RNA as a drug target

Neil D Pearson[1] and Catherine D Prescott[2]

Historically, the pharmaceutical industry has focused on proteins, rather than nucleic acids, as drug targets. But recent advances in the fields of RNA synthesis, structure determination and therapeutic target identification make the systematic exploitation of RNA as a drug target a realistic goal.

Addresses: [1]Department of Medicinal Chemistry, SmithKline Beecham Pharmaceuticals, New Frontiers Science Park (North), Third Avenue, Harlow, Essex CM19 5AW, UK. [2]RNA Research Group, 1250 South Collegeville Road, P.O. Box 5089, Collegeville, PA 19426-0989, USA.

E-mail: neil_pearson-1@sbphrd.com and cathy_prescott-1@sbphrd.com

Chemistry & Biology June 1997, **4**:409–414
http://biomednet.com/elecref/1074552100400409

Compared with the structural information available for DNA–protein complexes, relatively little is known about the structures of RNA–protein complexes. But RNA–protein recognition is arguably a variation on the well-documented protein–protein recognition paradigms. RNA tertiary structure is extremely diverse and provides the basis for specific recognition by RNA-binding proteins. In contrast, the structure of DNA double helices is essentially invariable, the major groove is readily accessible and DNA–protein recognition is more dependent on sequence than tertiary structure.

Many features of RNA make it attractive as a therapeutic target. It is important in many functions of the cell, participating in nearly all macromolecular processes; it is structurally flexible, both in secondary structure and in tertiary structure, suggesting that selective binding should be possible; and it lacks a cellular repair mechanism. Drugs targeted towards ribonucleoprotein (RNP) particles have demonstrated therapeutic success in the antibacterial area. Here, we will argue that recent advances in the fields of RNA synthesis, structure determination and therapeutic target identification offer new opportunities for drug discovery, especially in the areas of antibacterial and antiviral targets.

Targeting RNA by design: new twists on old themes

Elucidating the tertiary structure of a therapeutic target yields a wealth of information for drug design. But large RNAs are notoriously difficult to crystallise and nuclear magnetic resonance techniques for structure determinations are restricted to relatively small fragments of RNA. Fortunately, large RNAs are often functional when their dissociated parts are non-covalently recombined [1]. Using a reductionist approach, it has therefore been possible to make many functionally relevant RNA fragments available for structural analysis.

The detailed structural information available on many protein enzymes has successfully aided mechanism-based drug design. The ever increasing level of crystallographic and biochemical data for ribozymes such as RNase P, hammerhead and group I intron [2], affords another opportunity for mechanism-based drug design. For example, inhibitors of bacterial RNase P, an essential endoribonuclease that mediates phosphodiester bond cleavage of a variety of RNA substrates, are potential drug candidates.

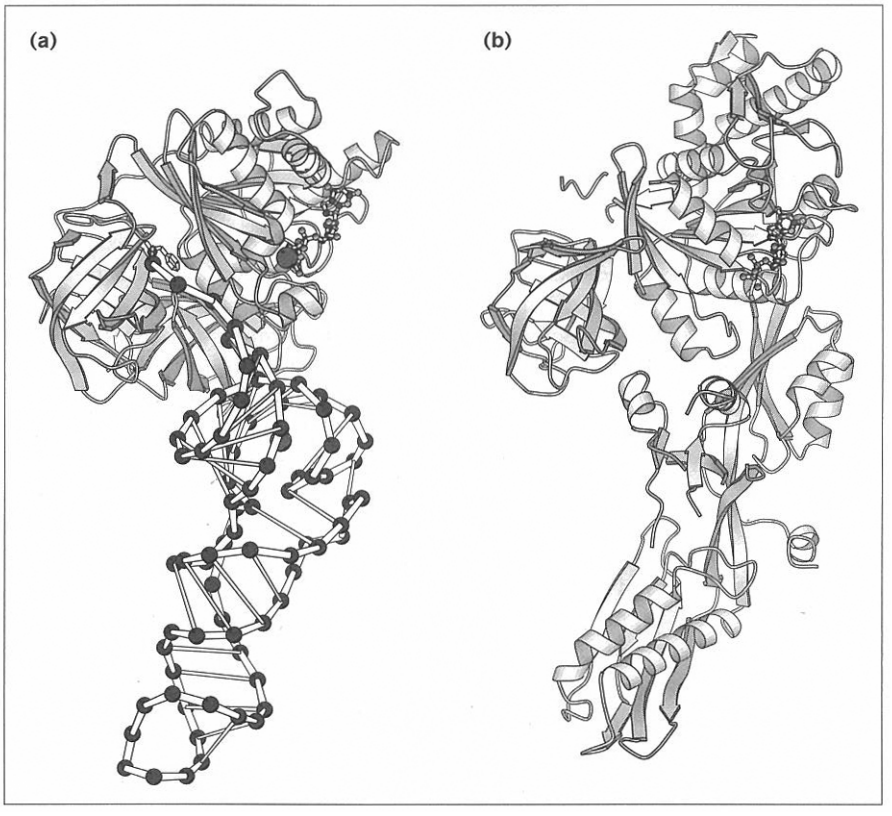
An understanding of the catalytic properties of RNA has led to the development of 'designer' ribozymes that site-specifically cleave RNA [2]. The potential for a small molecule to fulfil this role is supported by studies on bleomycin, which showed that cleavage of RNA occurs in a more site-selective manner than for DNA [3]. Improving the selectivity for RNA versus DNA may ultimately produce site-selective RNA cleaving agents as antibiotics.

Motif mimicry is another popular approach to drug design. Examples in which proteins mimic RNA suggest that RNA and proteins have common structural requirements for functionality. Familiar strategies for designing small molecules that mimic protein motifs may therefore be applicable to RNA mimics. For example, the ternary complex formed by phenylalanyl tRNA, the non-hydrolysable GTP analogue GDPNP, and the translation elongation factor Tu (EF-Tu) can be readily superimposed on the complex of elongation factor G (EF-G), the tRNA translocase, with GDP (Figure 1; [4]). Both of these complexes thus appear to bind to structurally similar sites on the ribosome, indicating that common structural elements exist between the ribosome's pre-translocation state (when the EF-Tu ternary complex binds) and the post-translocation state (when the EF-G–GDP complex is released). It has been suggested that domain IV of elongation factor G may be a mimic of the anticodon loop of tRNA. It may therefore be possible to alter codon-anticodon recognition using drugs designed to mimic a protein lead, and this may produce antibacterial agents.

Exploiting RNA conformational flexibility

The ability of RNA molecules to exchange between alternative conformations can be essential for biological function. But the existence of multiple states may create problems for the structure-based design approach mentioned above. Nevertheless, opportunities for therapeutic

Figure 1



The structure of the *Thermus thermophilus* elongation factor EF-G. **(a)** EF-G complexed with GDP (EF-G–GDP); **(b)** the ternary complex of Phe-tRNA, *Thermus aquaticus* elongation factor EF-Tu and the non-hydrolysable GTP analogue, GDPNP (EF-Tu–GDPNP–Phe-$tRNA^{Phe}$). Domain II in EF-G was superimposed on the corresponding domain II in EF-Tu and the structures are shown in a schematic representation. The overall shape of the structures is similar, suggesting common structural requirements for their function on the ribosome [4]. Figure kindly provided by Poul Nisson and reproduced with permission [30].

intervention by restricting conformational exchange should not be overlooked.

It is possible to imagine many ways in which a drug might be able to achieve conformational restriction of RNA, either by interacting directly with the ligand-binding surface or by making use of allosteric effects. Strategies for preventing RNA–ligand complex formation include stabilisation of the unbound conformation of RNA, the stabilisation of any intermediate state between the free and bound conformation of RNA, and the induction and stabilisation of unnatural conformations.

The stabilisation of transient RNA–RNA or RNA–protein complexes can also be exploited for drug development. For example, the binding of the aminoglycoside antibiotics in the major groove of the decoding site of 16S ribosomal RNA is proposed to induce an increase in miscoding by enhancing the affinity of the codon–anticodon complex for the minor groove (Figure 2; [5]). This example also serves to illustrate that small molecules that bind in the major or minor groove of bulged regions within A-form helical RNA can have inhibitory effects either directly or allosterically.

The possibility that any RNA–drug complex may use multiple inhibition mechanisms increases the importance of detailed structural information. Recent advances (e.g. [5–7]) in the synthesis of RNA have already been shown to facilitate the structural determination of RNA and RNA–drug complexes.

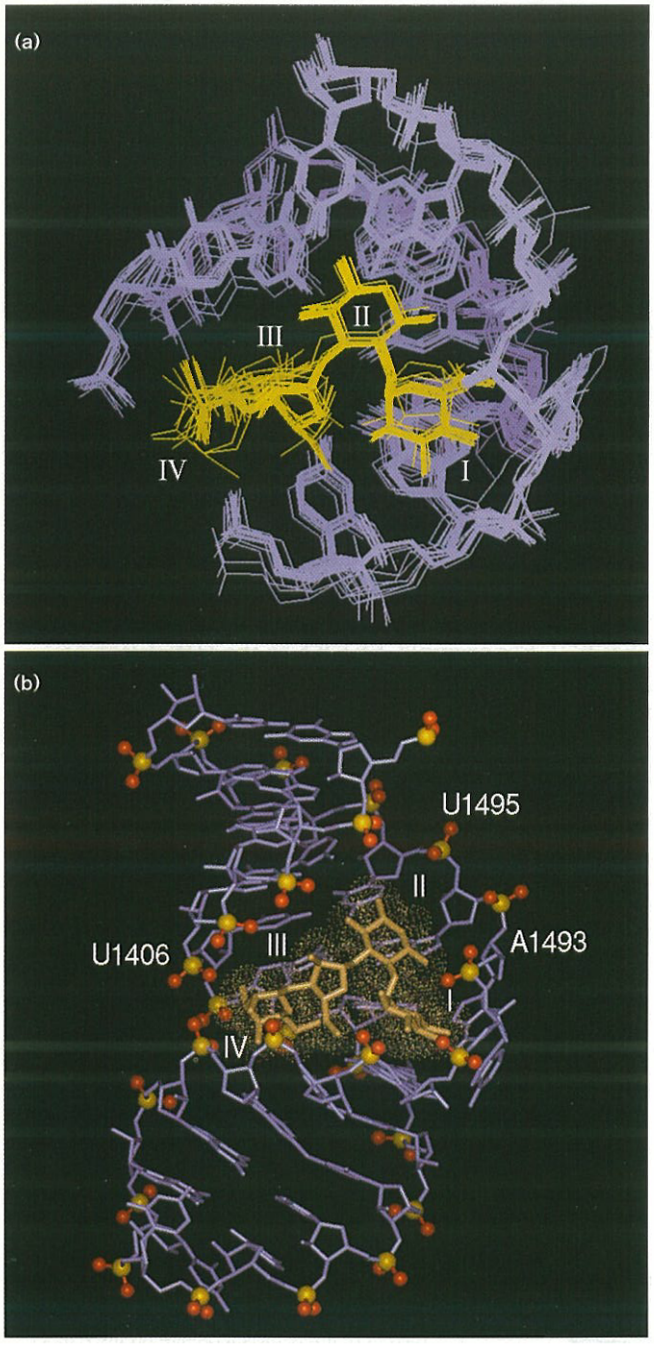
A new reflection on X-ray crystallography: (D,L)-RNA and (L)-RNA

The unnatural enantiomer of RNA, (L)-RNA (see [8] for chemical synthesis), can be combined with the natural enantiomer, (D)-RNA, to produce racemic mixtures. Because there are 100 additional crystal packing arrangements available for racemates of optically active molecules compared with homochiral molecules, this combination may facilitate crystallisation thereby allowing structural studies to be performed. In addition, centrosymmetric arrangements of racemates greatly aid structure determination using direct methods, simplifying X-ray crystallography by avoiding the need to generate heavy-atom derivatives [9]. Furthermore, the exceptional stability of (L)-RNA to nucleases may prove valuable in extended crystallisation studies.

Site-specific cross-linking and labelling of RNA

Many examples of the site-specific attachment of reporter and reactive groups using phosphoramidite coupling chemistry have been described (e.g. [8,10]). Of particular

Figure 2



The structure of the A-site RNA–paromomycin complex. **(a)** A best-fit superposition of 20 final simulated annealing structures viewed from the major groove of the RNA. Only the core of the antibiotic-binding site within the RNA (nucleotides U1406 to A1410 and U1490 to U1495) is shown in purple and paromomycin is yellow. The heavy atoms of the core RNA and paromomycin have been superimposed. The four rings of paromomycin are numbered. **(b)** A single representative structure of the complex; all heavy atoms are displayed. The RNA is shown in purple and paromomycin is tan. The van der Waals surface of paromomycin is shown and the four rings are numbered as in (a). Figure kindly provided by D. Puglisi and reproduced with permission from [5]. Copyright 1996 American Association for the Advancement of Science.

note, are those phosphoramidites that allow post-synthetic modification and afford the opportunity to attach a wide range of groups using a single phosphoramidite building block. In addition, these groups can bear functionality incompatible with automated RNA synthesis. For example, the selective introduction and modification of 2′-amino groups in oligoribonucleotides has been exploited to attach disulphides to RNA [10]. Hence, any molecule containing a thiol or thiol-reactive electrophile can be linked to the RNA. This could facilitate screen design by allowing incorporation of a fluorophore, biotin or another functional RNA fragment.

X-ray crystallography can also be facilitated by the chemical synthesis of modified RNA. The presence of non-paired bases predisposes RNA to self-aggregation, particularly at the concentrations used during crystallisation. Such heterogeneity may prevent successful crystallisation and structure determination. The ability to selectively cross-link nucleotides can make an RNA pseudocyclic, preventing dimerisation and facilitating crystallographic studies. For example, convertible nucleosides have been used to cross-link an eight-nucleotide mini-stem–loop which retained functional recognition by the enzyme ricin [11]. This same method could be used to differentiate between alternative folding models of RNA fragments. In addition, it can be used to stabilise the termini of short RNA helices instead of adding extra base pairs, as is more usual.

Generic insights from single structures

The X-ray crystal structure of a 160-nucleotide domain of the *Tetrahymena* pre-rRNA group I intron illustrates that a wealth of information can be obtained from a single structure determination [12,13]. There are many features of this structure that can be exploited for target development, including the following.

In the *Tetrahymena* pre-rRNA group I intron, both a tetraloop docking into its receptor and intermolecular hydrogen bonds between the 2′-hydroxyls of adenosines in widened minor grooves may represent motifs that could be introduced into non-functional termini of RNA fragments to increase intermolecular interactions, thereby facilitating crystallisation. Furthermore, the structure of these motifs defines a function that could in turn define therapeutic targets. The 'adenosine platform' structural motif has a distinctive chemical reactivity [13] which will aid its identification in therapeutically important RNA targets. This may also provide universal targets for generic inhibition strategies.

Finally, the divalent metal ion binding sites in the major groove of helices in the *Tetrahymena* group I pre-rRNA intron fragment structure are of considerable interest. Divalent metal ions are directly involved in the catalytic activity of ribozymes such as RNase P, and establishing potential

binding modes of the catalytically essential metal ions could aid the design of antibacterial agents.

Functional subdomains as therapeutic targets

RNA performs a variety of functional roles which are mediated by signals that are a combination of both sequence and structure motifs. Detailed analysis of such RNA–ligand interactions is hampered by the size and complexity of ribonucleoprotein particles and the intrinsic flexibility of RNA. The approaches discussed above, of using smaller functional analogues or restricting the number of conformational states by forming RNA–ligand complexes, are also applicable to this problem.

Large RNP complexes can be fragmented into functional subdomains that are amenable to detailed investigation [14–16]. For example, the bacterial 30S ribosomal subunit naturally contains three subdomains encompassing the 5′, 3′ and central domains [17,18]. The 3′ domain can be further dissected as exemplified by the generation of an oligoribonucleotide analogue encompassing the decoding region located near to the 3′ terminus of 16S rRNA [14]. This RNA analogue has been shown to interact *in vitro* with both aminoglycoside antibiotics and RNA ligands (tRNA and messenger RNA) of the 30S subunit in a manner that resembles normal subunit function [5,14]. Thus, the presence of a sequence that produces a nucleotide bulge results in the formation of an accessible, deep and widened major groove that affords a site for antibiotic binding. These results offer further evidence that therapeutic targets can be discovered by studying defined, functionally relevant RNA fragments.

Cell-based identification of ligand-binding RNA fragments

A whole-cell approach has also been developed to identify RNA fragments that interact with ligands. The 'RNA fragment rescue' concept is based on the premise that expressing an RNA analogue that can bind and thereby sequester an antibiotic will ensure the viability of an otherwise sensitive bacterium [16]. Support for this concept was provided by the expression of a 16S rRNA fragment encompassing the spectinomycin-binding domain that conferred drug resistance. The potential of this strategy was illustrated further by the selection of additional RNA sequence analogues from a mixed pool of RNA fragments that conferred spectinomycin resistance [16]. The analogues shared a common primary sequence and predicted structural motifs that have been proposed to form the recognition domain for spectinomycin.

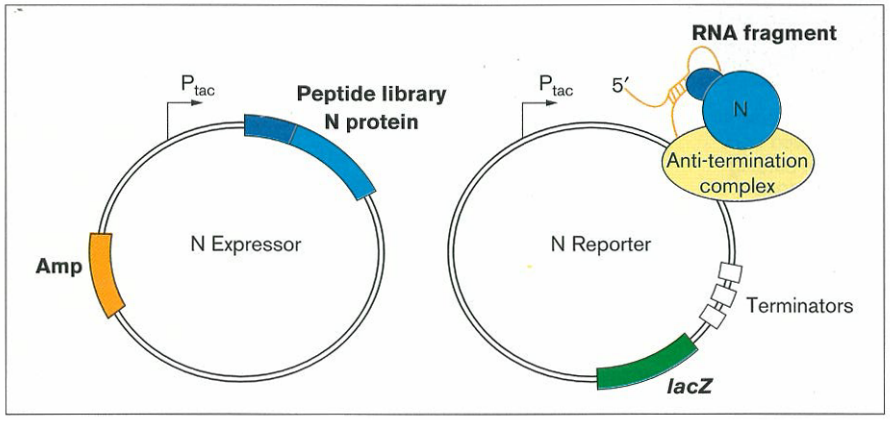
Selection of ligand-binding nucleic acid motifs

Methods that permit rapid screening, selection and identification of the key functional elements involved in intermolecular interactions are clearly advantageous to drug research. SELEX (the systematic evolution of ligands by exponential enrichment) is arguably the most powerful *in vitro* approach that enables the selection of minimal nucleic acid recognition motifs for ligand binding (for review see [19]). A wide variety of molecules have been targeted including nucleic acid binding proteins [19–23], non-nucleic acid binding proteins [24–26] and even antibiotics [27]. For example, SELEX to the HIV-1 Rev protein resulted in RNA ligands with striking sequence and secondary structural similarities to the high-affinity Rev-binding site on the Rev-responsive element (RRE) RNA stem IIB [28,29]. Disruption of the rev-RRE interaction may afford an antiviral agent.

Genetic strategies for defining functional protein–RNA complexes

The lack of information detailing how RNA-binding proteins specifically interact with their target can be attributed to the laborious procedures required to generate large numbers of proteins with defined amino acid substitutions, each of which needs to be purified and individually characterised. Two genetic strategies (translational repression and anti-termination) that permit the rapid screening of large numbers of protein and RNA variants

Figure 3



The two-plasmid system for measuring transcriptional anti-termination by the λ N protein [31]. The λ N protein binds to the box B hairpin of the N utilisation site on the nascent RNA transcript. N, along with other host factors, forms an anti-termination complex that allows RNA polymerase to transcribe through termination sites. In the two-plasmid system, λ N protein is expressed under the control of a tac promoter on a pBR322-derived plasmid and β-galactosidase, also under control of a tac promoter, is expressed on a pACYC-derived reporter plasmid, with Nut and termination sites upstream of *lacZ*. Figure adapted from [32] with permission.

have recently been reported. These strategies can be used to identify the key functional elements that modulate affinity and specificity of protein–RNA binding.

Translational repression arises following the formation of a stable protein–RNA complex that sterically hinders ribosome binding to a Shine–Dalgarno sequence. This concept was adopted to characterise the binding of the HIV Rev protein to HIV RRE [30]. Binding of HIV Rev protein to RRE inserted immediately upstream of the reporter (β-galactosidase) Shine–Dalgarno sequence resulted in decreased levels of reporter expression. A series of mutations introduced into both RRE and Rev were rapidly screened and key residues involved in Rev binding, as well as suppressor mutations in Rev that rescued the binding of altered RRE sequences, were successfully identified. The results supported existing evidence that the arginine-rich α-helical domain of Rev binds RRE and that sequence-specific interactions are mediated through amino acids on the non-arginine face of the helix.

RNA–protein interactions have also been investigated by exploiting the interaction between the amino-terminal domain of the λ N protein and the N utilisation site (Nut) RNA stem–loop that precedes the anti-termination site [31]. The λ N protein–Nut complex signals the recruitment of additional factors that form an anti-termination complex capable of transcribing through termination signals. The system was adapted to identify key residues involved in the recognition of RRE by replacing the amino-terminal portion of the N protein with an arginine-biased peptide library and exchanging the Nut RNA stem–loop with RRE (Figure 3; [32]). Specific RNA–peptide interactions resulted in expression of the β-galactosidase reporter gene located beyond termination sequences. The screen resulted in the selection of peptides that matched the Rev consensus sequence.

Both genetic screens provide a powerful means to rapidly select and identify critical residues involved in protein–RNA interactions and, in so doing, provide complexes suitable for the development of high-throughput screens for drug-lead identification.

Conclusions

The exploitation of RNA as a therapeutic target presents three major issues: size, complexity and intrinsic flexibility. Drug discovery necessitates the collaboration of synthetic chemists, structural chemists and biologists in the definition of the precise nature of drug–RNA complexes. Only then will the derivation of structure–activity relationships, which underpins drug discovery, be reliable.

References

1. Westhof, E., Masquida, B. & Jaeger, L. (1996). RNA tectonics: towards RNA design. *Fold. Des.* **1**, R78-R88.
2. Eckstein, F. & Lilley, D.M.J. (eds) (1996). *Catalytic RNA* Volume 10: *Nucleic Acids and Molecular Biology*. Springer-Verlag, Berlin, Germany.
3. Hecht, S.M. (1994). RNA degradation by bleomycin, a naturally occurring bioconjugate. *Bioconj. Chem.* **5**, 513-526.
4. Nissen, P., *et al.*, & Nyborg, J. (1995). Crystal structure of the ternary complex of Phe-tRNA^{Phe}, EF-Tu, and a GTP analog. *Science* **270**, 1464-1472.
5. Fourmy, D., Recht, M.I., Blanchard, S.C. & Puglisi, J.D. (1996). Structure of the A site of *Escherichia coli* 16S ribosomal RNA complexed with an aminoglycoside antibiotic. *Science* **274**, 1367-1371.
6. Scott, W.G., *et al.*, & Klug, A. (1995). Rapid crystallization of chemically synthesized hammerhead RNAs using a double screening procedure. *J. Mol. Biol.* **250**, 327-332.
7. Scott, W.G., Finch, J.T. & Klug, A. (1995). The crystal structure of an all-RNA hammerhead ribozyme: a proposed mechanism for RNA catalytic cleavage. *Cell* **81**, 991-200.
8. Nolte, A., Klussmann, S., Bald, R., Erdmann, V.A. and Furste, J.P. (1996). Mirror-design of L-oligonucleotide ligands binding to L-arginine. *Nat. Biotechnol.* **14**, 1116-1119.
9. Toniolo, C., Peggion, C., Crisma, M., Formaggio, F., Shui, X. & Eggleston, D.S. (1994). Structure determination of racemic trichogin A IV using centrosymmetric crystals. *Struct. Biol.* **1**, 908-914.
10. Sigurdsson, S.Th. & Eckstein, F. (1996). Site specific labelling of sugar residues in oligoribonucleotides: reactions of aliphatic isocyanates with 2′ amino groups. *Nucleic Acids Res.* **24**, 3129-3133.
11. Allerson, C.R. & Verdine, G.L. (1995). Synthesis and biochemical evaluation of RNA containing an intrahelical disulfide crosslink. *Chem. Biol.* **2**, 667-675
12. Cate, J.H., *et al.*, & Doudna, J.A. (1996). Crystal structure of a group I ribozyme domain: principles of RNA packing. *Science* **273**, 1678-1685.
13. Cate, J.H., *et al.*, & Doudna, J.A. (1996). RNA structure mediation by adenosine platforms. *Science* **273**, 1696-1699.
14. Purohit, P. & Stern, S. (1994). Interactions of a small RNA with antibiotic and RNA ligands of the 30S subunit. *Nature* **370**, 659-662.
15. Schroeder, R. (1994). Dissecting RNA function. *Nature* **370**, 597-598.
16. Thom, G. & Prescott, C.D. (1997). The selection and characterization of an RNA recognition motif for spectinomycin. *Bioorg. Med. Chem.*, in press.
17. Yuki, A. & Brimacombe, R. (1975). Nucleotide sequences of *Escherichia coli* 16-S RNA associated with ribosomal proteins S7, S9, S10, S14 and S19. *Eur. J. Biochem.* **56**, 23-34.
18. Samaha, R.R., O'Brien, T., O'Brien, B. & Noller, H.F. (1994). Independent *in vitro* assembly of a ribonucleoprotein particle containing the 3′ domain of 16S rRNA. *Proc. Natl Acad. Sci. USA* **91**, 7884-7888.
19. Gold, L. (1995). Conformational properties of oligonucleotides. *Nucleic Acids Symp. Ser.* **33**, 20-22.
20. Tuerk, C. & Gold, L. (1990). Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* **249**, 505-510.
21. Bartel, D.P., Zapp, M.L., Green, M.R. & Szostak, J.W. (1991). HIV-1 Rev regulation involves recognition of non-Watson-Crick base pairs in viral RNA. *Cell* **67**, 529-536.
22. Schneider, D., Tuerk, C. & Gold, L. (1992). Selection of high affinity RNA ligands to the bacteriophage R17 coat protein. *J. Mol. Biol.* **228**, 862-869.
23. Tuerk, C., MacDougal, S. & Gold, L. (1992). RNA pseudoknots that inhibit human immunodeficiency virus type 1 reverse transcriptase. *Proc. Natl Acad. Sci. USA* **89**, 6988-6992.
24. Jellinek, D., Lynott, C.K., Rifkin, D.B. & Janjic, N. (1993). High-affinity RNA ligands to basic fibroblast growth factor inhibit receptor binding. *Proc. Natl Acad. Sci. USA* **90**, 11227-11231.
25. Nieuwlandt, D., Wecker, M. & Gold, L. (1995). *In vitro* selection of RNA ligands to substance P. *Biochemistry* **34**, 5651-5659.
26. Binkley, J., Allen, P., Brown, D.M., Green, L., Tuerk, C. & Gold, L. (1995). RNA ligands to human nerve growth factor. *Nucleic Acids Res.* **23**, 3198-3205.
27. Wallis, M.G., Von Ahsen, U., Schroeder, R. & Famulok, M. (1995). A novel RNA motif for neomycin recognition. *Chem. Biol.* **2**, 543-552.
28. Heaphy, S., Dingwall, C., Ernberg, I., Gait, M. & Skinner, M.A. (1990). HIV-1 regulator of virion expression (Rev) protein binds to an RNA stem-loop structure located within the Rev response element region. *Cell* **60**, 685-693.

29. Heaphy, S., Finch, J.T., Gait, M., Karn, J. & Singh, M. (1991). Human immunodeficiency virus type 1 regulator of virion expression, Rev, forms nucleoprotein filaments after binding to a purine-rich 'bubble' located within the Rev-responsive region of viral mRNAs. *Proc. Natl Acad. Sci. USA* **88**, 7366-7370.
30. Jain, C. & Belasco, J.G. (1996). A structural model for the HIV-1 Rev-RRE complex deduced from altered-specificity Rev variants isolated by a rapid genetic strategy. *Cell* **87**, 115-125.
31. Franklin, N.C. (1993). Clustered arginine residues of bacteriophage lambda N protein are essential to antitermination of transcription, but their locale cannot compensate for boxB loop defects. *J. Mol. Biol.* **231**, 343-360.
32. Harada, K., Martin, S.S. & Frankel, A.D. (1996). Selection of RNA-binding peptides *in vivo*. *Nature* **380**, 175-179.