In Vitro Selected Receptors Rationalized: The First 3D Structures of RNA Aptamer/Substrate Complexes**

Martin Egli*

The development of new drugs is nowadays often preceded by the determination of the three-dimensional structure of the receptor targeted. If the specific receptor is a protein, knowledge of the geometrical features of its binding site and the distribution of the chemical groups docking the ligand can provide valuable input for a host of approaches, including computational methods, organic synthesis, and combinatorial chemistry, all aimed at identifying the optimal inhibitor. Such structure-based design contributed to the development of HIV protease inhibitors, which promise dramatic improvement in the treatment of AIDS. Although this approach is far from becoming routine, the opposite strategy, namely tailoring a receptor for a chemically and structurally well-characterized molecule, is an even more formidable challenge. Particularly in an aqueous environment, mastering the many small contributions from hydrogen bonds and hydrophobic interactions is very difficult. With the advent of in vitro selection, advanced by tools such as polymerase chain reaction (PCR) and reverse transcription, it has become possible to isolate ribonucleic acid receptors with a desired property, ranging from the specific recognition of a chosen substrate to the ability to catalyze a particular chemical reaction.[1-3]

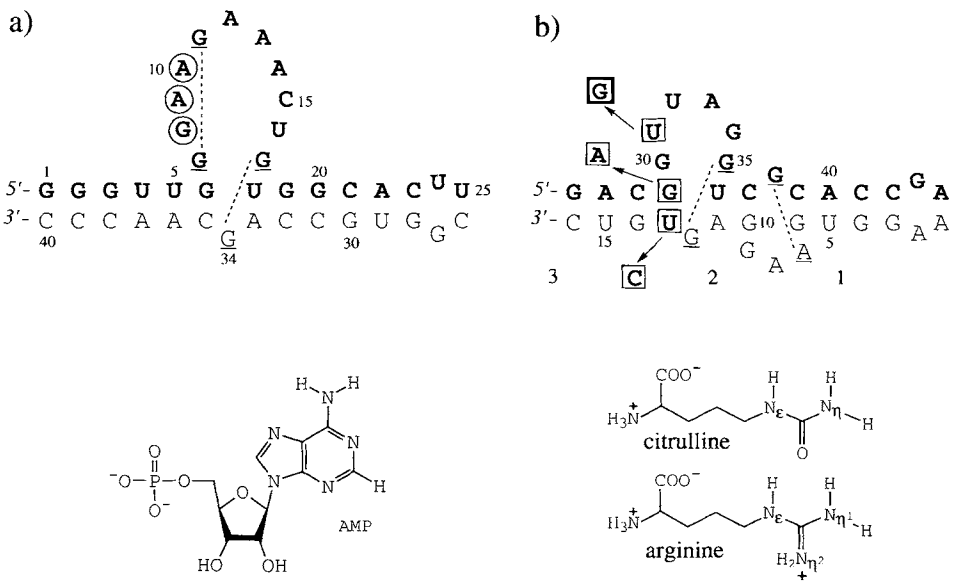

Figure 1. Base sequence, numbering, and proposed secondary structure (now essentially confirmed by solution NMR experiments) for the RNA aptamer of ATP (a) and of arginine and citrulline (b). Bases of one strand are in bold to facilitate tracing of the folded chain depicted in Figures 2 and 3. Dashed lines indicate the formation of noncanonical homopurine base pairs in G·G and G·A (these bases are underlined) in the 3D structures. In the ATP aptamer (a) the circled nucleotides form a GNRA tetraloop motif together with the AMP substrate. In (b) the three variant nucleotides that are critical for arginine and citrulline specificity, respectively, are boxed, and the numbers 1–3 under the receptor mark the stem regions.

Starting from randomized pools with roughly 10^{10} to 10^{15} molecules, RNA and DNA species have been recovered that recognize simple dye molecules, biologically important cofactors, amino acids, and antibiotics.[4] The list of selected catalytically active nucleic acid molecules includes RNA with polynucleotide kinase activity,[5] a self-alkylating ribozyme,[6] RNAs and DNAs with ligase activity,[7-9] and both RNAs and DNAs that undergo metal ion mediated autolytic cleavage.[10-12] Selected and optimized RNA species that are able to bind small organic molecules with dissociation constants in the mM to nM range, so-called RNA aptamers, are quite compact, with lengths of around 40 nucleotides, and their deduced secondary structures display some common features, notably asymmetric internal loops (see Figure 1). Although chemical methods are very effective in determining which portions of the RNA binder are essential for substrate recognition, they are not suited for establishing the details of the chain folding and the binding site geometry. Recently, four NMR studies of RNA–substrate complexes in solution have opened the door to a 3D view of RNA's aptameric repertoire. Three of these studies were conducted on binders of the biologically important cofactors adenosine triphosphate (ATP) (Figure 1a, two independent studies[13,14]) and flavin mononucleotide (FMN),[15] which were previously selected in the research groups led by Jack Szostak,[16] and Michael Famulok,[17] respectively. The fourth study concerns RNA aptamers for the amino acids citrulline and arginine,[18] which were selected earlier on by the latter researchers (Figure 1b).[19]

The solution NMR structure of the complex between the ATP aptamer and the adenosine monophosphate (AMP) substrate reveals that the RNA molecule assumes an L shape (Figure 2). The internal loop folds into two tightly packed hairpin loops, each capped by a noncanonical G·G base pair; one of these guanine residues is the bulged G34 (Figure 1a). The binding

[*] Prof. M. Egli
Department of Molecular Pharmacology & Biological Chemistry
Northwestern University Medical School
303 East Chicago Avenue, Chicago, IL 60611-3008 (USA)
Fax: Int. code +(312)503-0796
e-mail: m-egli@nwu.edu

[**] I am grateful to Dr. Michael Famulok and Prof. Dinshaw Patel for sending me the coordinates of the aptamer complexes of arginine/citrulline and ATP, respectively.

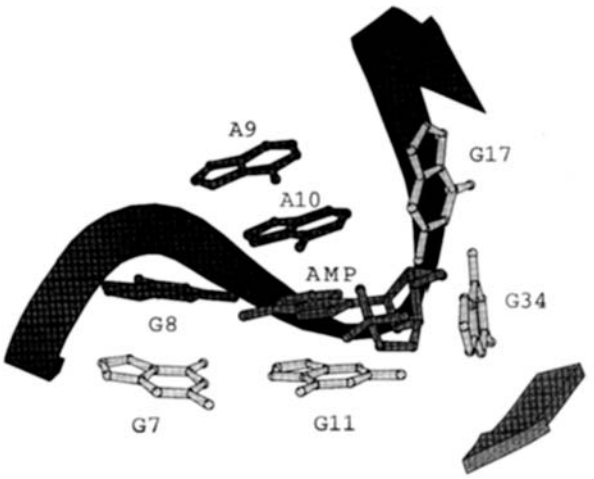

Figure 2. Sketch of the substrate binding site of the ATP aptamer with the bound AMP molecule (center). The RNA backbones are drawn as ribbons: residues G6 to U18 are dark gray, and A33 to C35 are light gray. The folding of the large portion of the asymmetric loop into two adjacent hairpins is clearly visible. The substrate binding site is located at their junction, and the AMP is part of a stable G8-A9-A10-A(MP) loop motif (bases are dark gray). The G·G base pairs capping each hairpin and enclosing the substrate are light gray.

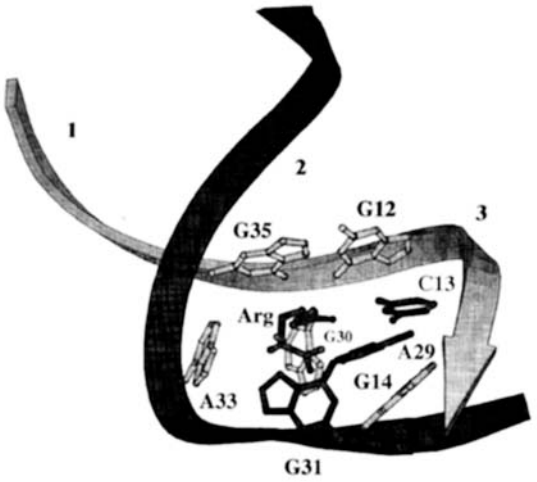

Figure 3. Sketch of the substrate binding site of the arginine aptamer with the bound arginine molecule (center). The RNA backbones are drawn as ribbons: residues A27 to A40 are dark gray, and U5 to U15 are light gray. The large segment of the asymmetric internal loop, visible on the left in the foreground, faces the minor groove formed by stems 2 and 3. The three variant residues 13, 29, and 31 are dark gray, and other residues defining the binding pocket, including the G·G base pair that caps the internal loop, are light gray.

pocket for the substrate, which is located virtually at the junction between the two hairpin loops, is formed by residue A10 from above and residues G7 and G11 from below. The latter two bases form an asymmetic homopurine pair; G7 faces the Hoogsteen edge of G11 with its Watson–Crick edge. The walls of the binding pocket are provided by residues G8, A12 (not shown in Figure 2), and G17. The first residue is paired with the adenine base of the substrate through two hydrogen bonds (to N1 and N6 of AMP); the exocyclic amino groups of A12 and G17 each form a hydrogen bond with N3 of AMP. While A12 is stacked onto G34, G17 and G34 form a Hoogsteen-type pair with G34 adopting a *syn* conformation. In addition to these specific hydrogen bonding contacts to the base portion of the substrate, several stacking interactions contribute to the stabilization of the AMP molecule in the binding pocket. The most remarkable feature of this aptamer complex is that the RNA ties the substrate into a thermodynamically stable GNRA loop motif (N = any base, R = purine), which includes G8, A9, A10, and the adenine residue of AMP (Figures 1a and 2). The RNA has found an ingenious way here to stabilize the substrate by letting it take care of the formation of at least part of its own binding site.

Selection of the L-arginine binder was achieved with an RNA pool predisposed for L-citrulline, and the RNA molecule acquired arginine specificity by mutation at three nucleotide positions. One of these sites was located in the longer portion of the internal loop (U31 to G31) and the others capped that loop at the adjacent stem 3 (G29·U13 to A29·C13, Figure 1b). In both complexes, the Watson–Crick and Hoogsteen edges of bases from residues of this loop region face the virtually continuous minor groove formed by the coaxially stacked stems 2 and 3 (Figures 1b and 3) and thus generate a pocket for the substrate. Nucleotides G9 (not shown in Figure 3), G12, and G35 form the roof of the cavity, G14 and G(U)31 insert themselves underneath, and C(U)13 and A33 provide the walls, while G30 seals off the back.

Two of the variant residues in the aptamers, C(U)13 and G(U)31, are in direct contact with the guanidino and urea portions of the arginine and citrulline substrates, respectively. However, the orientations of these substrate moieties relative to the minor groove (stems 2 and 3) are quite different. The guanidino group of arginine is practically coplanar with the base plane of residue C13 in stem 3, but the urea part of citrulline is rotated by roughly 90° and is thus nearly perpendicular to the base plane of residue U13. Nucleotide 13 is directly involved in the specific recognition of the substrate, and the O2 and N3 atoms of C13 in the arginine aptamer form hydrogen bonds with the protons at N_ε and $N_{\eta 1}$, respectively. In the citrulline complex, the altered arrangement of hydrogen bond donors and acceptors in the substrate is now matched by U13, which forms contacts to N_ε through its O2 atom and to the carbonyl oxygen through its N3 atom. Nucleotide 13 in turn is held in place by a hydrogen bond to the covariant nucleotide 29. Although the substrate moieties are rotated relative to one another in the binding pockets of the two complexes, the base planes of residues G31 and U31 have similar orientations within the pocket. Thus, the different frameworks of guanine and uracil allow an optimal overlap between the lone pair orbitals of their carbonyl oxygens and the π^* orbitals of the guanidino and urea groups, respectively, providing similar stabilizing contributions in the two aptamer complexes. These structures have clearly revealed how RNA can complex small-molecule ligands with high stability. They have also provided insight into the structural origins of recognition by directly linking variant positions to substrate-specific hydrogen bonding interactions and stereoelectronic effects.

The structures of the ATP and amino acid aptamers demonstrate two radically different ways of how RNA uses one of its common secondary structure elements, an asymmetric internal loop, to acquire a specific function, namely recognizing a small, charged organic molecule by means of hydrophilic and hydrophobic interactions. Both internal loops are rich in purines, many of which were conserved among the molecules isolated

  0570-0833/97/3605-0481 $ 15.00+.25/0

after several rounds of selection. The loop of the ATP aptamer is longer than that of the arginine aptamer, allowing this binder to adopt more complex folds. Nevertheless, the actual double hairpin fold displayed by the ATP aptamer is unexpected, and it would have been clearly impossible to predict the formation of the prominent hairpin motif by the GAA segment together with the substrate.

The structure of the flavin mononucleotide (FMN) binder reveals yet another way of accommodating a substrate: adoption of a noncanonical duplex by an internal loop region.[15, 17] There, formation of a stretch of noncanonical purine–purine base pairs widens the major groove but leaves stacking intact and generates an intercalation site for the FMN chromophore. The aromatic isoalloxazine system is flanked by a base triple and a G·G base pair, resulting in extensive stacking interactions between substrate and RNA. Additional stability is provided by the formation of two hydrogen bonds between the uracyl-like edge of the intercalated isoalloxazine ring with the Hoogsteen edge of an adenosine residue. Thus, functionally important homopurine base pairs, particularly between guanines, are found in all three RNA aptamer complexes. A common feature of the ATP and arginine/citrulline binders is the recruitment of the bulged guanine residue facing the large portion of the loop. A G·G pair forms, sealing the loop and extending base stacking beyond regular Watson–Crick paired stem regions.

Perhaps not surprisingly, the regions in the substrates that surround the site used to immobilize them on the column during selection are not contacted by the RNA. Thus, the N7 and C8 atoms of the AMP molecule are facing the solvent and the α-amino groups of arginine and citrulline are only loosely bound. It is noteworthy that there is no contact between the RNA aptamer and the phosphate group of AMP. One could argue that the actual interaction mode may depend on the concentration of metal cations in the selection experiments and that the phosphate groups in the folded RNA backbone may not tolerate the accomodation of the substrate phosphate in a binding cleft. However, the stacking of a phosphate group on an adjacent base in the stable U-turn motif in tRNA and the hammerhead ribozyme demonstrates convincingly that RNA can deal with such a situation as well. Moreover, closely spaced phosphate groups are tolerated in alternative nucleic acid structures when compensated for by a variety of stabilizing contributions.[20] Another notable feature of the arginine aptamer complex is the absence of an interaction between the guanidino group of the substrate and a phosphate. This is quite different from the situation in protein–DNA complexes, where such interactions occur frequently and help to orient a variety of sequence-recognizing protein folding motifs with respect to the DNA major and minor grooves. The simultaneous contacting of the guanidino system by several bases in the RNA aptamer complex is also strikingly different from the more common, sequence-specific, 1:1 interaction between arginine and the Hoogsteen side of guanine in protein–DNA complexes. This demonstrates a fundamental difference between protein–DNA and protein–RNA interactions. Rather than merely probing for sequence, proteins may recognize RNA through a certain three-dimensional structural motif or folding pattern.

What, then, have we learned from these structures? Seemingly similar RNA secondary structure motifs may generate divergent tertiary and quarternary structures. The backbones in internal loops can wind around a stem in a continuous manner, be sequestered into subdomains, or be zippered up by means of noncanonical base pairs into a stem structure. Just as the clover leaf diagram for tRNA only faintly resembles the actual 3D shape of the molecule, loops may confer different degrees of kinking and result in a range of topologies. However, certain motifs are recurring, despite the small number of available 3D structures of RNA, for example the GNRA loop and a variety of homopurine base pairs at functionally important sites. As might have been expected, RNA and proteins both reach into the same toolbox to master substrate binding and recognition. The formation of the GAAA loop is only possible when the substrate is docked and is thus an induced fit mechanism. Another similarity is the use of the same structural motif; with perhaps a small variation, in an entirely different functional context. Still, the 3D conformational versatility of RNA remains largely unexplored. The door has just been opened a crack; now let's push it wide open!

German version: *Angew. Chem.* **1997**, *109*, 494–497

Keywords: hydrogen bonding · molecular recognition · RNA aptamers

[1] A. D. Ellington, J. W. Szostak, *Nature* **1990**, *346*, 818–822.
[2] D. P. Bartel, J. W. Szostak, *Science* **1993**, *261*, 1411–1418.
[3] G. F. Joyce, *Curr. Opin. Struct. Biol.* **1994**, *4*, 331–336.
[4] M. Famulok, J. W. Szostak, *Angew. Chem.* **1992**, *104*, 1001–1011; *Angew. Chem. Int. Ed. Engl.* **1992**, *31*, 979–988; K. W. Uphoff, S. D. Bell, A. D. Ellington, *Curr. Opin. Struct. Biol.* **1996**, *6*, 281–288.
[5] J. R. Lorsch, J. W. Szostak, *Nature* **1994**, *371*, 31–36.
[6] C. Wilson, J. W. Szostak, *Nature* **1995**, *374*, 777–782.
[7] E. H. Ekland, J. W. Szostak, D. P. Bartel, *Science* **1995**, *269*, 364–370.
[8] B. Cuenoud, J. W. Szostak, *Nature* **1995**, *375*, 611–614.
[9] E. H. Ekland, D. P. Bartel, *Nature* **1996**, *382*, 373–376.
[10] D. L. Robertson, G. F. Joyce, *Nature* **1990**, *344*, 467–468.
[11] T. Pan, O. C. Uhlenbeck, *Nature* **1992**, *358*, 560–563.
[12] R. R. Breaker, G. F. Joyce, *Chem. & Biol.* **1995**, *2*, 655–660.
[13] F. Jiang, R. A. Kumar, R. A. Jones, D. J. Patel, *Nature* **1996**, *382*, 183–186.
[14] T. Dieckmann, E. Suzuki, G. K. Nakamura, J. Feigon, *RNA* **1996**, *2*, 628–640.
[15] P. Fan, A. K. Suri, R. Fiala, D. Live, D. J. Patel, *J. Mol. Biol.* **1996**, *258*, 480–500.
[16] M. Sassanfar, J. W. Szostak, *Nature* **1993**, *364*, 550–553.
[17] P. Burgstaller, M. Famulok, *Angew. Chem.* **1994**, *106*, 1163–1167; *Angew. Chem. Int. Ed. Engl.* **1994**, *33*, 1084–1087.
[18] Y. Yang, M. Kochoyan, P. Burgstaller, E. Westhof, M. Famulok, *Science* **1996**, *272*, 1343–1346.
[19] M. Famulok, *J. Am. Chem. Soc.* **1994**, *116*, 1698–1706.
[20] I. Berger, M. Egli, A. Rich, *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 12116–12121.

  0570-0833/97/3605-0482 $ 15.00+.25/0