they are 3′ to a bulged residue.[54] Ethidium has a high affinity for DNA and therefore does not have the desired RNA binding selectivity. In the search for molecules with enhanced RNA affinity relative to DNA, Ratmeyer *et al.*[59] reported that an ethidium analogue with a *m*-carboxyphenyl substituent (Figure 6B) binds strongly to RNA containing a bulged residue in comparison to RNAs and DNAs without bulges. Therefore, this ethidium analogue shows promise as a selective RNA-binding agent. Perhaps the negatively charged carboxyl group is a factor in the binding specificity. This compound was used to target a base-bulged duplex from the TAR sequence (Figure 11) of HIV-1, an RNA with known affinity for the HIV-1 Tat protein. Related studies with cationic compounds and TAR have also been reported by Bailly and co-workers.[60]

Subsequent studies demonstrated improved binding specificity by the use of linked intercalators. Specific dyes were covalently linked to nucleic acid bases in an attempt to enhance the binding specificity at RNA base bulges, while limiting intercalative binding into DNA.[61] The effects on the melting temperatures (T_m's) were used to measure relative binding by base-linked ethidium and acridine dyes. The largest enhancement in T_m was observed for a thymine-linked ethidium (Figure 12A) bound to a duplex RNA with a bulged A residue. Similarly, a 2,6-diaminopurine-linked acridine (Figure 12B) bound favorably to duplexes containing A or U bulged sites.

In further studies with diphenylfuran derivatives, Ratmeyer *et al.*[62] demonstrated that these compounds

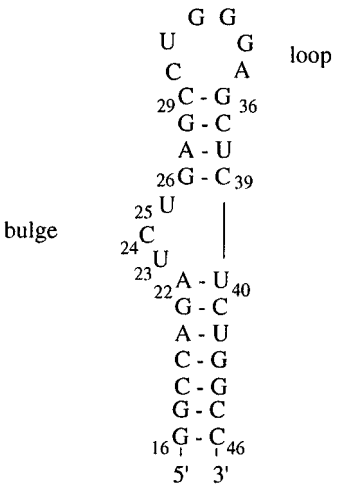


Figure 11. The sequence and proposed secondary structure of an HIV-1 TAR RNA fragment with indications of bulge and loop motifs.

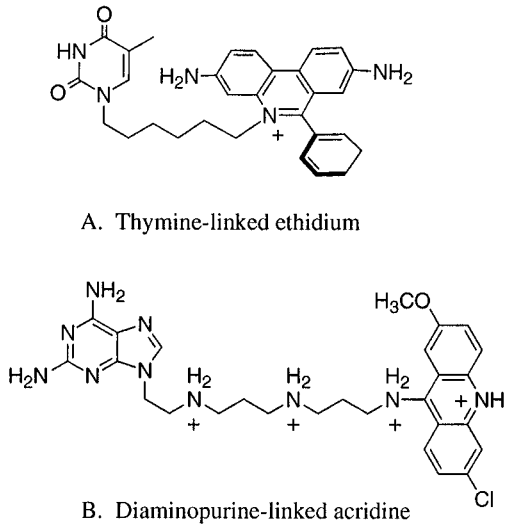


Figure 12. The structures of ethidium- and acridine-based conjugates.

bind to RNAs with specific tertiary structures more avidly than duplex RNAs. Specifically, the di- and tetracationic compounds 2, 7, and 8 (Figure 7E) show the greatest level of discrimination (K's $\approx 10^6-10^7$ M^{-1}, and 100-fold less for poly(rA)·poly(rU)). In addition, these compounds can inhibit the interactions of the viral HIV-1 Rev protein with an RRE RNA fragment at concentrations less than 1 μM. Circular dichroism studies with RRE RNA hairpins indicate that these compounds bind at the asymmetric internal loop of the RNA by a nonclassical binding mode. In addition, these drugs cause a change in the RNA conformation that is likely to be critical for the Rev-RRE inhibition. These results are very promising, but additional studies are clearly needed in order to understand the exact binding modes and basis for specificity. A detailed understanding of these molecular interactions will assist in the future development of these and related cationic molecules into an effective class of RNA-targeting drugs.

3. Guanosine

An early example of small molecule recognition by a macromolecular RNA is the binding of guanosine and its analogues by the self-splicing group I intron from *Tetrahymena*.[63,64] This well-known ribozyme contains a specific binding site for the nucleoside. In addition, the exogenous guanosine actually participates in the splicing reaction. Bass and Cech offered the first evidence that catalytic RNAs can utilize noncanonical, in addition to standard Watson−Crick, base pairs to bind substrates. A highly specific guanosine recognition site has been identified by the use of mutant versions of the ribozyme and guanosine analogues. The K_m for guanosine is 32 μM while deoxyguanosine and dideoxyguanosine are competitive inhibitors of the splicing reaction (K_I's = 1.1 and 5.4 mM, respectively). These results indicate that the ribose hydroxyls are necessary for optimal binding of guanosine at the active site. Interestingly, Yarus showed that L-arginine and a related fragment, guanylurea, can also bind specifically and reversibly at or near the guanosine-binding site.[65,66] The α-amino and carboxyl groups are not important for binding,

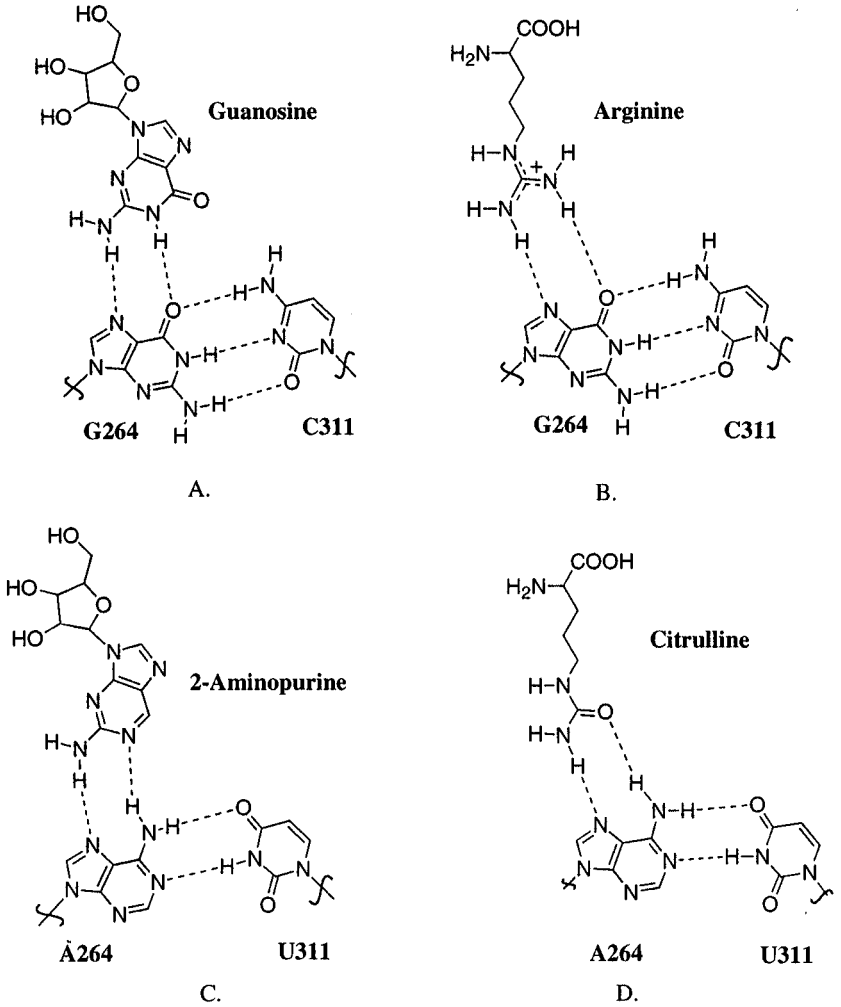


Figure 13. The proposed binding of guanosine (A) and arginine (B) to the base pair G264−C311 of the *Tetrahymena* group I intron and of 2-aminopurine (C) and citrulline (D) to a mutant base pair A264−U311.[67]

suggesting the importance of the guanidinium side chain for specific RNA binding and competitive inhibition of the splicing reaction. Yarus has developed a model for the binding in which the guanidino group of either guanosine or arginine forms similar hydrogen bonds in the GTP site. Further work by Michel *et al.*[67] with guanosine analogues and RNA substitutions showed that a phylogenetically invariant guanosine residue in the G264−C311 base pair forms a base-triple interaction with either guanosine or arginine (Figure 13, parts A and B, shows the guanosine-binding site nucleotides and proposed hydrogen-bonding interactions). Furthermore, the specificity of the ribozyme was altered when the G−C pair was mutated to A−U. Specifically, citrulline and 2-aminopurine were preferred over arginine and guanosine, respectively, thus providing credence for the base-triple model (Figure 13C,D). These studies demonstrate the role of hydrogen bonding in small-molecule specificity for RNA. Yarus and Majerfeld have also suggested the importance of substrate stacking interactions with a neighboring residue for selection of the appropriate ligand.[68]

4. Amino Acids

Arginine exhibits several key structural features that make it an ideal RNA-binding molecule. It contains positive charges, planar hydrogen-bonding patterns that strongly resemble the nucleotide bases, and a π system that is well suited for stacking interactions with the bases. Examples of specific binding of arginine by RNA will be discussed in this section. In addition, several RNA-binding proteins exist in which the binding to RNA is governed by specific arginine residues. The protein binding will be discussed later, but it will be noted here that there are several different types of arginine-binding sites found in naturally occurring RNAs. In general, these sites appear to involve internal loops and bulge segments of RNA whose sequences include arginine-coding triplets.

Following the work with the *Tetrahymena* group I intron mentioned above, a second arginine-binding site was identified. Specific binding of the HIV-1 Tat protein to TAR, an RNA hairpin with a six-nucleotide loop and trinucleotide bulge (Figure 11) located at the 5′ end of the viral mRNA, is mediated by a single arginine within a nine-residue basic peptide.[69] TAR mutants without the bulge residues show decreased affinity for Tat and peptide fragments.[24] Free arginine also binds to TAR, and this fact was exploited to obtain an NMR structure of the RNA in the presence of the amino acid analogue argininamide.[70] Upon argininamide binding in the major groove, the bulge region of TAR changes its conformation. In particular, binding of the ligand leads to an unstack-

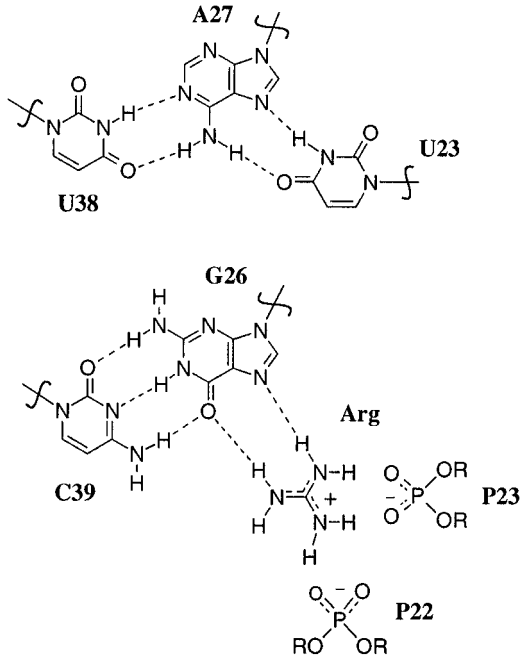


Figure 14. Schematic representation of the proposed base triple between U23 and A27−U38 of TAR RNA and the interaction of the guanidinium side chain of argininamide with G26 and phosphates 22 and 23.[70]

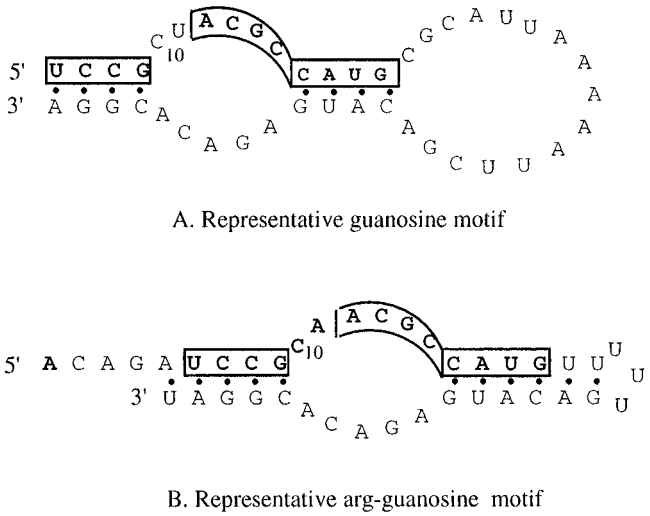


Figure 15. Consensus sequences and proposed secondary structures for ligand-binding RNAs: (A) guanosine and (B) arginine−guanosine motifs. The flanking sequences that represent the PCR primer sites have been omitted for simplification. The consensus sequences are shown in boldface. The sequences that are conserved between the selected RNAs are boxed. C10 of the guanosine motif can vary; one to two nucleotides of any kind except U can occur at this position. Position 11 of the guanosine motif is a pyrimidine. C10 and A11 of the joint site are conserved.[73]

ing of the three bulged nucleotides and coaxial stacking of the RNA helical stems. Nucleotides that are essential for Tat binding (U23−A27−U38) form a triple-base interaction (shown schematically in Figure 14) and help stabilize the argininamide as it forms hydrogen bonds with the adjacent G26 and a neighboring phosphate group. The specificity of the arginine-TAR interaction appears to be largely mediated by the RNA tertiary structure and the ability of the RNA to undergo a conformational change. From this work, three general observations can be made: (1) molecular contacts with the RNA occur at a bulge site where the major groove is wider and more accessible to the ligand and a specific binding site has been generated;[24] (2) the initial arginine contacts induce a conformational change within the RNA bulge region; and (3) specific hydrogen-bonding and stacking interactions between the RNA and amino acid side chain help stabilize the binding interaction.

On the basis of the findings that arginine could bind specifically to the group I intron and TAR RNA, Connell *et al.*[71] then used *in vitro* selection methods to evolve L-arginine-binding RNA motifs. Three motifs were discovered with high specificity for arginine (K_d's $\approx$ 0.2−0.4 mM). The binding sites contain specific internal loops and bulged residues, each with distinguishable specificities as determined by analogue binding affinities. Similar to the group I intron, all three motifs exhibit affinity for guanosine 5′-monophosphate. One site is stereoselective, preferring D-arginine over L-arginine. Tao and Frankel also performed *in vitro* selection experiments with arginine and obtained TAR-like RNAs with three-nucleotide loops.[72] Thus, it is evident that different selection conditions will yield different RNAs. These results reflect the ability of arginine to bind to RNAs with alternate conformations. The selected arginine-binding RNAs might represent binding sites for arginine-binding proteins or ligands with guanidino groups, and may therefore provide useful models for understanding the different binding properties of arginine and related ligands. Several key features from these studies should be emphasized. First, the fact that all of the selected arginine-binding motifs also bind guanosine 5′-monophosphate supports the idea that there is an overlapping set of RNA motifs that recognize structurally similar ligands. Second, there are many ways to fold RNA to create specific binding sites for arginine, and likely for other small molecules as well.

As demonstrated with the group I intron, a single nucleotide is sufficient to expand or alter the binding properties of an RNA molecule. In addition, an RNA binding site can have dual specificity. For example, the group I intron has affinity for different ligands (*e.g.*, guanosine and arginine) that share common features. The nucleoside and amino acid have a common hydrogen-bonding network, and therefore have the ability to bind to the same site of the RNA.[65] Similarly, *in vitro* selection experiments led to RNAs that bind to both of the two ligands (guanosine and arginine) or to one ligand only (guanosine).[71,73] A comparison of the selected RNAs reveals that the binding sites are virtually identical with the exception of two sites within an asymmetric internal loop (6:5) and a 5′ extension of the joint site (Figure 15). Thus, it is apparent that a single or double nucleotide substitution can either expand or limit the binding specificity of an RNA. Two general conclusions can be drawn from these studies: (1) apparently unrelated RNA sequences can fold to perform the same function, in this case binding to arginine or guanosine; and (2) an RNA binding site can tolerate simple substitutions that will generate new specificities while retaining the old binding properties.

Similarly, an L-citrulline binding RNA can be evolved through *in vitro* selection into an L-arginine

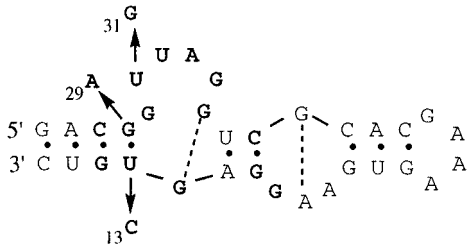


Figure 16. Secondary structure proposed for a citrulline-specific aptamer. The conserved bases are represented in bold. At positions 13, 29, and 31 are indicated the three bases that are critical for arginine specificity (C, A, and G, respectively). The noncanonical base-pairing interactions are represented by dotted lines.[75]

binding RNA.[74] The two ligands differ in that one has a neutral urea group (hydrogen-bonding acceptor) and the other contains a basic guanidino group (hydrogen-bond donor) (see Figure 13, parts D and B, respectively). The initial selection for L-citrulline revealed RNAs with K_d's ≈62−68 μM that can fold into the same binding motif containing asymmetric internal loops. Reselection against arginine with a single mutagenized L-citrulline binder resulted in an arginine-binding pool. Thus, a general arginine-binding motif was derived that is only slightly different from the original L-citrulline-binding motif. Isolated L-citrulline and L-arginine aptamers bind to their cognate amino acids with K_d's $\sim 10^{-5}$ M without affinity for the noncognate amino acids. Both RNAs contain a two-stem region flanking two internal asymmetric loop regions (6:1 and 1:3) and a third stem, as shown in Figure 16. Furthermore, one aptamer from the arginine-binding pool differed from the starting 44-nucleotide (later shortened to 33 nucleotides) L-citrulline binder by only three specific mutations, U13C, G29A, and U31G.

Further studies with the selected L-citrulline- and L-arginine-binding RNAs revealed that nearly every base in the conserved region is important for binding to the ligand. In addition, a conformational change in the RNA occurs upon amino acid binding and the noncanonical base pairs are stabilized by the complexed ligand. NMR studies[75] revealed that the asymmetric internal loops are not ordered in the free RNA, but fold into a compact structure upon amino acid binding. During this conformational change, a specific hydrogen-bonding contact with a Watson−Crick base pair is formed, thus forming a triple contact that is reminiscent of the triple-base interaction in the TAR−arginine complex. Furthermore, a cluster of conserved purine residues forms a hydrogen-bonding network, including a G12−G35 mismatch pair that is important for forming a scaffold for amino acid binding. The three variant nucleotides (U/C13, G/A29, and U/G31) make specific hydrogen-bonding contacts and stacking interactions with the individual amino acids: two form a triple contact with the amino acid, and the third caps the binding cleft. On the basis of these studies, two key points can be made. First, the RNA is able to conform into a compact structure within a well-defined binding pocket upon ligand binding. The ability of the RNA to undergo a conformational change depends on noncanonical base pairs and single-stranded residues within an internal loop. Second, specificity is achieved by using a relatively few number of residues while the general scaffold remains unchanged. In other words, a specific set of nucleotides determines the binding of the amino acids L-citrulline and L-arginine, and a smaller subset of these residues determines the specificity for the individual ligands.

In another *in vitro* selection experiment, Famulok and Szostak found RNAs that recognize tryptophan agarose in a stereospecific manner.[76] Selections against D-tryptophan−agarose resulted in RNAs that were specific with a K_d of 18 μM. The selected RNAs did not bind to L-tryptophan−agarose, even at higher substrate concentrations ($K_d \geq 12$ mM). Even though the complexity of the pool was high (∼100−1000 different sequences) with no obvious sequence homology, these experiments demonstrated that RNA has the ability to bind aromatic amino acids. The stereospecificity demonstrates that there must be a specific interaction in the binding pocket and that RNA is able to discriminate between chiral molecules. Similarly, Majerfeld and Yarus[77] have isolated RNAs with specificity for the side chain of L-valine. The selected consensus RNA contains a highly conserved asymmetric (4:10) internal loop adjacent to a mismatched base pair (G−U). This RNA also binds to its cognate amino acid in a stereoselective manner and can distinguish between the aliphatic amino acid side chains on the basis of their size and configuration. This example demonstrates that RNA can select ligands on the basis of aliphatic interactions, in addition to hydrogen bonding and stacking interactions as seen in previous examples. This type of interaction may have a biological role and could possibly be utilized in the design of new RNA-binding ligands. However, further studies are necessary to determine the molecular structure of the valine−RNA complex and to fully understand this high level of molecular discrimination.

5. Cofactors

In vitro selection techniques have also been used to produce RNAs with specific binding to ribose-containing cofactors, in particular ATP, cyanocobalamin, flavins, and nicotinamide (Figure 17). In 1993, Sassanfar and Szostak[78] reported the selection of small RNAs (40 nucleotides) with high affinity (K_d < 50 μM) for the common biological cofactor ATP. Similar to the arginine-binding RNAs, the ATP-binding RNAs contain a consensus sequence that exhibits tight binding and close contacts with the ligand (ATP) and undergoes a conformational change upon ATP binding. The isolated RNA aptamers also bind to ADP, AMP, and the adenosine moiety of FAD. Modification of four adenosine positions, including the 2′-hydroxyl, leads to a decreased affinity for the selected RNAs. In addition, cloning and sequencing showed that 11 nucleotide positions were invariant, and led to the predicted secondary structure consisting of an asymmetric internal loop (11:1) (Figure 18). More specifically, a purine-rich 11-nucleotide loop lies opposite a single G residue and is flanked by Watson−Crick stems.

Two independent groups have recently solved the structures of AMP−RNA complexes by using NMR

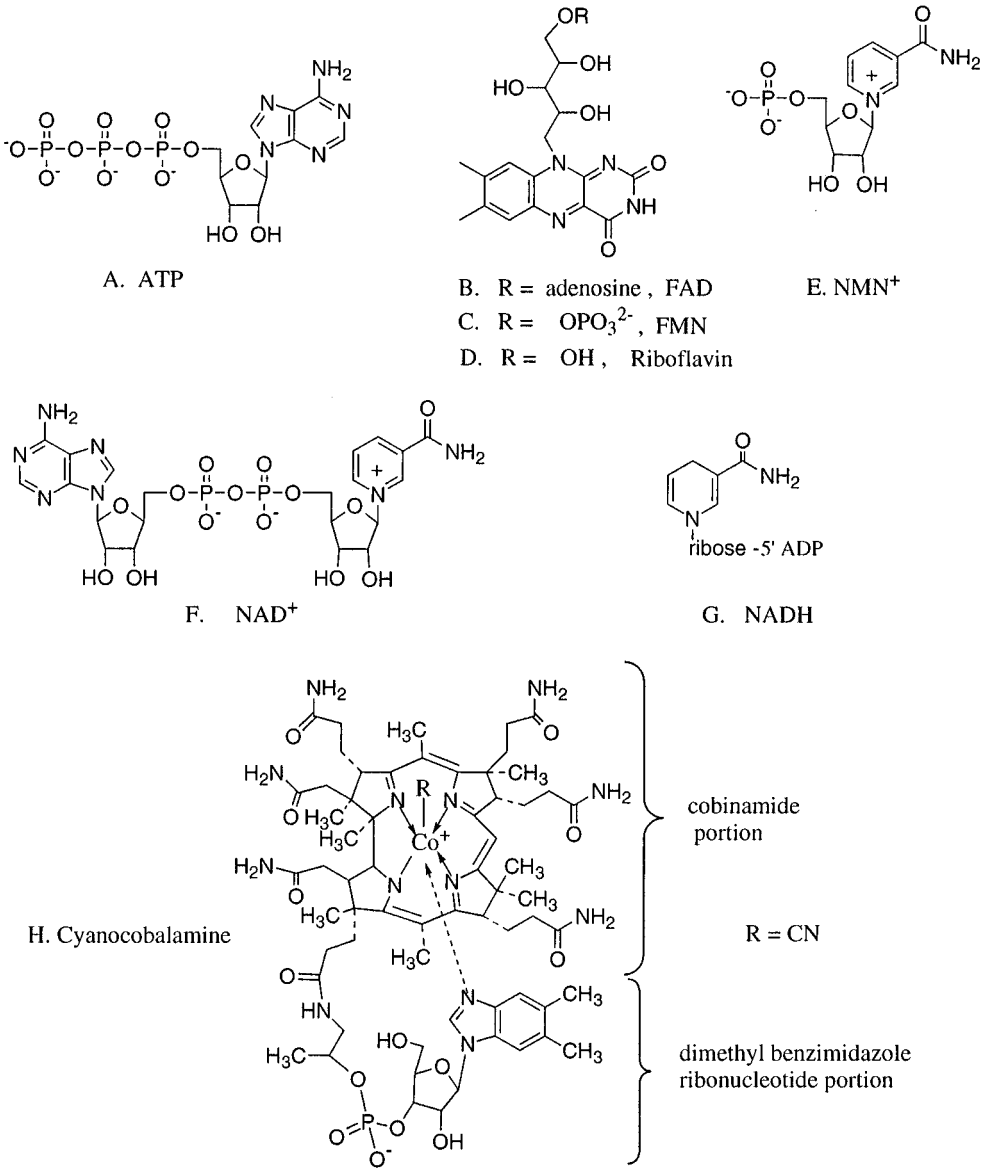


Figure 17. Structure of ribose-containing cofactors: (A) adenosine triphosphate (ATP), (B) flavin adenine dinucleotide (oxidized form) (FAD), (C) flavin mononucleotide (oxidized form) (FMN), (D) riboflavin, (E) nicotinamide mononucleotide (NMN^+), (F) nicotinamide adenine dinucleotide (oxidized form) (NAD^+), (G) nicotinamide adenine dinucleotide (reduced form) (NADH), and (H) cyanocobalamine (vitamin B12).

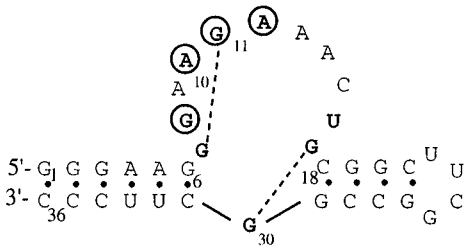


Figure 18. Sequence of the 36-nucleotide AMP-binding RNA. Conserved nucleotides are shown in bold and those that directly interact with the ligand are circled. Dotted lines represent non-Watson–Crick base pairs.[78–82]

spectroscopy (reviewed recently by Feigon *et al.*[79]).[80–82] The two RNAs examined were virtually identical in terms of length and sequence within the internal loop, and both form the same binding pocket for the cofactor by employing the invariant loop residues. The key features of both tertiary structures are two asymmetric G–G base pairs. These mismatches close the loops and help form a hairpin turn that strongly resembles a U-turn from tRNA[18] or a GNRA (where N is any nucleotide and R is a purine residue) tetraloop.[83] In the later, the fourth residue is the AMP ligand. The two orthogonal Watson–Crick stems are each capped with a G–G mismatch pair (G7–G11 and G17–G30), and a G–AMP base pair is formed at the junction between the helices. The stacking of the neighboring purine rings (G8, A9, and A10) also helps stabilize the AMP intercalation site. These high-resolution structures have provided researchers with many new insights into small molecule RNA recognition, as well as confirmed prior findings: (1) binding of the ligand occurs at a site of noncanonical base pairing; (2) the RNA undergoes a significant conformational change upon AMP binding; (3) the binding of the ligand (AMP) is stabilized by hydrogen-bonding and neighboring stacking interactions; and (4) the RNA contacts nearly half of the total accessible surface area of the ligand, thus deeply integrating the ligand into an extremely stable tertiary fold.

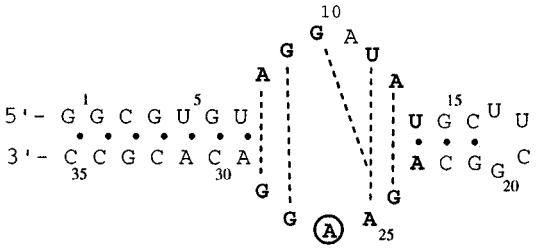


Figure 19. Sequence and secondary structure of the FMN-binding 35-nucleotide RNA aptamer. The conserved nucleotides are shown in bold. The dotted line represents nonstandard base pairs that were determined by NMR studies. The A residue that makes a direct hydrogen-bonding contact with FMN is circled.[87]

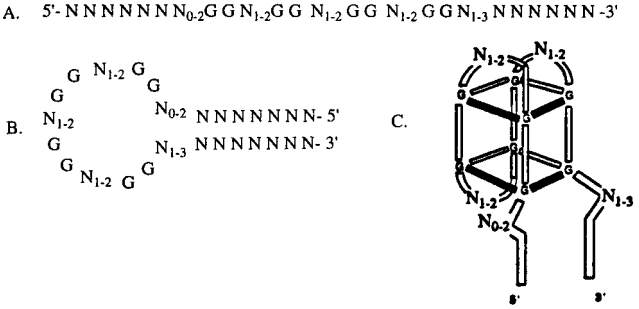


Figure 20. Consensus sequence (A) and proposed secondary structure (B) of riboflavin-binding RNAs. (C) The proposed tertiary structure of the riboflavin aptamer, a G-quartet, is shown.[86]

Similarly, several groups have used *in vitro* selection methods to identify RNAs that bind specifically to cyanocobalamin (vitamin B12), flavin adenine dinucleotide or mononucleotide (FMN or FAD, respectively), and nicotinamide mononucleotide, adenine dinucleotide, or adenine dinucleotide phosphate (NMN^+, NAD^+, or $NADP^+$, respectively).[84−86] The structures of these ligand cofactors are shown in Figure 17. The 35-nucleotide FMN-binding RNA (Figure 19) contains an asymmetric internal loop (6:5) which is conserved among all isolated FMN-binding aptamers.[84] Furthermore, the K_d's for FMN-binding RNAs were ~0.5 μM for FMN, 7,8-dimethylalloxazine (an FMN fragment), and FAD. Thus, a small RNA aptamer has relatively high affinity for a specific ligand and is able to discriminate among the closely related structures.

Recently, Fan *et al.*[87] reported the NMR structure of the FMN−RNA complex, the first high-resolution structure of a selected RNA aptamer−ligand complex. As with other RNA-binding ligands, the specific binding interactions lead to a distinct conformational change of the RNA, in this case involving a purine-rich loop. The bound cofactor essentially locks the RNA into a specific compact structure, which is apparent by the narrowing of the NMR resonances. As would later be apparent with other ligand−RNA interactions, Fan and co-workers showed that non-canonical base pairs are critical for stabilizing the folded tertiary structure and for maintaining the RNA−ligand specificity. In particular, FMN inserts between the base pairs by an intercalative binding mode and the uracil-like edge of the ligand forms a Hoogsteen-type base pair with a conserved adenosine (A26). In addition, a G10−U12−A25 triple-base platform stacks below the FMN−A pair and a G9−G27 mismatch pair sits above, both of which provide additional stabilizing interactions. The remaining bases participate in two G−A mismatches (G28−A8 and G24−A13), with A11 looped out of the major groove. Thus, there are several general comments that can be made about the structure. First, purine-rich loops and non-Watson−Crick base pairs (mismatches and triple-base interactions) are used to stabilize the ligand interaction. Second, the use of unique base-stacking and hydrogen-bonding interactions with the ligand confers stability and specificity in the resulting complex. The otherwise floppy RNA motif is stabilized in the presence of the ligand. It remains to be seen if these are general features of all ligand−RNA interactions, although they have been found in several of the high-resolution structures examined to date.

FAD-binding RNAs were also isolated with a 13-nucleotide consensus sequence that is proposed to form a purine-rich loop (5-PuAAAGGAAGUGUA-3′) within a stem-loop secondary structure.[84] The *in vitro*-selected RNAs are known to recognize the flavin portion of FAD (K_d = 23 μM) rather than the adenosine portion ($K_d^{ATP} \geq$ 4.0 mM). In contrast, NAD^+-binding RNAs[84] closely resemble the ATP-binding RNAs identified by Sassanfar and Szostak.[78] The consensus sequence 5′-GGAAGAAACUG-3′ is found in the proposed asymmetric internal loop (10:1), and the base opposite the large loop is a G. Thus, two completely independent selections for structurally related ligands gave similar RNAs. It still remains to be determined if this second RNA motif forms a similar structure with bound NAD^+ as with AMP.[80−82]

In a related study, Lauhon and Szostak selected for RNAs with specificity for riboflavin and nicotinamide mononucleotide (NMN^+).[86] In this case, a structural motif was identified that contains intramolecular G-quartets with high affinity for oxidized riboflavin (Figure 20) ($K_d \approx 1-5$ μM). These aptamers do not reveal any discrimination between oxidized and reduced forms of the ligand. Preliminary structural data suggest a stacking interaction between the flavin and the guanine quartets, as well as a stabilization of the unusual quartet structure. In contrast, RNAs selected to bind to the nicotinamide portion (NMN) of NAD^+ show a 10-fold enhanced binding of NAD^+ over NADH. The major binding determinant in this case could be the positive charge of the nicotinamide ring. Indeed, experiments with analogues such as *N*-methylnicotinamide have indicated that the positively charged nitrogen is an important recognition element. Together these results demonstrate that RNA can bind to cofactors with micromolar affinities and in some cases discriminate between the redox states of the cofactor. These selected RNAs also employ tertiary structures, such as G quartets and internal loops, to bind their given ligands.

Recently, Burgstaller and Famulok discovered the ability of the isoalloxazine moiety of FMN to recognize and induce strand scission of G−U mismatch pairs.[88] The cleavage to the 3′ side of the uracil residue is dependent upon light, the presence of metal ions (*e.g.*, Mg^{2+}, Ca^{2+}, Sr^{2+}, Ba^{2+}, Zn^{2+}, and

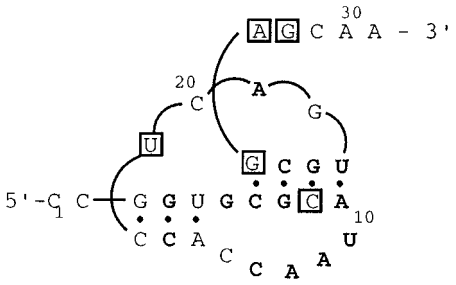


Figure 21. The consensus sequence and proposed secondary structure of the 35-nucleotide cyanocobalamine aptamer.[85] The absolutely invariant bases are shown in bold and the bases that vary only once are boxed.

Cd^{2+}), and appears to be oxidative in nature. The isoalloxazine moiety represents one of several small molecules with the ability to recognize a nonstandard base pair in RNA (see section IV.B.1). Such molecules may provide insights as to how proteins or other potential drugs might target such sites.

Receptors for the cofactor cyanocobalamine (vitamin B12) have also been isolated.[85] The major binding sequence exhibits a K_d of ~320 nM for the ligand and binds to the related cobinamide dicyanide with a K_d of 8.8 μM. Tighter binding to the cyanocobalamin is evidence that contacts exist between the aptamer and the dimethylbenzimidazole ribonucleotide portion of the ligand as well as the cobinamide portion. The receptor did not bind to adenosylcobalamin (coenzyme B12) (Figure 17H, where R = adenosyl), suggesting the importance of the axial cyanide group in binding to RNA. The cyanide may contact the RNA directly, or the face of the cyanocobalamin has to be accessible for interactions with the aptamer. Further selection experiments afforded a highly conserved 31-base region with 14 absolutely invariant bases and five that vary only once (Figure 21). A 35-nucleotide RNA representing this sequence binds to cyanocobalamin with enhanced affinity (K_d = 88 nM) and increased specificity (*i.e.*, decreased affinity for cobinamide dicyanide (K_d = 20 μM)). A pseudoknot structure, as shown schematically in Figure 21, has been proposed, and chemical modification experiments indicated that the RNA undergoes a conformational change upon ligand binding which is further stabilized by LiCl.

Overall, cofactors are able to bind with high affinity and specificity to RNA, and come close to matching the binding to protein enzymes. The ultimate goal is to develop RNAs that contain these cofactor binding motifs in conjunction with catalytic abilities. Further structural characterization of the aptamer–ligand complexes should aid in the design of such cofactor-dependent catalytic RNAs, or ribozymes.

6. Aminoglycoside Antibiotics

Antibiotics interact with several known biological targets to affect specific cellular processes involving RNA. Ribosomal RNA was known to be one of these targets since the 1960s[89] and some of the binding sites have been mapped on 16S and 23S rRNAs.[90] For further details regarding the earlier studies, the reader can refer to several reviews.[91–93] The focus of this review will be aminoglycoside antibiotics, for which the most structural data are available. These RNA-binding ligands are composed of amino sugars linked to a deoxystreptamine ring. The functionalities on these sugars are generally amino and hydroxyl groups.

Streptomycin (Figure 22A), neomycin (Figure 22B), and closely related analogues are of particular interest. These aminoglycoside antibiotics bind to unique sites on the ribosome[90,94] and interfere with RNA functions. Similarly, the group I intron self-splicing reaction is inhibited by binding of these drugs at or near the guanosine binding site.[95–99] von Ahsen and co-workers have suggested that binding of certain aminoglycosides leads to a disruption of the structural contacts within the ribozyme that are responsible for splicing. Aminoglycoside antibiotics, specifically neomycin, can also inhibit the hammerhead ribozyme cleavage reaction.[100] Some of the characteristics of neomycin that allow it to bind favorably to RNA are conformational rigidity and a polycationic nature.[101,102] Although there are no obvious common sequences between 16S rRNA, 23S rRNA, group I introns, and the hammerhead ribozyme, they might contain common tertiary structures that allow a similar recognition mechanism by the aminoglycoside ligands.

Neomycin also inhibits binding of the HIV Rev protein to its viral RNA recognition element known as RRE[103] and similarly can inhibit a Tat peptide–TAR RNA interaction.[104] Neomycin can bind to the TAR RNA (Figure 11) in the absence of peptide, but these interactions appear to be governed by the duplex and to some extent the loop regions, rather than the bulge nucleotides which are important for Tat binding.[24,104–106] Similarly, neomycin binds to RRE in the absence of Rev, and chemical modification studies indicate that the protein and drug have a similar binding site. Neomycin interacts with a 67-nucleotide RNA at an asymmetric purine-rich internal loop (6:4) involving a subset of sites bound by Rev (Figure 23A). Evidence suggests that an RNA conformational change occurs upon binding of the drug and that the irregular RNA structures (*e.g.*, base mismatches) are necessary for recognition by neomycin. The NH_2 substituents of the aminoglycoside are also important for specificity. For example, paromomycin (Figure 22C), which differs from neomycin by a single amino to hydroxyl substitution, shows a 100-fold lower affinity for RRE. Thus, it appears that neomycin and related antibiotics have more than one possible binding mode to RNA, although each of these modes may be highly specific within a single RNA species.

More recently, Park and co-workers have taken a combinatorial approach to prepare a library of neomycin B mimetics and screen for RRE binding.[107] Several members of this library were shown to have better binding activities than neamine, which alone interacts only weakly with RRE. This combinatorial synthesis and screening approach has the exciting possibility of quickly leading to new drugs that target specific RNA sequences or structures, such as those found in viral RNAs. Additionally, this method can be used to find aminoglycoside antibiotics with improved efficiency if applied to such systems as those described below.

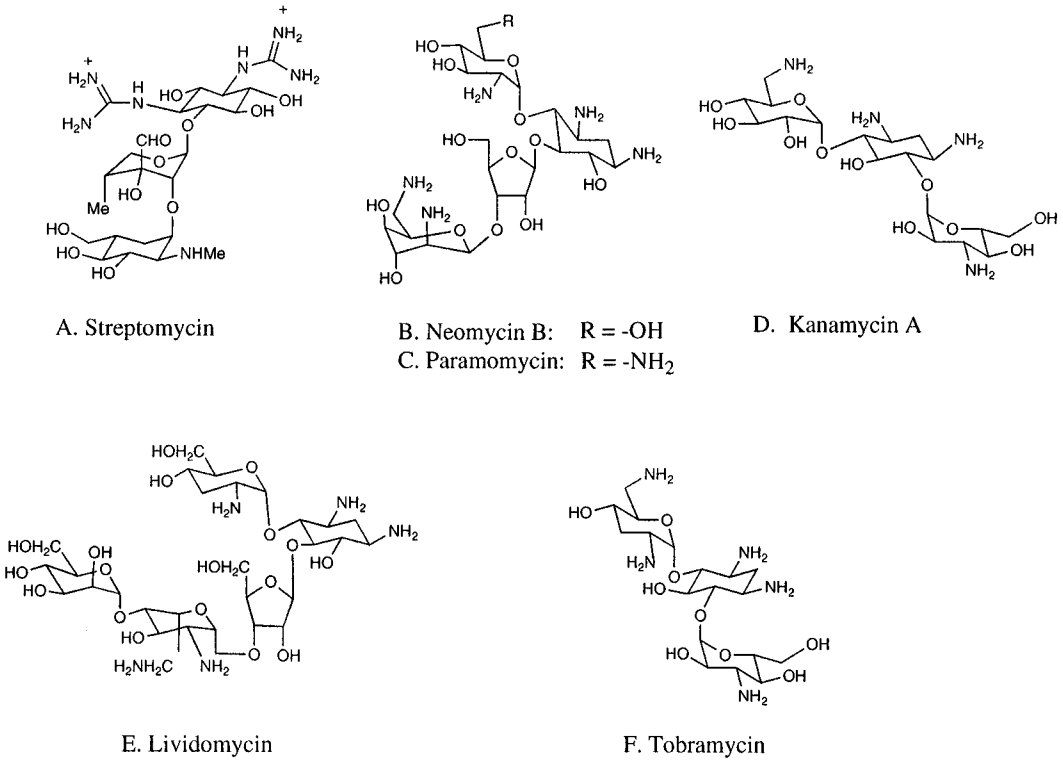


Figure 22. Chemical structures of some representative aminoglycoside antibiotics that bind to RNA are shown.

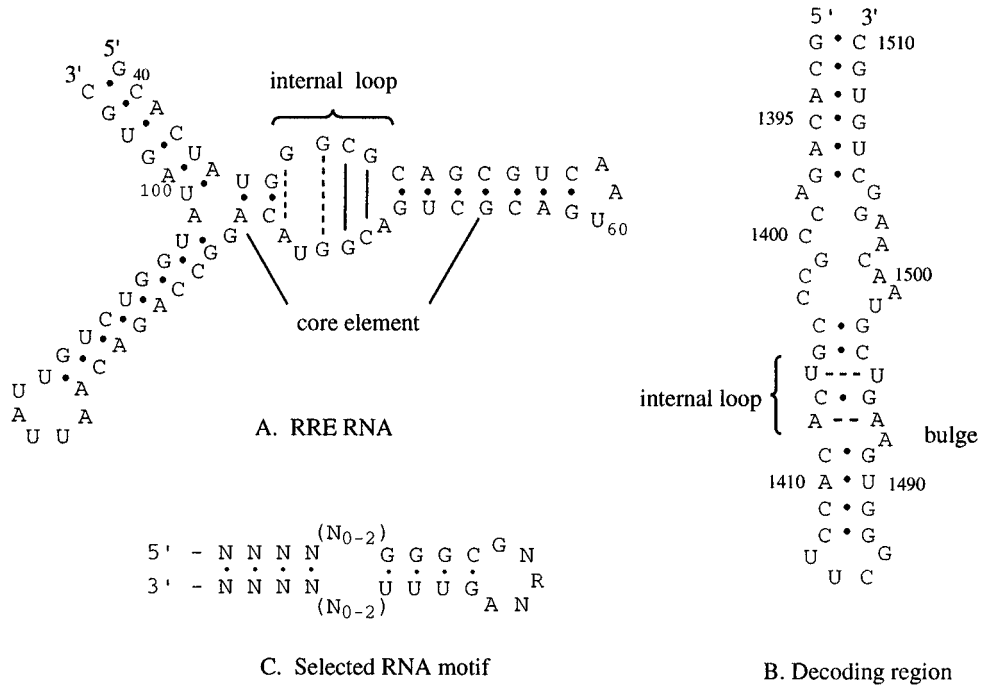


Figure 23. Secondary structure motifs that are recognized by neomycin: (A) A 67-nucleotide RNA (RRE) containing the Rev-binding site (core element) is shown.[103] The base pairs that were identified by using NMR and a minimal Rev−RRE complex are shown (noncanonical base pairs are shown with dotted lines and the Watson−Crick base pairs are indicated by solid lines).[150,151] (B) A 49-nucleotide fragment representing the decoding region of *E. coli* 16S rRNA binds to neomycin, paromomycin, and related aminoglycoside antibiotics. Noncanonical base pairs are indicated with dotted lines. NMR studies were performed on a 27-nucleotide RNA (from residues 1404 to 1499, where the last two base pairs in the helix were G−C).[108,112] (C) The consensus neomycin-binding motif, where N is any nucleotide and R a purine, was obtained from *in vitro* selection experiments.[113]

In 1994, Purohit and Stern demonstrated that certain small rRNA domains can fold and function outside the context of the ribosomal RNA and proteins.[108] In particular, the decoding region of *Escherichia coli* 16S rRNA, a 49-nucleotide domain (shown in Figure 23B) located near the 3′ end of the full-length ribosomal RNA, can interact with aminoglycoside antibiotics. This A-site subdomain of the decoding region RNA is able to bind neomycin and the closely related paromomycin. The patterns of chemical modification indicate that the drug binds similarly to the model RNA as to the full-length 16S rRNA and 30S ribosomal subunit. Mutagenesis analysis in conjunction with chemical probing experi-

ments have also been used to further define the binding site residues.[109] In addition, a longer RNA fragment containing the decoding region has been shown by fluorescence assays to bind to kanamycin B, paromomycin, tobramycin, and gentamycin C with K_d's ~1−2 μM and to neomycin B with a K_d of 132 nM.[110]

More recently, Recht, Fourmy and co-workers developed a high-resolution model for the paromomycin−16S rRNA interaction by using NMR spectroscopy and a 27-nucleotide fragment.[111,112] The antibiotic binding site is an asymmetric loop (3:4) containing noncanonical base pairs. The specific interactions with paromomycin occur in these regions of unusual tertiary structure. The two noncanonical base pairs that have been identified are U1406−U1495 and A1408−A1493, leaving a bulged A1492 residue. The antibiotic binds in the major groove of this region and induces a conformational change in the RNA. The determinants for recognition appear to be A1408, A1493, U1495, and the G−C base pair which stacks between the U−U and A−A mismatch pairs. The bulged A residue creates asymmetry within the loop region and is also important for recognition by paromomycin. The loop region is dynamic in the absence of drug, but A1408, A1492, and A1493 still stack within the helix.

Several key features are evident from the NMR studies with the 16S rRNA analogue and paromomycin. First, the noncanonical base pairs are important for recognition by the ligand. Second, a specific conformation in the RNA is stabilized upon binding of the antibiotic and involves the mismatched base pairs and bulged residue. The drug itself adopts an L-shape conformation and binds within the major groove of the RNA in the well-defined A-bulge pocket. Specific hydrogen-bonding contacts stabilize the complex. Third, the specific interactions with the A1408 residue are significant with regard to organism selectivity. Prokaryotes are more susceptible to the aminoglycoside antibiotics than eukaryotes which have a G substituted at this position. For this site, it appears that the A1408−A1493 pair is essential to form the proper binding pocket for the drug. In addition, A1408 is methylated in resistant strains. Thus, the structure provides an explanation for specific binding of aminoglycosides to rRNA, specificity for certain organisms, and means of resistance to this class of drugs.[112] Given the relatively simple structure of the decoding analogue, it should be possible to construct other small model RNAs and study their interactions with small ligands or drugs.

Neomycin has overlapping sites with paromomycin in the decoding region. This result is in contrast to the studies with RRE in which there was a differential binding between the closely related antibiotics. Thus, the two studies suggest that neomycin and paromomycin have more than one binding mode or RNA binding site, which is consistent with the previously mentioned RNA−ligand studies. *In vitro* selection experiments have also been used to understand the sequence constraints of the decoding region of 16S rRNA in binding to neomycin.[113] The 47-nucleotide domain of the decoding region of 16S rRNA was mutated at 30% per base position and selected for binding to neomycin. Perhaps surprisingly, the isolated clones do not have any sequence homology to the starting sequence. Unlike the starting structure, the new sequences can fold into a hairpin structure. This selected RNA sequence more closely resembles RNA from previous *in vitro* selection experiments (Figure 23C).[114] The conserved 13-nucleotide sequence can be folded into a stem−loop structure with a GNRNA loop sequence (where N is any nucleotide and R is a purine), a variable-stem sequence, and a conserved three base-pair stretch of G−U wobble pairs. The change in specificity can be explained from the affinity data. The selected RNAs exhibit tighter neomycin binding than the starting 16S rRNA analogue. Thus, under the conditions of the selection experiment, a much tighter binding species is identified. Overall, these results indicate that a highly optimized RNA sequence for neomycin binding can be isolated; whether this sequence occurs naturally and is a biologically relevant target site remains to be determined. Although the original 16S rRNA motif may not be the optimal target for neomycin, perhaps it is recognized in the absence of a higher affinity site. Alternatively, other factors may be necessary for enhanced binding at this particular site *in vivo*.

Although there appears to be variability in the neomycin-binding sequences and secondary structures, all of the binding sites appear to involve a "loosening" of the major groove, either by bulged nucleotides, mismatched base pairs, or loop residues. No primary sequence homologies exist between all of the neomycin-binding RNAs, suggesting that the recognition process has a structural basis. These structures might be considered as the "minimal requirements" for binding, while other factors will affect specificity in binding, such as the ability to discriminate between the closely related neomycin B and paromomycin.[114] Lato and co-workers made a similar conclusion by using *in vitro* selection to determine the diversity of kanamycin A and lividomycin-binding sites that can be generated by RNA sequences.[115] The structures of kanamycin A and lividomycin are shown in Figure 22, parts D and E. These results demonstrate that there are many ways to fold RNA into high-affinity, specific aminoglycoside-antibiotic binding sites. Examination of the selected RNA sequences and tertiary structures may allow the identification of potential target sites in natural RNAs that have not been previously utilized. Such an application would require a comparison of the selected RNA sequences with databases of natural RNA sequences.

Wang and Rando reported that tobramycin-binding RNAs can be isolated by using *in vitro* selection techniques.[116] Tight-binding RNAs with predicted stem−loop structures have been identified. A single RNA exhibits strong binding to tobramycin (Figure 22F) (K_d = 0.77 nM) and a high level of discrimination, binding to structurally related aminoglycosides with 10^3 to 10^4 lower affinity.[117] As a result of the tight binding and high level of specificity, this RNA−tobramycin complex has proven to be useful for determining some additional rules that govern specific aminoglycoside interactions at particular RNA

binding sites. Specifically, Jiang and co-workers have studied a tobramycin−RNA aptamer complex by NMR spectroscopy.[118] A mismatch (U−U) at an RNA hairpin stem−loop junction and a neighboring bulged A base appear to be critical for major groove opening and tobramycin interactions. A single C residue in the loop is important for the partial encapsulation of the drug at its binding site. Furthermore, a three-residue turn in the RNA hairpin showed a strong similarity in structure to the yeast $tRNA^{Phe}$ anticodon and TΨC hairpin loops, and an even stronger resemblance to a conserved rRNA hairpin loop. Thus, these results provide the first evidence that recurring RNA structures may also serve as common drug-binding motifs.

7. Antitumor Antibiotics

The enediyne and bleomycin (BLMs) antitumor antibiotics deserve special attention. Although they are classically known as DNA binders and cleaving agents, several preliminary studies have indicated RNA as a viable target for these drugs. Members of both classes have the ability to recognize specific RNA sequences or structures and induce strand scission. Oxidative cleavage of the RNA by the metal-coordinated Fe^{II}-BLM will be discussed later. Bleomycin A2 (Figure 24A) was recently found to promote sequence-specific cleavage of yeast $tRNA^{Phe}$ through phosphodiester hydrolysis in the absence of transition-metal ions.[119] The facile hydrolysis reaction occurs to the 3′ side of the pyrimidine at pyrimidine−purine sites. One suggested mechanism is that BLM binds to RNA and induces a conformational change so that the reactive sites become exposed to external cleaving agents such as Mg^{2+} or to a site within BLM itself.

Similarly, neocarzinostatin (NCS) (Figure 24B), an enediyne-containing drug, is able to induce highly efficient, site-specific cleavage at the bulge site of TAR RNA.[120] Cleavage occurs in the absence of a thiol activator, and the thiol-generated NCS intermediate that efficiently cleaves DNA is not recognized by RNA. Thus, the structurally different intermediates appear to have differential recognition of the RNA and DNA targets. In addition, the RNA structure, rather than sequence, determines the extent of cleavage, although in this case it is actually a poorer substrate than the DNA analogue. It is possible that highly specific sites for NCS binding and cleavage of RNA have yet to be identified. A second group of investigators has shown that NCS can recognize a variety of RNA structures, such as two sites within the precursor $tRNA^{His}$, a 14-nucleotide hairpin RNA, and a proposed RNA pseudoknot.[121] The related esperamicin (ESP) and calicheamicin (CAL) (Figure 24C,D) also exhibit specificity for RNA hairpins, but require added thiol for RNA cleavage to occur. It is not known at this time whether specific structures of the drugs are required in order to bind and/or cleave at specific RNA sites. Clearly, the enediyne−RNA and BLM−RNA binding interactions and cleavage mechanisms require further investiga-

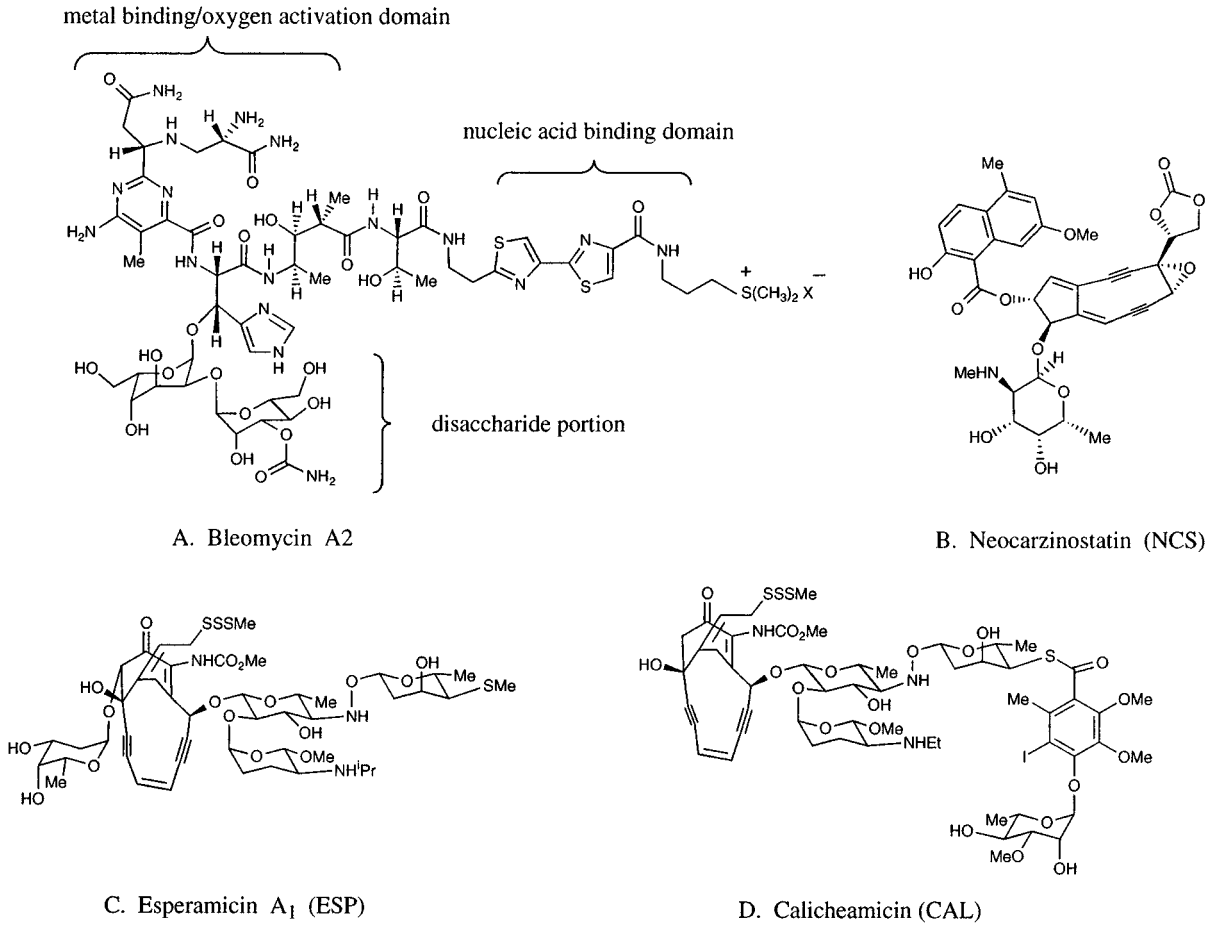


Figure 24. Chemical structures of antitumor antibiotics.

tions in order to better understand the recognition processes and to consider any therapeutic relevance of these interesting natural organic ligands.

8. Alkaloids

Another example of RNAs that exhibit a high level of molecular discrimination comes from an *in vitro* selection experiment with theophylline (Figure 25A). This naturally occurring alkaloid is an important bronchodilator. Jenison *et al.*[122] used an elegant "counter-SELEX" method and RNAs starting with 40 random nucleotides to find aptamers with very high affinity (K_d = 100 nM) for the drug. In addition to tight binding, the selected RNAs also display a 10000-fold lower affinity for the structurally related caffeine molecule (Figure 25B) even though the two alkaloids differ only by a single methyl group at the N-7 position. This high level of discrimination approaches that displayed by protein enzymes. The consensus RNA sequence (Figure 25C) contains a CCU bulge on one side of a three base-pair stem and a highly conserved six-nucleotide symmetric (3:3) internal loop. By using theophylline analogues, Jenison and co-workers concluded that the N-7 hydrogen was crucial for the discrimination, and likely forms a specific hydrogen bond within the RNA binding pocket. NMR spectroscopy reveals that a single RNA aptamer undergoes a significant conformational change and an alteration of base pairing upon binding of the ligand. These results are particularly important in that they support the possibility of using RNAs as diagnostic tools. One such application would be the monitoring of theophylline levels in serum in the presence of drug analogues such as caffeine or theobromine.[122]

In summary, recent information on organic ligand–RNA interactions has provided us with a better understanding of these molecular complexes. There are several recurring themes from these studies. First, the RNAs use nonstandard base pairs, particularly purine–purine (G–G, A–A, and G–A) pairs, within single-stranded loop or bulge regions. In general, such interactions serve to open the otherwise narrow major groove and make it more accessible to small molecules. Second, alterations in RNA conformation occur upon ligand binding, while the use of hydrogen bonding and stacking interactions confer specificity. These structural changes combined with specific interactions are analogous to the induced-fit model for substrates binding to their corresponding protein enzymes. Third, some RNAs might use common tertiary structure motifs to bind small molecules (*e.g.*, the GNRA U-turn motif). Lastly, the ability of several of these organic ligands to inhibit RNA function is significant with respect to drug discovery efforts.

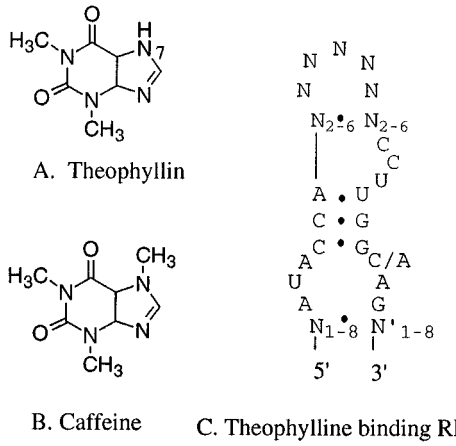


Figure 25. Structures of alkaloids theophylline (A) and caffeine (B), and (C) the proposed secondary structure of a theophylline-binding RNA. N is any nucleotide and N′ implies the complementary nucleotide.

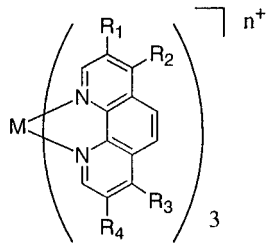
B. Inorganic Ligands

The advantage of inorganic complexes is their relative ease of synthesis and the ability to alter the ligands or metal, in turn altering the shape, RNA-recognition properties, and reactivity of the complex. Several general comments can be made regarding the inorganic complexes that have been studied for their RNA-binding abilities: they are usually coordinatively saturated, inert to substitution, and rigid and well defined in structure, and they cannot make direct contacts by coordination to RNA. These metal complexes typically bind by noncovalent interactions such as van der Waals, electrostatic, or hydrogen-bonding contacts. In addition, metal complexes often have the ability to induce RNA strand scission by a number of mechanistic pathways, thus marking their binding sites. It should be noted, however, that not all bound species are equally efficient at RNA cleavage, therefore cleavage does not equal binding in all cases.

1. Rhodium and Ruthenium Polypyridals

Initial studies with tris(1,10-phenanthroline)ruthenium(II) [$Ru(phen)_3^{2+}$] and tris(3,4,7,8-tetramethylphenanthroline)ruthenium(II) [$Ru(TMP)_3^{2+}$] (Figure 26A,B) demonstrated that these complexes can bind to and promote strand scission of RNA in both a specific and a nonspecific manner.[123] Reactions with a racemic mixture of $Ru(phen)_3^{2+}$ and the structurally well-characterized yeast $tRNA^{Phe}$ lead to nonspecific modification of guanine residues. The complex also promotes cleavage at the TΨC-loop residues T54 and Ψ55, thus indicating a close association, or specificity, for this region of the tRNA relative to other sites. $Ru(TMP)_3^{2+}$ also shows a preference for cleavage at G residues on tRNA, but only reacts at a subset of the $Ru(phen)_3^{2+}$ sites and prefers binding at helical regions of the RNA. Overall, these results demonstrate the potential for using ruthenium polypyridal complexes as agents for specific binding to RNA.

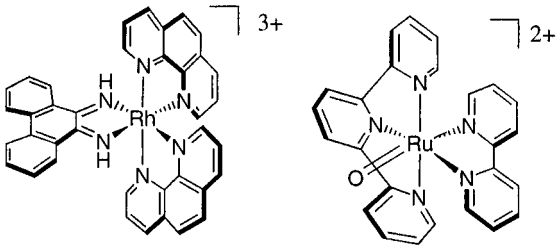
Studies with tris(4,7-diphenyl-1,10-phenanthroline)rhodium(III) [$Rh(DIP)_3^{3+}$] (Figure 26C) showed that the addition of the phenyl group substituents to the phenanthroline ring adds specificity in RNA binding. In addition, the substitution of rhodium for ruthenium produces a complex capable of direct strand scission rather than cleaving by a diffusible species. $Rh(DIP)_3^{3+}$ recognizes two major sites on yeast $tRNA^{Phe}$.[123] The first site (Ψ55) resides in a pocket between two loop regions of the tRNA and resembles the cleavage site on DNA cruciforms. The second site occurs at the 3′ side of a wobble-paired U



A. M = Ru(II), n=2, R_1=R_2=R_3=R_4=H, $Ru(phen)_3^{2+}$

B. M = Ru(II), n=2, R_1=R_2=R_3=R_4=CH_3, $Ru(TMP)_3^{2+}$

C. M = Rh(III), n=3, R_1=R_4=H, R_2=R_3=Ph, $Rh(DIP)_3^{3+}$



D. $Rh(phen)_2phi^{3+}$ E. $Ru(tpy)(bpy)O^{2+}$

Figure 26. Structures of rhodium and ruthenium polypyridal complexes.

residue (G−U mismatch) within a double-helical region of RNA. The specificity for the G−U mismatch is conserved in yeast $tRNA^{Asp}$, as well as synthetic RNA "microhelices" and 5S rRNAs from *Xenopus* oocytes and *E. coli*.[124] This high level of discrimination is attributed to shape-selective binding by the metal complex and unique stacking interactions on one side of the G−U mismatch. Furthermore, the complex is sensitive to the flanking sequences surrounding the G−U mismatch and shows differential reactivity with different G−U containing RNAs. This interaction may serve as a model for protein−RNA interactions because the G−U mismatch has been identified as a determinant for several RNA-binding proteins, such as *E. coli* $tRNA^{Ala}$ aminoacyl synthetase. In addition, it may be possible to inhibit specific RNA−protein interactions at G−U sites with these complexes.

The rhodium complex, bis(phenanthroline)(9,10-phenanthrenequinone diimine)rhodium(III) [$Rh(phen)_2phi^{3+}$] (Figure 26D), is able to bind and cleave RNA at accessible major groove sites.[123] In particular, the complex is able to recognize triple-base sites and certain stem−loop junctions in tRNA.[125] In addition, the rhodium complex is able to recognize regions of *Xenopus* oocyte 5S rRNA that are critical for binding by the transcription factor IIIA (TFIIIA) or L5 protein.[126,127] The major sites of cleavage by $Rh(phen)_2phi^{3+}$ are clustered in the loop E and helix III-loop C hairpin regions. The conformation of loop E, an asymmetric internal loop (4:5), has been determined by NMR spectroscopy and contains a mismatched base pair (G−A), a reverse-Hoogsteen A−U pair, and a putative triple-base interaction.[128] This unusual tertiary structure is apparently a recognition element for the metal complex, as well as an important determinant of TFIIIA binding to

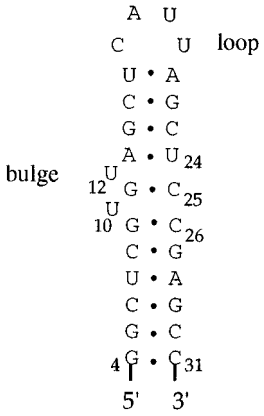


Figure 27. The sequence and proposed secondary structure of a BIV TAR RNA fragment.

5S rRNA. Similarly, the helix III-loop C hairpin site is a recognition site for the ribosomal protein L5.[127] The possible recognition elements are bulged adenosine residues, a stem−loop junction, and a putative tetraloop structure. Overall, it appears that the open or accessible sites that are recognized by $Rh(phen)_2phi^{3+}$ are also important for protein binding in the major groove of the RNA. As with $Rh(DIP)_3^{3+}$, this complex may be useful for inhibiting protein binding at specific sites.

$Rh(phen)_2phi^{3+}$ was found to cleave the HIV TAR RNA at a site (C39 and U40) opposite the trinucleotide bulge site, whereas no cleavage is apparent on bulgeless or one-nucleotide bulge TAR RNAs. In addition, Tat inhibits cleavage by the metal complex, suggesting that the two species have overlapping binding sites.[129] In contrast, the cleavage efficiency by $Rh(phen)_2phi^{3+}$ on HIV TAR RNA is unaffected by the presence of arginine. Therefore, either the complex binding is not inhibited by arginine binding, or it competes effectively with arginine for the same binding site.

The Δ isomer of $Rh(phen)_2phi^{3+}$ binds with an affinity of 2×10^6 M^{-1} at a proposed triple-base site of BIV TAR RNA (cleavage occurs at U24 of the putative U10−A13−U24 base triple).[130] The sequence of the proposed secondary structure of TAR RNA fragment from BIV is shown in Figure 27. In this case, the complex is able to inhibit binding by the BIV Tat peptide. NMR studies have demonstrated that the base-triple interaction is induced upon binding of the Tat peptide to the bulge site.[131] Since the interaction of the metal complex with the TAR RNA does not require prior binding of the Tat peptide, it has been suggested by Lim and Barton that the metal complex induces a similar change in the RNA conformation in order to bind with such high specificity. Furthermore, enantioselectivity for the Δ isomer only occurs with the BIV TAR RNA, and not with HIV TAR RNA, suggesting that the BIV RNA single-base bulge site is more restrictive than the HIV RNA trinucleotide-bulge site. Mutation of the A13−U24 base pair to a U13−A24 pair in the BIV TAR RNA abolishes binding by $Rh(phen)_2phi^{3+}$, thus demonstrating the significance of the potential base triple formation for recognition by the metal complex.

$Rh(phen)_2phi^{3+}$ also exhibits specificity for another RNA hairpin, namely the iron regulatory element

(IRE) from a region of mRNA encoding ferritin, and a base-paired flanking region (FL). Recognition occurs at residue U35 in the FL region. This site contains unpaired or mispaired bases which presumably causes a distortion of the duplex region and allow access to the bulky metal complex.[132] Interestingly, this site is also functionally significant and resides adjacent to a CUC/GAG conserved triplet that modulates binding by the regulatory protein IRP. Mutations at the triplet site that affect $Rh(phen)_2phi^{3+}$ recognition also cause changes in the ability of the RNA to regulate translation. Another related complex, $Ru(tpy)(bpy)O^{2+}$ (tpy = 2,2′,2″-terpyridine, bpy = 2,2′-bipyridine) (Figure 26E), has the ability to recognize a few sites on RNA in a highly selective manner, presumably because of its size, shape, charge, and unique thermal cleavage properties.[132] A major site on IRE+FL is a $CAG_{14}UGU$ hexaloop in which G14 is cleaved specifically. No other reagents examined to date have reactivity or known binding activity at this site. Thus, the ruthenium complex is able to identify uniquely a specific structure within the IRE hairpin. Again, this site has functional significance and is important for binding to the regulatory protein IRP.

On the basis of these studies, $Rh(phen)_2phi^{3+}$, $Rh(DIP)_3^{3+}$, $Ru(tpy)(bpy)O^{2+}$, and related complexes should be helpful in identifying structures in ribozymes, viral RNAs, rRNAs, and mRNAs with potential functional significance. Modifications of the ligands may be useful in the design of even more specific complexes that may serve as therapeutic agents. Further studies are necessary in order to identify the high-resolution structures of these metal complexes bound specifically to RNA. In addition, the need to modify these inorganic ligands to minimize their DNA-binding and DNA-cleaving potential is apparent. Such studies should aid in the development of an extremely useful class of RNA-binding molecules. They might also be used as specific RNA inactivators by targeting and cleaving undesirable RNAs, such as viral mRNAs.

2. Nickel Complexes

Chen *et al.*[133,134] have demonstrated that the inorganic complex (2,12-dimethyl-3,7,11,17-tetraazabicyclo-[11.3.1]heptadeca-1(17),2,11,13,15-pentaenato)nickel-(II) perchlorate [NiCR] (Figure 28A) is an effective agent for oxidation of guanine residues of RNA in the presence of an oxidant such as potassium monopersulfate. The complex exhibits preferential reactivity at accessible guanine N-7 sites, such as those found in unpaired regions (*e.g.*, mismatches, bulges, or hairpin loops). Thus, NiCR is a sensitive probe for

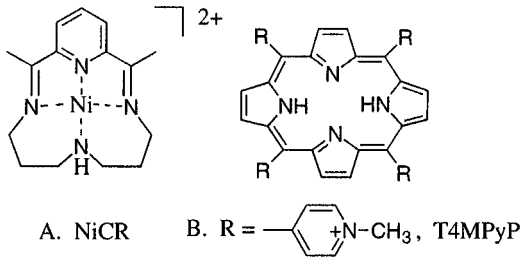


Figure 28. Structures of inorganic complexes (A) NiCR and (B) a porphyrin cation, T4MPyP.

the folded structure of RNA and has been used to examine tertiary structures in yeast $tRNA^{Phe}$ and the *Tetrahymena* group I intron. Interestingly, the major site of interaction with the group I intron occurs at the cofactor (guanosine) binding site. The specific binding interactions of the NiCR complex with these RNAs has yet to be characterized at high resolution, but it appears that the RNA structure is a major determinant for binding specificity.

3. Iron Complexes

MPE-Fe^{II} (methidiumpropyl−EDTA−Fe^{II}) (Figure 9) is an intercalator moiety tethered to a metal-chelating EDTA. Upon the addition of Fe(II), a reducing agent such as dithiothreitol, and hydrogen peroxide, the ferrous ions bound to EDTA can generate diffusible short-lived radicals that have the ability to promote strand scission of the RNA backbone. This reagent binds preferentially to double-stranded sites at or near single-base bulges or at the ends of helices. MPE-Fe^{II} binds to a 345-base RNA fragment representing the S8/S15 protein binding site of *E. coli* 16S rRNA at several helical sites, one at a single-base bulge, and the others at the end of a helix.[54−56] Interestingly, the sites of binding correspond to the S15 binding site. Thus, the tertiary fold of the RNA can generate binding sites for these iron complexes, as well as for the parent ethidium molecule. The binding affinity at these sites was found to be $\sim 10^7$ M^{-1}. MPE-Fe^{II} has also been used to investigate eukaryotic ribosomal structure.[135] Cleavage occurs in relatively few regions of the RNA, and thus provides useful information regarding ribosomal RNA folding and tertiary structure.

Bleomycin (Figure 24A) also cleaves RNA oxidatively in the presence of Fe(II) (Fe-BLM).[136] The RNA cleavage is approximately 10-fold more selective than DNA cleavage. The Fe-BLM complex clearly recognizes specific RNA tertiary structures rather than specific sequences. For example, the cleavage of a *Bacillus subtilis* $tRNA^{His}$ precursor occurs at a single site, but other RNAs without specific binding sites were refractory to cleavage. Even structurally related RNAs such as the *E. coli* $tRNA^{Tyr}$ precursor exhibit differential reactivity, demonstrating the high level of discrimination possible for this natural ligand. At least two of the known cleavage sites of Fe-BLM occur at junctions between single- and double-stranded regions of the RNA substrates or at regions involved with tertiary interactions. Similarly, treatment of yeast 5S rRNA or a 347-nucleotide fragment corresponding to the 5′ end of HIV-1 reverse transcriptase mRNA with Fe-BLM affords several cleavage products. These sites of interaction also occur at junctions between single- and double-stranded RNA or at single-nucleotide bulges. In addition, Fe-BLM has been utilized to study the hairpin loop structure of the IRE (iron regulatory element) of ferritin mRNA. The metal complex is very specific for the RNA hairpin structure and is also sensitive to changes in the RNA structure, such as those mediated by protein binding at a neighboring site.[137]

As with the other metal complexes, some Fe-BLM binding sites may not be identified if they do not lead

to productive strand scission. The effects of polyamines (spermidine), Mg^{2+}, and NaCl on Fe-BLM reactivity suggest that the RNA conformation is a major factor in determining specificity of the complex.[136,138−140] The nature of the structural and conformational requirements of RNA for Fe-BLM binding and cleavage has yet to be characterized. These interactions between Fe-BLM and RNA need to be defined more clearly in order to fully understand the antitumor properties of this and related compounds and to consider them as RNA-targeting drugs. Preliminary models have been constructed by using crystallographically characterized RNAs such as yeast $tRNA^{Asp}$ which is cleaved at a single major site by FE-BLM. The metal complex was docked into the minor groove of the tRNA at a TΨCG stem-loop site and the binding was modulated by hydrogen-bonding and electrostatic interactions.[139]

4. Porphyrin–Metal Complexes

Foster and co-workers have investigated the interactions of yeast $tRNA^{Phe}$ with free porphyrin, T4MPyP (*meso*-tetrakis(4-*N*-methylpyridyl)porphine) (Figure 28B), and its metal complexes (copper(II), manganese(III), and zinc(II)) by UV−visible spectroscopy, circular dichroism, thermal denaturation, and NMR.[141,142] They concluded that a highly specific binding site in the tRNA exists and the binding mode is neither classical intercalation nor electrostatic surface binding. Furthermore, binding of the drug occurred in tertiary regions of the RNA. These interactions have recently been characterized in more detail by Celander and Nussbaum.[143] The porphyrin cation and its analogues recognize hinge regions in RNA where the helices are coaxially stacked. The interaction is enhanced in the absence of stabilizing counterions, suggesting that the stacked helices are destabilized and allow more favorable binding by the porphyrin molecule. Although earlier reports suggested that porphyrins could be potential drugs, perhaps by altering the structure and function of specific RNAs, the tight binding of these molecules to DNA prohibits their usefulness. On the other hand, these porphyrin cations may be extremely useful for monitoring RNA folding interactions such as the coaxial stacking of contiguous helices.

Interestingly, *Conn et al.*[144] have recently demonstrated the ability of an *in vitro*-selected RNA to bind specifically to a mesoporphyrin IX molecule. Furthermore, this 35-nucleotide RNA can catalyze the insertion of Cu(II) into the porphyrin ring, thereby mimicking the human enzyme ferrochelatase. The RNA enzyme inserts Fe(II) into the porphyrin with a similar specificity constant ($k_{cat}/K_m = 2100\ M^{-1}\ s^{-1}$) as the protein enzyme ($k_{cat}/K_m = 1290\ M^{-1}\ s^{-1}$). The predicted secondary structure of the RNA is a stem−loop motif with a long stem (16 base pairs), a large loop (25 nucleotides), and several putative bulge and base mismatch sites. Overall, this work demonstrates the ability of RNA to bind specifically to a porphyrin molecule, to be able to detect subtle changes in porphyrin structure (metalated versus non-metalated or alkylated), and to catalyze the metalation reaction. Further studies are necessary to characterize the porphyrin−RNA interactions, which have strong implications for drug design. In this case, one could imagine the design of an RNA with specific binding motifs that could target a ligand (*e.g.*, porphyrin) *in vivo* and catalyze a biologically important reaction on that ligand, such as metal insertion.

In comparison to the organic systems, the inorganic ligand−RNA interactions have not been examined at high-resolution, however, some general observations can be made. First, the binding generally occurs at sites exhibiting tertiary structure, such as internal loops, bulge sites, or helix junctions. This idea is consistent with the binding of organic ligands to RNA. Second, the recognition sites are often correlated with protein-binding sites, suggesting that the inorganic complexes have potential applications for inhibition of essential RNA−protein interactions. The metal complexes may use a similar structural basis as proteins for the recognition of their RNA targets. In particular, they appear to utilize sites where the major groove is opened by a specific tertiary structure and is more accessible for hydrogen-bonding contacts or van der Waals interactions with the ligand. The RNA-cleaving abilities of the metal complexes may also increase their potential as drug candidates.

C. Peptides

The recognition of RNA by proteins is essential for many cellular functions, such as regulation of gene expression and protein synthesis. The interactions between RNA and proteins have been characterized in some detail by a small number of high-resolution structure studies. These studies have been reviewed[10] and therefore will not be discussed here. Two RNA−peptide examples will be discussed, however, because they serve as relevant models for small molecule−RNA interactions. One can imagine the design of RNA-targeting drugs by using the rules discovered for RNA−peptide complexes.

1. Tat

Replication of HIV requires binding of the viral protein Tat to its RNA target sequence TAR. Peptides derived from Tat bind to a TAR contact site which spans approximately five base pairs and includes a trinucleotide pyrimidine bulge and hexanucleotide loop. Biochemical studies indicate that the effect of the bulge site is to make an accessible Tat site in the otherwise deep and narrow major groove of the duplex RNA.[24] Fragments of the HIV Tat protein (24 amino acids), representing the basic arginine-rich domain, bind to a representative TAR RNA fragment (Figure 11) with a dissociation constant in the subnanomolar range.[105,145] Furthermore, L-arginine is able to block the Tat peptide−TAR interaction, implicating the important role of a single arginine residue.[69,146] NMR studies with the amino acid analogue argininamide and TAR RNA fragment have already been mentioned. In short, the study showed that the RNA undergoes a conformational change upon ligand binding and that the complex is stabilized by specific hydrogen bonding, stacking, and electrostatic interactions.[70,106] The structural model suggests that the guanidinium group interacts specif-

ically with the trinucleotide bulge and is stabilized by an adjacent triple-base interaction (U23−A27−U38).

A more detailed NMR study of the Tat−TAR interactions was carried out by using a Tat-derived peptide and was compared to the results with a single amino acid. Overall, the high-resolution structure was consistent with previous results using argininamide except for the assignment of the base-triple interaction. Aboul-ela *et al.*[147] have shown that U13 is positioned near G26 and A27 in the major groove of the RNA rather than being stacked as in the free TAR RNA. U23 and G26 are in close contact with the guanidinium side chain of a specific arginine residue and there are multiple contacts between the RNA and the Tat peptide, but a complete triple-base interaction is not observed. In summary, several important features of the argininamide−TAR and Tat peptide−TAR complexes should be noted. First, the ligand (either arginine or peptide) binds specifically at the UCU bulge of TAR RNA fragments and induces an RNA conformational change. In particular, A22 and U23 stacking interactions are disrupted and U23 becomes intimately involved with ligand binding. Second, a higher concentration of argininamide than Tat peptide is required for binding to TAR. The TAR−argininamide interactions involve a subset of contacts observed in the Tat peptide−TAR complex, and there is a subtle difference in the structures of the complexes formed by these two ligands. Finally, the Tat peptide or argininamide ligands both contact RNA in the major groove at an accessible region, namely the bulge site, and recognize an array of functional groups that are available for specific binding interactions in this groove.

Recently, a related system from bovine immunodeficiency virus (BIV) was examined. In particular, a BIV Tat peptide was shown to interact with a 28-nucleotide TAR RNA stem−loop region that contains two single-nucleotide bulges separated by one base pair.[148] In this case, the Tat−TAR interaction is strongly dependent on several amino acid residues (Arg70, Gly71, Thr72, Arg73, Arg77, and Ile79) rather than a single arginine moiety. An NMR structure of a 14-residue BIV Tat peptide and TAR RNA revealed that the protein binds in the major groove of the RNA, and specific contacts with those amino acids are apparent.[149] As with the HIV system, the free RNA undergoes a conformational change upon ligand binding. Specifically, the U10 bulged nucleotide becomes unstacked as it loops out of the helix, and there is a minor distortion of the RNA helical structure at the stem junction, thus leading to an opening of the major groove at the peptide binding site. The peptide binds deeply in the major groove and makes specific hydrogen bonding and hydrophobic contacts with the RNA. Ye *et al.*[131] have provided further details on the structure of the complex and shown that the U10 residue becomes involved with a triple-base interaction with A13 and U24. This specific interaction positions U10 for contacts with the peptide and neighboring residues.

The NMR structures of the Tat−TAR complexes have provided important information regarding RNA−ligand interactions. Some of the features observed, such as RNA conformational changes, recognition of open sites involving RNA tertiary structure, and specific contacts in the RNA major groove, have been seen in other systems and are likely to emerge again. It should also be noted that we have focused mainly on the RNA structure. The peptides or small molecules can also change their shapes or structures upon binding to the RNA, thus providing two approaches to adaptive binding.

2. Rev

Another peptide−RNA example from HIV that has been examined is the Rev−RRE mRNA complex. This system has also been discussed in previous sections with respect to inhibition by small molecules. The binding of a minimal 34-nucleotide Rev responsive element (RRE) motif (residues 41−79) of HIV mRNA and a 22-amino acid basic fragment of the viral Rev protein has been studied by using NMR spectroscopy.[150,151] As seen with the Tat−TAR system, the RNA undergoes a conformational change upon ligand binding. Two purine−purine mismatch pairs (G48−G71 and G47−A73) (see Figure 23A) are formed in the complexed RNA which stack on each other and cause the two remaining residues (U72 and A68) to loop out into solution. Furthermore, specific RNA−ligand contacts occur in the major groove involving arginine, asparagine, and threonine residues of the peptide and the internal loop nucleotides. Overall, it can be seen once again that the non-Watson−Crick base pairs provide recognition sites within the widened RNA major groove for specific ligands, and the RNA structure is stabilized in the RNA−ligand complex.

Interestingly, an *in vitro* selected RNA aptamer binds similarly to a 17-mer Rev peptide, involving RNA conformational changes, G47−A73 and A48−A71 mismatch base pairs, and bulged nucleotides.[152] However, subtle differences can also be observed between the two high-resolution structures involving specific amino acid−RNA contacts as well as local RNA structure. These results may also explain the neomycin resistance that Werstuck and co-workers noted with *in vitro* selected RRE variants.[153] In their study, an A48−A71 substitution was sufficient to inhibit binding by the drug, yet produced a stronger protein−RNA complex. Thus, substitutions that favor protein binding may actually inhibit antibiotic binding, suggesting different modes of binding for the two classes of molecules. Taken together, these results suggest the possibility of designing drugs with unique specificities and binding modes. Target molecules designed to inhibit these protein−RNA interactions will be most effective if they can closely mimic the protein binding interactions or cause alternative conformations to be stabilized that are nonproductive in protein binding.

V. Conclusions and Future Challenges

It is now apparent that RNA has the ability to form an array of tertiary structures and specific binding pockets for small molecule substrates (or ligands) and catalytic cofactors. The future development of RNA-targeting drugs will rely on a deeper understanding

of these binding processes. At the present time, the paucity of high-resolution structures of RNA−ligand complexes has made it difficult to discover general classes of motifs that recognize small molecules. Yet, as can be seen here, there is a rapidly increasing number of known specific RNA-binding molecules and more structural information has emerged. In addition, more biologically relevant RNA targets have been identified, and the high-resolution methods to study them and their potential interactions with ligands are developing at a fast pace. In comparison to the free RNAs, RNA−ligand complexes have proven to be even easier to study by NMR because of increased stability and decreased conformational heterogeneity of the bound RNAs. Many of the preliminary binding studies mentioned in this review will serve as starting points for further structural studies involving small molecule−RNA recognition. In addition, some general clues regarding RNA−ligand interactions have already been obtained from high-resolution NMR studies on *in vitro* selected RNAs and their ligand counterparts. As the data set expands, structural relationships between classes of small molecules and their corresponding RNA binding sites may be realized.

Thus far, it has been demonstrated that ligands generally bind to RNA by using specific tertiary motifs, such as internal loops or base bulges, in which the normally deep and narrow major groove has become accessible for binding interactions. Furthermore, ligands may exploit regions of flexibility in the RNA, in which the binding site can be adapted for specific and tight binding by a small molecule. Because proteins appear to use similar rules for binding to RNA, the natural protein-binding sites are desirable targets for drug action. Perhaps both natural products and designed ligands can be used to exploit the structure-specific recognition in the same way that protein, RNAs, or other natural biological effectors (*e.g.*, antibiotics) bind to RNA. Thus, a knowledge of the structural basis of these RNA−ligand interactions will ultimately provide an impetus for rational drug discovery.

Finally, several topics have not been discussed here, but are also important to consider when studying potential RNA-targeting drugs. First, an understanding of the kinetics of ligand binding and a knowledge of how ligands identify their preferred binding site(s) among many potentially lower affinity sites need to be developed. Second, the role of cellular factors and *in vivo* conditions on the RNA−ligand interactions needs to be determined. Third, the role of modified RNA nucleosides in mediating the ligand-binding processes needs to be assessed. Modified nucleosides are highly abundant in naturally occurring RNAs, and can potentially alter or regulate the RNA tertiary structures, as well as ligand-binding sites. Regardless, an increased understanding of RNA−ligand interactions will emerge in the near future as more structural information is obtained. Thus, we believe the discovery of highly specific RNA-targeting inhibitors or biological effectors is imminent, and current studies may eventually lead to improved antibiotics, antiviral agents, or other RNA-specific drugs with clinical applications.

VI. Acknowledgments

The authors acknowledge support by Wayne State University. We thank Professor J. K. Barton and A. C. Lim of Caltech for a preprint of their work. We are grateful to P. Grohar and Professor A. Nicholson for critical reading of the manuscript, and to H. Allawi, Professor J. SantaLucia, I. Massova, and Professor S. Mobashery for use of their Silicon Graphics computers and assistance with figure preparations.

VII. References

(1) Kim, S. H.; Sussman, J. L.; Suddath, F. L.; Quigley, G. J.; McPherson, A.; Wang, A. H. J.; Seeman, N. C.; Rich, A. *Proc. Natl. Acad. Sci. USA* **1974**, *71*, 4970.
(2) Westhof, E.; Dumas, P.; Moras, D. *J. Mol. Biol.* **1985**, *184*, 119.
(3) Pley, H. W.; Flaherty, K. M.; McKay, D. B. *Nature* **1994**, *372*, 68.
(4) Cate, J. H.; Gooding, A. R.; Podell, E.; Zhou, K.; Golden, B. L.; Kundrot, C. E.; Cech, T. R.; Doudna, J. A. *Science* **1996**, *273*, 1678.
(5) Patel, D. J.; Shapiro, L.; Hare, D. *Q. Rev. Biophys.* **1987**, *20*, 78.
(6) Varani, G.; Aboul-ela, F.; Allain, F. H.-T. *Prog. Nucl. Magn. Reson. Spectrosc.* **1996**, *29*, 51.
(7) Pardi, A. *Methods Enzymol.* **1995**, *261*, 350.
(8) Famulok, M.; Szostak, J. W. *Angew. Chem., Int. Ed. Engl.* **1992**, *31*, 979.
(9) Szostak, J. W. *Trends Biochem. Sci.* **1992**, *17*, 89.
(10) Nagai, K. *Curr. Opin. Struct. Biol.* **1996**, *6*, 53.
(11) Burd, C. G.; Dreyfuss, G. *Science* **1994**, *265*, 615.
(12) Glück, A.; Wool, I. G. *J. Mol. Biol.* **1996**, *256*, 838.
(13) Szewczak, A. A.; Moore, P. B. *J. Mol. Biol.* **1995**, *247*, 81.
(14) Batey, R. T.; Inada, M.; Kujawinski, E.; Puglisi, J. D.; Williamson, J. R. *Nucleic Acids Res.* **1992**, *20*, 4515.
(15) Doudna, J. A.; Grosshans, C.; Gooding, A.; Kundrot, C. E. *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 7829.
(16) Saenger, W. *Principles of Nucleic Acid Structure*; Springer-Verlag: New York, 1984.
(17) Alden, C. J.; Kim, S.-H. *J. Mol. Biol.* **1979**, *132*, 411.
(18) Quigley, G. J.; Rich, A. *Science* **1976**, *19*, 796.
(19) Wyatt, J. R.; Tinoco, I., Jr. In *The RNA World*; Gesteland, R. F., Atkins, J. F., Eds.; Cold Spring Harbor Laboratory Press: Plainview, 1993; pp 465.
(20) Scott, W. G.; Finch, J. T.; Klug, A. *Cell* **1995**, *81*, 991.
(21) Heus, H. A.; Pardi, A. *Science* **1991**, *253*, 191.
(22) Cheong, C.; Varani, G.; Tinoco, I., Jr. *Nature* **1990**, *346*, 680.
(23) Varani, G.; Cheong, C.; Tinoco, I., Jr. *Biochemistry* **1991**, *30*, 3280.
(24) Weeks, K. M.; Crothers, D. M. *Cell* **1991**, *66*, 577.
(25) Weeks, K. M.; Crothers, D. M. *Science* **1993**, *261*, 1574.
(26) Puglisi, J. D.; Wyatt, J. R.; Tinoco, I., Jr. *J. Mol. Biol.* **1990**, *214*, 437.
(27) Binkley, J.; Allen, P.; Brown, D. M.; Green, L.; Tuerk, C.; Gold, L. *Nucleic Acids Res.* **1995**, *23*, 3198.
(28) Green, L.; Waugh, S.; Binkley, J. P.; Hostomska, Z.; Hostomsky, Z.; Tuerk, C. *J. Mol. Biol.* **1995**, *247*, 60.
(29) Jack, A.; Ladner, J. E.; Rhodes, D.; Brown, R. S.; Klug, A. *J. Mol. Biol.* **1977**, *111*, 315.
(30) Luzzati, V.; Masson, F.; Lerman, L. S. *J. Mol. Biol.* **1961**, *3*, 634.
(31) Quigley, G. J.; Teeter, M. M.; Rich, A. *Proc. Natl. Acad. Sci. USA* **1978**, *75*, 64.
(32) Frydman, B.; de los Santos, C.; Frydman, R. B. *J. Biol. Chem.* **1990**, *265*, 20874.
(33) Frydman, L.; Rossomando, P. C.; Frydman, V.; Fernandez, C. O.; Frydman, B.; Samejima, K. *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 9186.
(34) Frydman, B.; Westler, W. M.; Samejima, K. *J. Org. Chem.* **1996**, *61*, 2588.
(35) Fernandez, C. O.; Frydman, B.; Samejima, K. *Cell., Mol. Biol.* **1994**, *40*, 933.
(36) Tao, T.; Nelson, J. H.; Cantor, C. R. *Biochemistry* **1970**, *9*, 3514.
(37) Liebman, M.; Rubin, J.; Sundaralingam, M. *Proc Natl. Acad. Sci. USA* **1977**, *74*, 4821.
(38) Jones, C. R.; Bolton, P. H.; Kearns, D. R. *Biochemistry* **1978**, *17*, 601.
(39) Jones, C. R.; Kearns, D. R. *Biochemistry* **1975**, *14*, 2660.
(40) Lamos, M. L.; Lobenstine, E. W.; Turner, D. H. *J. Am. Chem. Soc.* **1986**, *108*, 4278.
(41) Pilch, D. S.; Kirolos, M. A.; Liu, X.; Plum, G. E.; Breslauer, K. J. *Biochemistry* **1995**, *34*, 9962.
(42) Tanious, F. A.; Veal, J. M.; Buczak, H.; Ratmeyer, L. S.; Wilson, W. D. *Biochemistry* **1992**, *31*, 3103.
(43) Zhao, M.; Janda, L.; Nguyen, J.; Strekowski, L.; Wilson, W. D. *Biopolymers* **1994**, *34*, 61.

(44) Zhao, M.; Ratmeyer, L.; Peloquin, R. G.; Yao, S.; Kumar, A.; Spychala, J.; Boykin, D. W.; Wilson, W. D. *Bioorg., Med. Chem.* **1995**, *3*, 785.
(45) McConnaughie, A. W.; Spychala, J.; Zhao, M.; Boykin, D.; Wilson, W. D. *J. Med. Chem.* **1994**, *37*, 1063.
(46) Latham, J. A.; Cech, T. R. *Science* **1989**, *245*, 276.
(47) Celander, D. W.; Cech, T. R. *Biochemistry* **1990**, *29*, 1355.
(48) Murakawa, G. J.; Chen, C.-h. B.; Kuwabara, M. D.; Nierlich, D. P.; Sigman, D. S. *Nucleic Acids Res.* **1989**, *17*, 5361.
(49) Mazumder, A.; Chen, C.-h. B.; Gaynor, R.; Sigman, D. S. *Biochem. Biophys. Res. Commun.* **1992**, *187*, 1503.
(50) Pearson, L.; Chen, C.-h. B.; Gaynor, R. P.; Sigman, D. S. *Nucleic Acids Res.* **1994**, *22*, 2255.
(51) Tuerk, C.; Gold, L. *Science* **1990**, *249*, 505.
(52) Ellington, A. D.; Szostak, J. W. *Nature* **1990**, *346*, 818.
(53) Helfgott, D. C.; Kallenbach, N. R. *Nucleic Acids Res.* **1979**, *7*, 1011.
(54) White, S. A.; Draper, D. E. *Biochemistry* **1989**, *28*, 1892.
(55) White, S. A.; Draper, D. E. *Nucleic Acids Res.* **1987**, *15*, 4049.
(56) Kean, J. M.; White, S. A.; Draper, D. E. *Biochemistry* **1985**, *24*, 5062.
(57) Tanner, N. K.; Cech, T. R. *Nucleic Acids Res.* **1985**, *13*, 7759.
(58) Tanner, N. K.; Cech, T. R. *Nucleic Acids Res.* **1985**, *13*, 7741.
(59) Ratmeyer, L. S.; Vinayak, R.; Zon, G.; Wilson, W. D. *J. Med. Chem.* **1992**, *35*, 966.
(60) Bailly, C.; Colson, P.; Houssier, C.; Hamy, F. *Nucleic Acids Res.* **1996**, *24*, 1460.
(61) Wilson, W. D.; Ratmeyer, L.; Cegla, M. T.; Spychala, J.; Boykin, D.; Demeunynck, M.; Lhomme, J.; Krishnan, G.; Kennedy, D.; Vinayak, R.; Zon, G. *New J. Chem.* **1994**, *18*, 419.
(62) Ratmeyer, L.; Zapp, M. L.; Green, M. R.; Vinayak, R.; Kumar, A.; Boykin, D. W.; Wilson, W. D. *Biochemistry* **1996**, *35*, 13689.
(63) Bass, B. L.; Cech, T. R. *Nature* **1984**, *308*, 820.
(64) Bass, B. L.; Cech, T. R. *Biochemistry* **1986**, *25*, 4473.
(65) Yarus, M. *Science* **1988**, *240*, 1751.
(66) Yarus, M. *Biochemistry* **1989**, *28*, 980.
(67) Michel, F.; Hanna, M.; Green, R.; Bartel, D. P.; Szostak, J. W. *Nature* **1989**, *342*, 391.
(68) Yarus, M.; Majerfeld, I. *J. Mol. Biol.* **1992**, *225*, 945.
(69) Calnan, B. J.; Tidor, B.; Biancalana, S.; Hudson, D.; Frankel, A. D. *Science* **1991**, *252*, 1167.
(70) Puglisi, J. D.; Tan, R.; Calnan, B. J.; Frankel, A. D.; Williamson, J. R. *Science* **1992**, *257*, 76.
(71) Connell, G. J.; Illangesekare, M.; Yarus, M. *Biochemistry* **1993**, *32*, 5497.
(72) Tao, J.; Frankel, A. D. *Biochemistry* **1996**, *35*, 2229.
(73) Connell, G. J.; Yarus, M. *Science* **1994**, *264*, 1137.
(74) Famulok, M. *J. Am. Chem. Soc.* **1994**, *116*, 1698.
(75) Yang, Y.; Kochoyan, M.; Burgstaller, P.; Westhof, E.; Famulok, M. *Science* **1996**, *272*, 1343.
(76) Famulok, M.; Szostak, J. W. *J. Am. Chem. Soc.* **1992**, *114*, 3990.
(77) Majerfeld, I.; Yarus, M. *Struct. Biol.* **1994**, *1*, 287.
(78) Sassanfar, M.; Szostak, J. W. *Nature* **1993**, *364*, 550.
(79) Feigon, J.; Dieckmann, T.; Smith, F. W. *Chem. Biol.* **1996**, *3*, 611.
(80) Dieckman, T.; Suzuki, E.; Nakamura, G. K.; Feigon, J. *RNA* **1996**, *2*, 628.
(81) Jiang, F.; Kumar, R. A.; Jones, R. A.; Patel, D. J. *Nature* **1996**, *382*, 183.
(82) Jiang, F.; Fiala, R.; Live, D.; Kumar, R. A.; Patel, D. J. *Biochemistry* **1996**, *35*, 13250.
(83) Jucker, F. M.; Pardi, A. *RNA* **1995**, *1*, 219.
(84) Burgstaller, P.; Famulok, M. *Angew. Chem., Int. Ed. Engl.* **1994**, *33*, 1084.
(85) Lorsch, J. R.; Szostak, J. W. *Biochemistry* **1994**, *33*, 973.
(86) Lauhon, C. T.; Szostak, J. W. *J. Am. Chem. Soc.* **1995**, *117*, 1246.
(87) Fan, P.; Suri, A. K.; Fiala, R.; Live, D.; Patel, D. J. *J. Mol. Biol.* **1996**, *258*, 480.
(88) Burgstaller, P.; Famulok, M. *J. Am. Chem. Soc.* **1997**, *119*, 1137.
(89) Gale, E. F.; Cundliffe, E.; Reynolds, P. E.; Richmond, M. H.; Waring, M. J. *The Molecular Basis of Antibiotic Activity*; John Wiley & Sons: New York, 1972; pp 278.
(90) Moazed, D.; Noller, H. F. *Nature* **1987**, *327*, 389.
(91) Noller, H. In *The RNA World*; Gesteland, R. F., Atkins, J. F., Eds.; Cold Spring Harbor Laboratory Press: New York, 1993; pp 137.
(92) Noller, H. F. *Annu. Rev. Biochem.* **1991**, *60*, 191.
(93) Davies, J.; von Ahsen, U.; Schroeder, R. In *The RNA World*; Gesteland, R. F., Atkins, J. F., Eds.; Cold Spring Harbor Laboratory Press: New York, 1993; pp 185.
(94) Woodcock, J.; Moazed, D.; Cannon, M.; Davies, J.; Noller, H. F. *EMBO J.* **1991**, *10*, 3099.
(95) von Ahsen, U.; Schroeder, R. *Nature* **1990**, *346*, 801.
(96) von Ahsen, U.; Davies, J.; Schroeder, R. *Nature* **1991**, *353*, 368.
(97) von Ahsen, U.; Schroeder, R. *Nucleic Acids Res.* **1991**, *19*, 2261.
(98) von Ahsen, U.; Davies, J.; Schroeder, R. *J. Mol. Biol.* **1992**, *226*, 935.
(99) von Ashen, U.; Noller, H. F. *Science* **1993**, *260*, 1500.
(100) Stage, T. K.; Hertel, K. J.; Uhlenbeck, O. C. *RNA* **1995**, *1*, 95.
(101) Botto, R. E.; Coxon, B. *J. Am. Chem. Soc.* **1983**, *105*, 1021.
(102) Clouet-d'Orval, B.; Stage, T. K.; Uhlenbeck, O. C. *Biochemistry* **1995**, *34*, 11186.
(103) Zapp, M. L.; Stern, S.; Green, M. R. *Cell* **1993**, *74*, 969.
(104) Mei, H.-Y.; Galan, A. A.; Halim, N. S.; Mack, D. P.; Moreland, D. W.; Sanders, K. B.; Truong, H. N.; Czarnik, A. W. *Bioorg., Med. Chem. Lett.* **1995**, *5*, 2755.
(105) Weeks, K. M.; Ampe, C.; Schultz, S. C.; Steitz, T. A.; Crothers, D. M. *Science* **1990**, *249*, 1281.
(106) Puglisi, J. D.; Chen, L.; Frankel, A. D.; Williamson, J. R. *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 3680.
(107) Park, W. K. C.; Auer, M.; Jaksche, H.; Wong, C.-H. *J. Am. Chem. Soc.* **1996**, *118*, 10150.
(108) Purohit, P.; Stern, S. *Nature* **1994**, *370*, 659.
(109) Miyaguchi, H.; Narita, H.; Sakamoto, K.; Yokoyama, S. *Nucleic Acids Res.* **1996**, *24*, 3700.
(110) Wang, Y.; Hamasaki, K.; Rando, R. R. *Biochemistry* **1997**, *36*, 768.
(111) Recht, M. I.; Fourmy, D.; Blanchard, S. C.; Dahlquist, K. D.; Puglisi, J. D. *J. Mol. Biol.* **1996**, *262*, 421.
(112) Fourmy, D. F.; Recht, M. I.; Blanchard, S. C.; Puglisi, J. D. *Science* **1996**, *274*, 1367.
(113) Famulok, M.; Hüttenhofer, A. *Biochemistry* **1996**, *35*, 4265.
(114) Wallis, M. G.; von Ashen, U.; Schroeder, R.; Famulok, M. *Chem. Biol.* **1995**, *2*, 543.
(115) Lato, S. M.; Boles, A. R.; Ellington, A. D. *Chem. Biol.* **1995**, *2*, 291.
(116) Wang, Y.; Rando, R. R. *Chem. Biol.* **1995**, *2*, 281.
(117) Wang, Y.; Killian, J.; Hamasaki, K.; Rando, R. R. *Biochemistry* **1996**, *35*, 12338.
(118) Jiang, L.; Suri, A. K.; Fiala, R.; Patel, D. J. *Chem. Biol.* **1997**, *4*, 35.
(119) Keck, M. V.; Hecht, S. M. *Biochemistry* **1995**, *34*, 12029.
(120) Kappen, L. S.; Goldberg, I. H. *Biochemistry* **1995**, *34*, 5997.
(121) Battigello, J.-M. A.; Cui, M.; Roshong, S.; Carter, B. J. *Bioorg., Med. Chem.* **1995**, *3*, 839.
(122) Jenison, R. D.; Gill, S. C.; Pardi, A.; Polisky, B. *Science* **1994**, *263*, 1425.
(123) Chow, C. S.; Barton, J. K. *J. Am. Chem. Soc.* **1990**, *112*, 2839.
(124) Chow, C. S.; Barton, J. K. *Biochemistry* **1992**, *31*, 5423.
(125) Chow, C. S.; Behlen, L. S.; Uhlenbeck, O. C.; Barton, J. K. *Biochemistry* **1992**, *31*, 972.
(126) Chow, C. S.; Hartmann, K. M.; Rawlings, S. L.; Huber, P. W.; Barton, J. K. *Biochemistry* **1992**, *31*, 3534.
(127) Scripture, J. B.; Huber, P. W. *J. Biol. Chem.* **1995**, *270*, 27358.
(128) Wimberly, B.; Varani, G.; Tinoco, I., Jr. *Biochemistry* **1993**, *32*, 1078.
(129) Neenhold, H. R.; Rana, T. M. *Biochemistry* **1995**, *34*, 6303.
(130) Lim, A. C.; Barton, J. K. *Bioorg., Med. Chem.*, in press.
(131) Ye, X.; Kumar, R. A.; Patel, D. J. *Chem. Biol.* **1995**, *2*, 827.
(132) Thorp, H. H.; McKenzie, R. A.; Lin, P.-N.; Walden, W. E.; Theil, E. C. *Inorg. Chem.* **1996**, *35*, 2773.
(133) Chen, X.; Woodson, S. A.; Burrows, C. J.; Rokita, S. E. *Biochemistry* **1993**, *32*, 7610.
(134) Burrows, C. J.; Rokita, S. E. *Acc. Chem. Res.* **1994**, *27*, 295.
(135) Han, H.; Schepartz, A.; Pellegrini, M.; Dervan, P. B. *Biochemistry* **1994**, *33*, 9831.
(136) Carter, B. J.; de Vroom, E.; Long, E. C.; van der Marel, G. A.; van Boom, J. H.; Hecht, S. M. *Proc. Natl. Acad. Sci. USA* **1990**, *87*, 9373.
(137) Theil, E. C. *New J. Chem.* **1994**, *18*, 435.
(138) Carter, B. J.; Holmes, C. E.; Van Atta, R. B.; Dange, V.; Hecht, S. M. *Nucleosides Nucleotides* **1991**, *10*, 215.
(139) Hecht, S. M. *Bioconjugate Chem.* **1994**, *5*, 513.
(140) Holmes, C. E.; Carter, B. J.; Hecht, S. M. *Biochemistry* **1993**, *32*, 4293.
(141) Foster, N.; Singhal, A. K.; Smith, M. W.; Marcos, N. G.; Schray, K. J. *Biochim. Biophys. Acta* **1988**, *950*, 118.
(142) Birdsall, W. J.; Anderson, W. R., Jr.; Foster, N. *Biochim. Biophys. Acta* **1989**, *1007*, 176.
(143) Celander, D. W.; Nussbaum, J. M. *Biochemistry* **1996**, *35*, 12061.
(144) Conn, M. M.; Prudent, J. R.; Schultz, P. G. *J. Am. Chem. Soc.* **1996**, *118*, 7012.
(145) Long, K. S.; Crothers, D. M. *Biochemistry* **1995**, *34*, 8885.
(146) Tao, J.; Frankel, A. D. *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 2723.
(147) Aboul-ela, F.; Karn, J.; Varani, G. *J. Mol. Biol.* **1995**, *253*, 313.
(148) Chen, L.; Frankel, A. D. *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 5077.
(149) Puglisi, J. D.; Chen, L.; Blanchard, S.; Frankel, A. D. *Science* **1995**, *270*, 1200.
(150) Battiste, J. L.; Tan, R.; Frankel, A. D.; Williamson, J. R. *Biochemistry* **1994**, *33*, 2741.
(151) Battiste, J. L.; Mao, H.; Rao, N. S.; Tan, R.; Muhandiram, D. R.; Kay, L. E.; Frankel, A. D.; Williamson, J. R. *Science* **1996**, *273*, 1547.
(152) Ye, X.; Gorin, A.; Ellington, A. D.; Patel, D. J. *Nature Struct. Biol.* **1996**, *3*, 1026.
(153) Werstuck, G.; Zapp, M. L.; Green, M. R. *Chem. Biol.* **1996**, *3*, 129.

CR960415W