Formation of an amino-acid-binding pocket through adaptive zippering-up of a large DNA hairpin loop

Chin H Lin[1], Weimin Wang[2], Roger A Jones[2] and Dinshaw J Patel[1]

Background: *In vitro* selection has identified DNA aptamers that target cofactors, amino acids, peptides and proteins. Structure determination of such ligand–DNA aptamer complexes should elucidate the details of adaptive DNA structural transitions, binding-pocket architectures and ligand recognition. We have determined the solution structure of the complex of a DNA aptamer containing a guanine-rich 18-residue hairpin loop that binds L-argininamide with ~100μM affinity.

Results: The DNA aptamer generates its L-argininamide-binding pocket by adaptive zippering up the 18-residue loop through formation of Watson–Crick pairs, mismatch pairs and base triples, while maximizing stacking interactions. Three of the four base triples involve minor-groove recognition through sheared G•A mismatch formation. The unique fold is also achieved through positioning of an adenine residue deep within the minor groove and through nestling of a smaller loop within the larger loop on complex formation. The accessibility to the unique L-argininamide-binding pocket is restricted by a base pair that bridges across one side of the major-groove-binding site. The guanidinium group of the bound L-argininamide aligns through intermolecular hydrogen-bond formation with the base edges of nonadjacent guanine and cytosine residues while being sandwiched between the planes of nonadjacent guanine residues.

Conclusions: The available structures of L-arginine/L-argininamide bound to their DNA and RNA targets define the common principles and patterns associated with molecular recognition, as well as the diversity of intermolecular hydrogen-bonding alignments associated with the distinct binding pockets.

Addresses: [1]Cellular Biochemistry and Biophysics Program, Memorial Sloan-Kettering Cancer Center, New York, NY, 10021, USA. [2]Department of Chemistry, Rutgers, The State University of New Jersey, Piscataway, NJ, 08855, USA.

Correspondence: Dinshaw J Patel
E-mail: pateld@mskcc.org

Key words: adaptive DNA structural transitions, L-argininamide-binding pocket, minor-groove base encapsulation, sheared G•A triple alignments

Received: **27 July 1998**
Accepted: **17 August 1998**

Published: **21 September 1998**

Chemistry & Biology October 1998, **5**:555–572
http://biomednet.com/elecref/1074552100500555

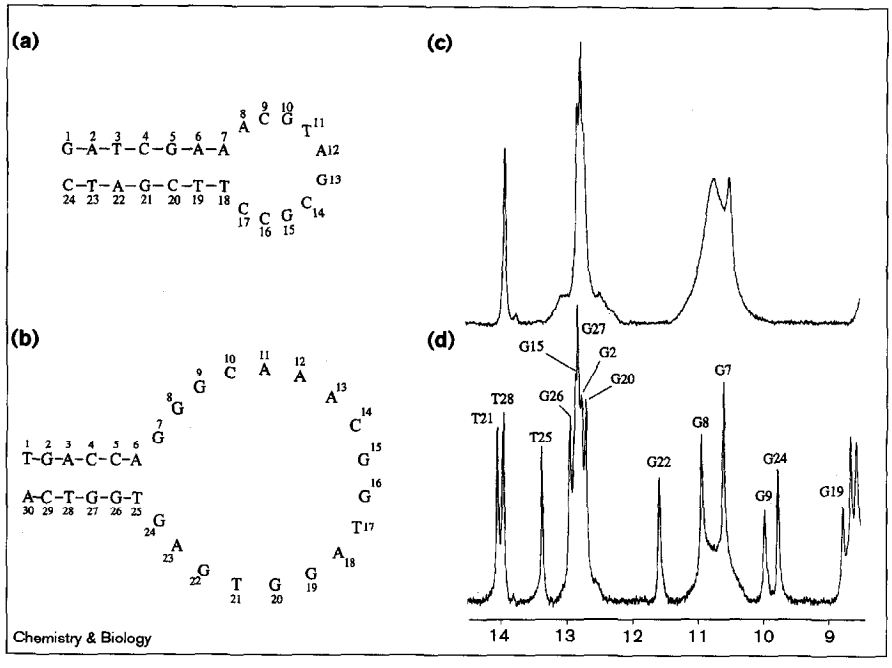
Introduction

Our laboratory is interested in defining the principles, patterns and diversity associated with the molecular recognition of higher order DNA and RNA folds. The potential for identifying novel folds for structural characterization has been greatly aided by the development of *in vitro* selection and evolution methodologies (reviewed in [1–4]). Such approaches identify unique nucleic-acid folds from random libraries of ~10^{14} RNA and DNA molecules on the basis of their ability to complex specific ligands of interest with high affinity and selectivity. Both RNA and DNA aptamers have been identified that target cofactors, antibiotics, amino acids, peptides and proteins [1–4], which in turn have been systematically evolved to generate catalytic ribozymes (reviewed in [5]) and DNAzymes (reviewed in [6]). The application of nuclear magnetic resonance (NMR)-based methods have contributed to our understanding of the structural basis for ligand–nucleic-acid aptamer recognition (reviewed in [7]), with the nucleic acids undergoing novel adaptive structural transitions on complex formation.

More literature is available to date on RNA aptamer complexes than on their DNA aptamer complex counterparts. Single-stranded DNA can also adopt a range of unusual folds aligned through base pair, mismatch, triple and tetrad formation, however, and much remains to be learnt about the architecture of DNA beyond the Watson–Crick double helix. The structural basis for molecular recognition in DNA aptamer complexes is a challenge worth pursuing, given the recent scientific contributions in the area of RNA-cleaving [8] and DNA-cleaving [9] divalent cation-dependent DNA enzymes identified through *in vitro* selection methodology.

The early research on DNA recognition has focused on the selection and structural characterization of complexes of DNA aptamers that target thrombin [10,11] and ATP [12,13]. More recently, two DNA aptamers with stem–loop secondary folds have been identified on the basis of their ability to target L-argininamide with ~100 μM affinity [14]. Our group has previously solved the solution structure of L-argininamide bound to one of these DNA aptamers (a 24-mer stem–loop sequence containing a ten residue hairpin loop, Figure 1a) [15] and now report the solution structure of L-argininamide bound to the second of these DNA aptamers (a 30-mer stem–loop

Figure 1



The sequence and numbering system of the L-argininamide binding **(a)** 24-mer and **(b)** 30-mer DNA aptamers [14]. The exchangeable proton NMR spectrum (8.5 to 14.5 ppm) of **(c)** the free 30-mer DNA aptamer and **(d)** the complex with L-argininamide (four equivalents of L-argininamide per DNA aptamer) in H_2O buffer, pH 6.4 at 4°C.

sequence containing a guanine rich 18-residue hairpin loop, Figure 1b). The 24-mer DNA aptamer contains a ten residue hairpin loop (Figure 1a), whereas the 30-mer DNA aptamer contains an 18-residue hairpin loop (Figure 1b). This offers the opportunity to compare the binding pockets and recognition elements in these two L-argininamide–DNA aptamer complexes and in turn compare them with the solution structures of the corresponding L-arginine–RNA aptamer [16,17] and L-argininamide–TAR RNA [18,19] complexes.

Results

Proton NMR spectra

The binding of L-argininamide to the 30-mer DNA aptamer (Figure 1b) exhibits a weak binding affinity of ~100 μM affinity [14] and hence full complex formation required addition of up to four equivalents of ligand per DNA aptamer. The proton NMR spectra (8.5 to 14.5 ppm) of the free 30-mer DNA aptamer and its complex on addition of four equivalents of L-argininamide in H_2O buffer, pH 6.4 at 4°C are plotted in Figure 1c and 1d respectively. We observe narrow imino protons from the stem region (12.0 to 14.0 ppm) and broad imino protons from the loop region (10.0 to 11.5 ppm) for the free 30-mer DNA aptamer (Figure 1c), indicative of an unstructured hairpin loop in the absence of bound ligand. By contrast, 14 narrow imino protons (including six between 8.5 and 12.0 ppm) are observed for the bound 30-mer DNA aptamer complex (Figure 1d), indicative of a structured DNA hairpin loop on complex formation. The exchange between free and L-argininamide bound 30-mer DNA aptamers was slow on the NMR time scale (monitored by DNA aptamer imino proton spectra) at 4°C.

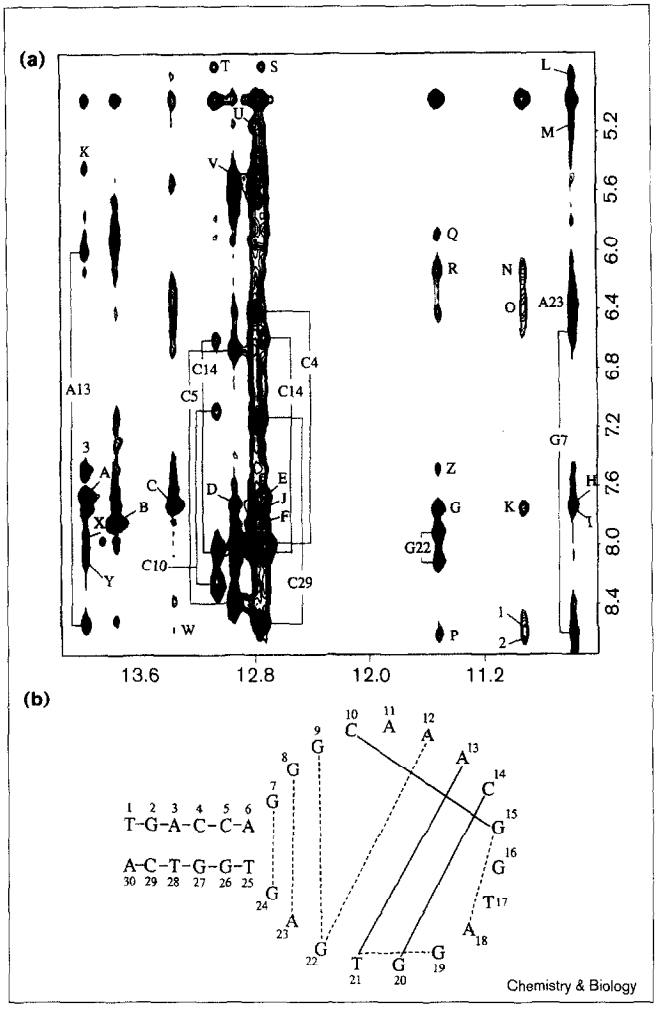
Imino proton spectra and restraints

The unambiguous assignments of the imino protons listed over the spectrum of the argininamide–DNA aptamer complex (Figure 1d) was achieved through a combination of approaches. The analysis of nuclear Overhauser effect spectroscopy (NOESY) data sets on the argininamide–DNA aptamer complex were supplemented by site-specific incorporation of ^{15}N-labeled purine residues and extensive use of primarily base-modified analogs at defined sites in the DNA aptamer.

Watson–Crick alignments between loop residues on complex formation

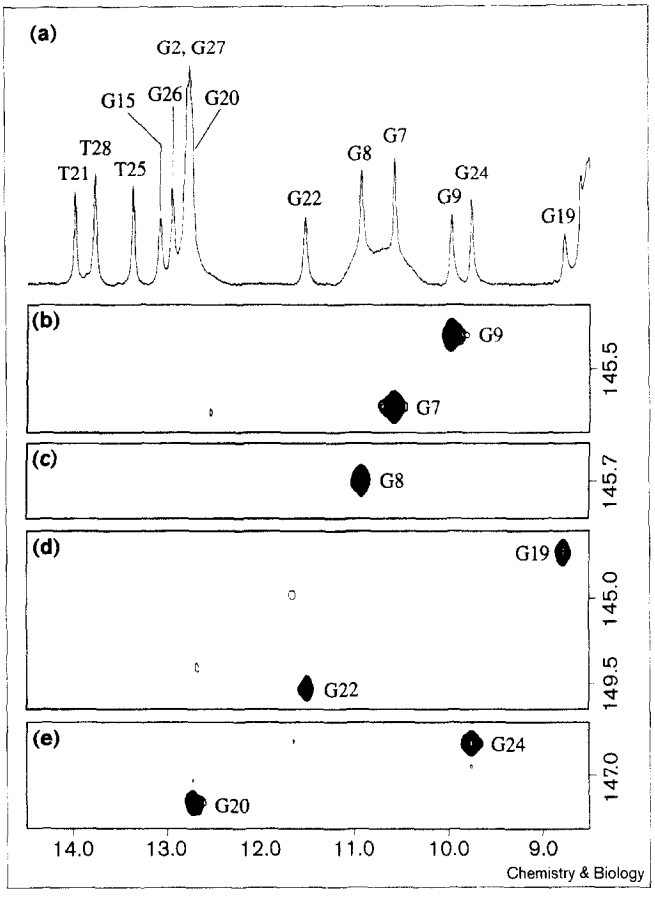
An expanded NOESY contour plot correlating NOEs between imino protons (10.4 to 14.2 ppm) and base and amino protons (4.6 to 8.8 ppm) of the argininamide–DNA aptamer complex in H_2O buffer at 4°C is plotted in Figure 2a. The observation of NOEs between guanine imino protons and cytidine amino protons are indicative of Watson–Crick G•C pair formation and such connectivities are observed for G2•C29, G27•C4 and G26•C5 in the stem segment and for G15•C10 and G20•C14 in the loop segment of the complex (these NOEs are labeled by cytidine residues in Figure 2a). The observation of NOEs between thymine imino protons and adenine amino and H2 protons are indicative of Watson–Crick A•T pair formation and such connectivities are observed for A3•T28 and A6•T25 in the stem segment and A13•T21 in the

Figure 2



(a) Expanded NOESY (135 ms mixing time) contour plots of the argininamide–DNA aptamer complex (four equivalents of L-argininamide per DNA aptamer) in H_2O buffer, pH 6.4 at 4°C. The expanded region correlates NOEs between the imino protons (10.4 to 14.2 ppm) and the amino and nonexchangeable base protons (4.6 to 8.8 ppm). The cross-peaks between guanine imino protons and cytosine amino protons across Watson–Crick G•C base pairs are labeled by residue (C4, C5, C10, C14 and C29) as are those between thymine imino proton and adenine amino protons across Watson–Crick A•T base pairs (A13). The cross peaks between guanine imino protons and their own amino protons are also labeled (G7 and G22). The labeled cross-peaks A to Z and 1 to 3 are assigned as follows: A, T21(NH3)-A13(H2); B, T28(NH3)-A3(H2); C, T25(NH3)-A6(H2); D, G26(NH1)-A6(H2); E, G20(NH1)-A13(H2); F, G2(NH1)-A3(H2); G, G22(NH1)-A23(H2); H, G7(NH1)-A6(H2); I, G7(NH1)-A23(H2); J, G2(NH1)-A30(H2); K, G8(NH1)-A23(H2); L, G7(NH1)-G24(H1′); M, G7(NH1)-G24(H8); N, G8(NH1)-T21(H1′); O, G8(NH1)-A11(H1′); P, G22(NH1)-A11(H8); Q, G22(NH1)-A23(H1′); R, G22(NH1)-T21(NH1′); S, G20(NH1)-C10(H5); T, G15(NH1)-C10(H5); U, G27(NH1)-C4(H5); V, G26(NH1)-C5(H5); W, T25(NH3)-G7(NH_{2b}); X, T21(NH3)-G22(NH_{2e}); Y, T21(NH3)-G22(NH_{2b}); Z, G22(NH1)-A12(NH_2); 1, G8(NH1)-G7(NH_{2b}); 2, G8(NH1)-A11(H8); 3, T21(NH3)-A12(NH_2).
(b) A schematic of the adaptive hydrogen-bonding alignments within the large hairpin loop of the DNA aptamer on complex formation with L-argininamide. Watson–Crick alignments are shown by full lines and mismatch alignments are shown by dashed lines.

Figure 3



An approach to imino and amino proton assignments in the argininamide–DNA aptamer complex based on site-specific incorporation of 1, 2, 7-^{15}N labeled guanine (designated #G), 2-^{13}C, 1, 2, 7-^{15}N labeled guanine (designated ##G) and 1, 2, 6-^{15}N labeled adenine (designated #A) into either T1–G15 or G16–A30 segments of the DNA aptamer. **(a)** The exchangeable proton NMR spectrum (8.5 to 14.5 ppm) of the argininamide–DNA aptamer (T1–G15 + G16–A30) complex (four equivalents of L-argininamide per DNA aptamer) in H_2O buffer, pH 6.4 at 4°C. Expanded segments of ^{1}H-^{15}N HSQC spectra on the following labeled samples: **(b)** ##G7, #G9 and #A11, **(c)** #G8 and #A13, **(d)** #G19, ##G22 and #A23, **(e)** #G20, #G24 and #A18.

loop segment of the complex (relevant cross peaks for NOEs to adenine H2 protons are labeled A to C in Figure 2a). Thus, complex formation results in an adaptive structural transition in the hairpin loop which involves formation of two G•C and one A•T Watson–Crick base pairs (shown schematically by solid lines in Figure 2b).

Site-specific incorporation of ^{15}N-labeled purines

A key determinant in resolving potential ambiguities in imino proton assignments was achieved following successful generation of the L-argininamide complex with the 30-mer DNA aptamer cleaved between residues G15 and G16 at the tip of the large hairpin loop. This enabled us to site-specifically incorporate ^{15}N-labeled guanines and adenines within either the T1 to G15 or G16 to A30 15-mer

Table 1

Summary of single-base substitution analogs of the argininamide–DNA aptamer complex that were investigated in this study to confirm proton assignments.

	A	I	U	O^6Me	2′-OMe	N^6Me	8-Br	Deaza	^{5}mC	Neb*	2-AA†	2-AP‡
G7	–	+/–		–				+/–				–
G8		–		+/–								+
G9		–										+
C10									+			
A11					–		–	–		–	–	
A12								–		+		
A13						–		+		–	+	
C14												
G15		–										
G16		+										
T17			+									
A18						–				–		
G19		+/–		–								–
G20	–	–										
T21			+									
G22	–	–					+					–
A23						+/–	–	–		+		
G24		+		–			–	–				

Substitutions that resulted in complex formation are represented by +, those that resulted in weak binding are represented by +/–, and those that resulted in no binding by –. *Neb, Nebularine; †2-AA, 2-aminoadenine; ‡2-AP, 2-aminopurine.

sequences. The three types of labeled residues involved site-specific incorporation of 1,2,7-^{15}N-labeled guanine (designated $^{\#}$G), 2-^{13}C, 1,2,7-^{15}N-labeled guanine (designated $^{\#\#}$G) and 1,2,6-^{15}N-labeled adenine (designated $^{\#}$A) into either T1 to G15 or G16 to A30 segments of the DNA aptamer. Expanded segments of ^{1}H-^{15}N heteronuclear single quantum coherence (HSQC) spectra containing labels at $^{\#\#}$G7, $^{\#}$G9 and $^{\#}$A11 in the T1 to G15 fragment (Figure 3b), $^{\#}$G8 and $^{\#}$A13 in the T1 to G15 fragment (Figure 3c), $^{\#}$G19, $^{\#\#}$G22 and $^{\#}$A23 in the G16 to A30 fragment (Figure 3d) and $^{\#}$G20, $^{\#}$G24 and $^{\#}$A18 in the G16 to A30 fragment (Figure 3e) readily permit definitive assignment of the labeled guanine imino (and amino, data not shown) protons in the spectrum of the argininamide–DNA aptamer complex (Figure 3a).

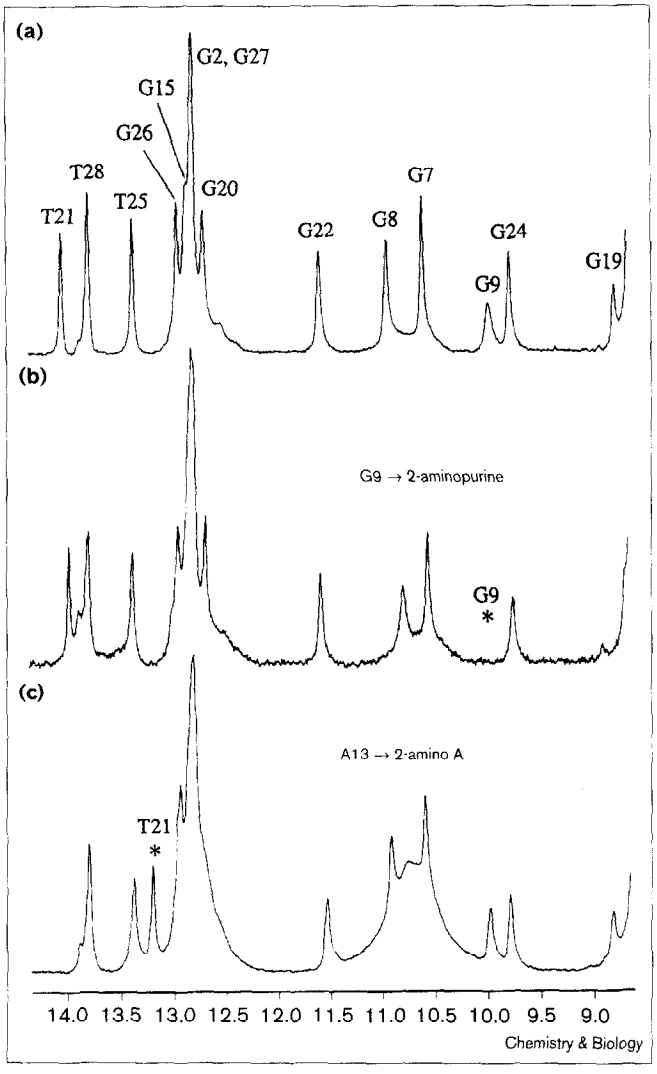
Site-specific incorporation of base analogs

We have screened a large number of primarily single-site-specific base analogs of the 30-mer DNA aptamer for complex formation with L-argininamide. The results of these experiments are shown in Table 1. Specific guanines in the DNA aptamer were replaced by adenine (completely different Watson–Crick edge), inosine (replacement of amino group by a proton at position 2), O^6-methylguanine (replacement of carbonyl by O-methyl at position 6 and N-H by N at position 1), 8-bromoguanine (replacement of hydrogen by bulky bromine at position 8), deazaguanine (replacement of N by C-H at position 7) and 2-aminopurine (replacement of carbonyl by hydrogen at position 6 and N-H by N at position 1). Specific adenines in the DNA aptamer were replaced by 2′-O-methyl adenine (replacement of hydrogen by O-methyl at the sugar 2′ position), N^6-methyl adenine (replacement of one amino group hydrogen by methyl at position 6), deazaadenine (replacement of N by C-H at position 7), nebularine (replacement of amino group by hydrogen at position 6) and 2-amino adenine (replacement of hydrogen by amino group at position 2). In addition, specific thymines were replaced by uracils (replacement of methyl group by hydrogen at position 5) and cytosines by 5-methylcytosines (replacement of hydrogen by methyl group at position 5).

The imino proton spectra (8.5 to 14.5 ppm) of complexes between argininamide and two DNA aptamer analogs containing substitutions of G9 by 2-aminopurine (Figure 4b) and A13 by 2-amino adenine (Figure 4c) are compared with the corresponding complex involving the unmodified DNA aptamer (Figure 4a). The imino proton patterns establish that the complex formation occurs with both of these DNA aptamer analogs. Replacement of guanine by 2-aminopurine results in a loss of the imino proton and therefore it is not surprising that one imino proton in the control spectrum assigned to G9 at 9.99 ppm (Figure 4a) disappears when G9 is replaced by 2-aminopurine (Figure 4b). The imino proton of T21 shifts from 14.00 ppm in the control spectrum (Figure 4a) to 13.18 ppm when A13 is replaced by 2-amino adenine (Figure 4c), which is suggestive of a base-pairing alignment between A13 and T21 on opposite sides of the large hairpin loop. This conclusion is consistent with the NOE-based demonstration of Watson–Crick A13•T21 base-pair alignment on complex formation.

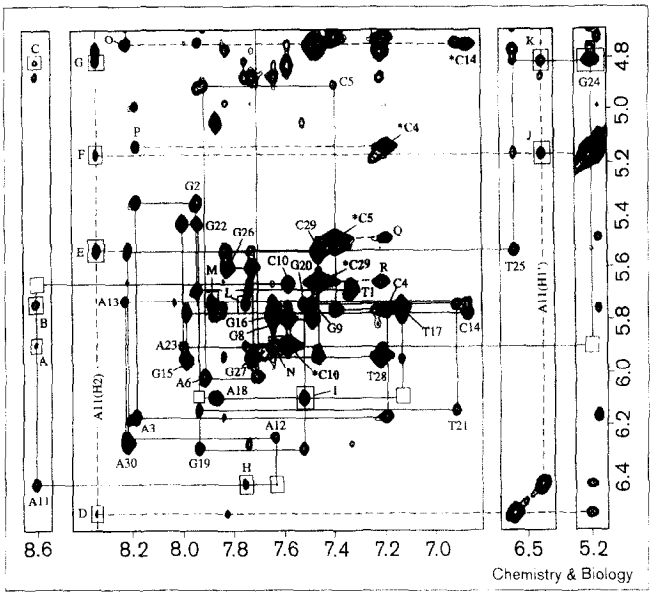
The above approaches have yielded definitive imino and amino proton assignments in the argininamide–DNA aptamer complex at 4°C and these chemical shifts are

Figure 4



An approach to exchangeable and nonexchangeable proton assignments in the argininamide-DNA aptamer complex based on site-specific base modification of DNA aptamer residues. **(a)** The exchangeable proton NMR spectrum (8.5 to 14.5 ppm) of the argininamide–DNA aptamer complex (approximately four equivalents of L-argininamide per DNA aptamer) in H_2O buffer, pH 6.4 at 4°C. Corresponding spectra of the complex containing site-specific substitutions of **(b)** 2-aminopurine at position G9 and **(c)** 2-aminoadenine at position A13.

Figure 5



Expanded NOESY (200 ms mixing time) contour plots of the argininamide–DNA aptamer complex (approximately four equivalents of L-argininamide per DNA aptamer) in D_2O buffer, pH 6.4 at 4°C. The NOE cross peaks between the base and their own sugar H1′ protons are labeled by residue position. The cytosine H5–H6 NOE cross peaks are labeled by asterisks. The NOEs are traced between the base protons and their own and 5′ flanking sugar H1′ protons with breaks (or very weak NOEs) observed in the connectivities at G9-C10-A11-A12, G16-T17-A18-G19 and A23-G24 steps (boxed regions). The NOE cross peaks A to R are assigned as follows: A, A11(H8)-A23(H1′); B, A11(H8)-G9(H1′); C, A11(H8)-G24(H1′); D, A11(H2)-T25(H6); E, A11(H2)-T25(H1′); F, A11(H2)-G24(H8); G, A11(H2)-G24(H1′); H, A23(H2)-A11(H1′); I, G20(H8)-A18(H1′); J, A11(H1′)-G24(H8); K, A11(H1′)-G24(H1′); L, A23(H2)-G9(H1′); M, A18(H2)-G20(H1′); N, A12(H8)-A23(H1′); O, A13(H8)-C14(H5); P, A3(H8)-C4(H5); Q, C4(H6)-C5(H5); R, T28(H6)-C29(H5). The NOE cross peaks involving the H8 (8.62 ppm), H2 (8.34 ppm) and H1′ (6.44 ppm) protons of A11 are boxed and labeled A to C, D to G and J to K, respectively. The H1′ proton of G8 (4.81 ppm), H8 proton of G24 (5.20 ppm) and the H6 proton of T25 (6.54 ppm) are shifted upfield of their normal range while the H8 proton of A11 (8.62 ppm) is shifted downfield of its normal range.

shown in the Supplementary material (Table S1) available through the online version of this paper.

Nonexchangeable proton spectra and restraints

The nonexchangeable protons in the argininamide–DNA aptamer complex have been assigned by several independent approaches. The primary approach has been to analyze the NOESY spectrum of the complex in D_2O buffer and use the site-specifically labeled analogs to resolve ambiguities in crowded regions of the NOESY contour plot. Thus, replacement of adenine by either nebularine or 2-aminopurine replaces the amino group by a hydrogen at position 6 and provides an extremely valuable marker for those DNA aptamer analogs that still bind L-argininamide. Similarly, replacement of adenine and guanine by their deazapurine analogs introduces a hydrogen at position 7, which can also be an extremely valuable marker for those deazapurine DNA aptamer analogs that still bind L-argininamide. The ability to ^{15}N-label the N^7 (and N^1) positions of purines results in a measurable two bond ^{15}N-^1H coupling to the H8 (and adenine H2) protons of those purines and has helped to verify the H8 (and adenine H2) proton assignments of loop residues G7, G8, G9, A11, A13, A18, G19, G20, G22, A23 and G24 in the complex for samples labeled with ^{15}N at the N^7 and N^1 positions.

Table 2

Intermolecular NOEs between the L-argininamide and the nucleic acid protons in the argininamide–DNA aptamer complex.

L-Argininamide*	Intermolecular NOEs
Sidechain protons	
Hα (3.99 ppm)	C10(H6)(w), C10(H1′)(m), C10(H2″)(w), C10(H4′)(m), C10(H5′,H5″)(m)
Hβ (1.86 ppm)	C10(H6)(vw), C10(H1′)(m), C10(H2′,H2″)(w), C10(H4′)(m), C10(H5′,H5″)(w), I16(H2)(w)†, A18(H8)(w), G19(H8)(m)
Hγ (1.61 ppm)	C10(H1′)(m), C10(H2′,H2″)(w), C10(H4′)(w), C10(H5′,H5″)(w), I16(H2)(w)*, G19(H8)(m)
Hδ (3.09, 3.16 ppm)	C10(H1′)(m), C10(H4′)(m), I16(H2)(w)†, T17(H1′)(w), T17(H4′)(w), A18(H8)(w), G19(H8)(m)
NHε (7.34 ppm)	C10(H1′)(s), C10(H2′, H2″)(w), C10(H3′)(w), C10(H4′)(w), A18(H8)(m), G19(H1′)(vw)
$NH_2η$ (6.62, 6.99 ppm)	C10(H1′)(m), A18(H1′)(m), G19(H8)(m), G19(H1′)(m), G20(H8)(m)
Backbone protons	
NH_2 (7.48 ppm)	C10(H6)(vw), C10(H1′)(w), C10(H4′)(s), C10(H5′,H5″)(m)

*$CONH_2$ (8.07 ppm). †I16 substituted Arg–DNA aptamer complex.

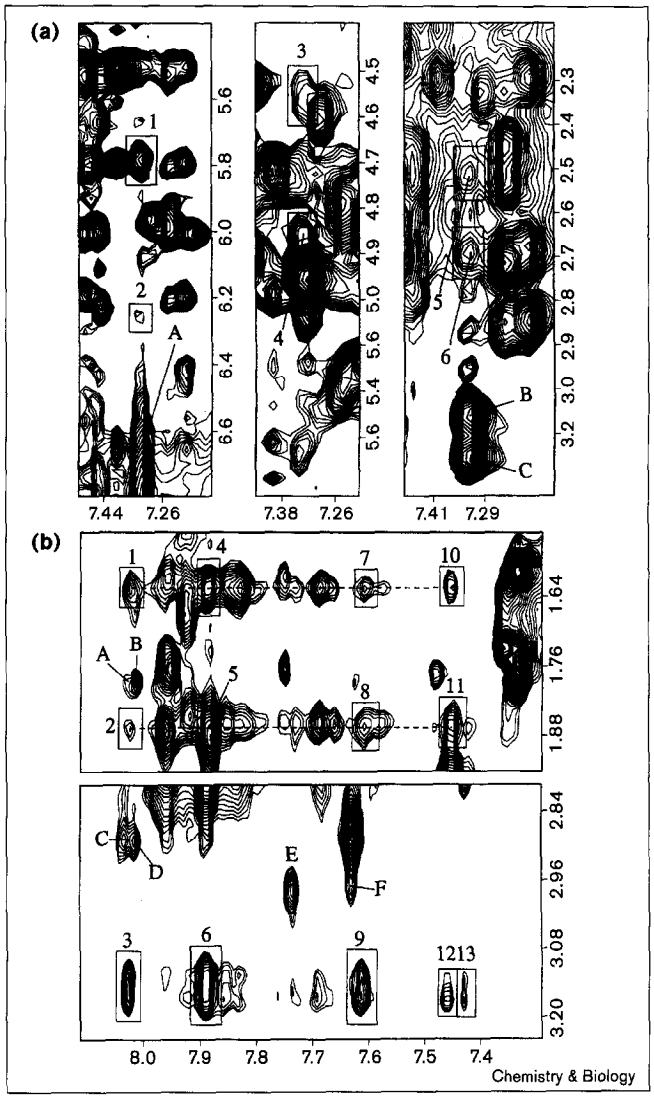
An expanded NOESY contour plot correlating NOEs between base protons (6.8 to 8.5 ppm and outlying regions) and sugar H1′ and cytosine H5 protons (4.7 to 6.6 ppm) of the argininamide–DNA aptamer complex in D_2O buffer at 4°C is plotted in Figure 5. We can trace sequential NOE connectivities between the base and its own and 5′-flanking sugar H1′ protons with breaks observed in the sequential connectivities (unlabeled boxed regions in Figure 5) at the G9-C10-A11-A12, G16-T17-A18-G19 and A23-G24 steps in the loop segment of the complex. Several base and sugar H1′ protons are upfield and downfield shifted on complex formation (listed in the legend to Figure 5) reflecting ring current contributions in the folded structure of the DNA aptamer in the complex. A thorough analysis of the entire NOESY spectrum has yielded the nonexchangeable proton assignments in the argininamide–DNA aptamer complex at 4°C and these chemical shifts are listed in Table S1 of the Supplementary material.

The A11 residue exhibits downfield shifted H8 (8.62 ppm) and H2 (8.34 ppm) proton chemical shifts in the complex. These A11 base protons together with the sugar H1′ proton exhibit NOEs to the sugar protons of G9 and the A23-G24-T25 segment (peaks A to C, D to G and J to K, Figure 5), suggesting an unusual alignment for this residue in the complex.

Mismatch-pair alignments between loop residues on complex formation

Only G24 amongst the 30-residues in the DNA aptamer adopts a *syn* alignment on complex formation with L-argininamide. This conclusion is based on the very strong NOE

Figure 6



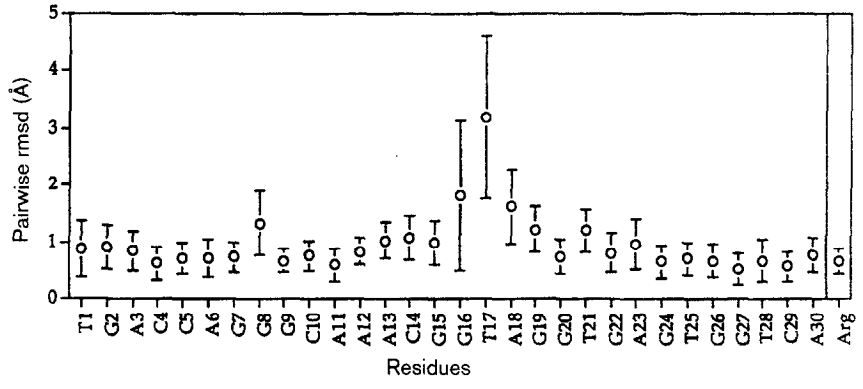
(a) Expanded NOESY (135 ms mixing time) contour plots identifying intermolecular NOEs in the argininamide–DNA aptamer complex (approximately four equivalents of L-argininamide per DNA aptamer) in H_2O buffer, pH 6.4 at 4°C. Intermolecular NOEs involving the L-argininamide NHε proton are labeled 1 to 6 and are assigned as follows: 1, Arg(NHε)-C10(H1′); 2, Arg(NHε)-G19(H1′); 3, Arg(NHε)-C10(H4′); 4, Arg(NHε)-C10(H3′); 5, Arg(NHε)-A18(H2′, H2″); 6, Arg(NHε)-C10(H2′). Intramolecular NOEs involving the L-argininamide NHε proton are labeled A to C and are assigned as follows: A, Arg(NHε)-Arg($NH_2η$); B,C, Arg(NHε)-Arg(Hδ). **(b)** Expanded NOESY (200 ms mixing time) contour plots identifying intermolecular NOEs in the argininamide-DNA aptamer complex (approximately four equivalents of L-argininamide per DNA aptamer) in D_2O buffer, pH 6.4 at 4°C. Intermolecular NOEs involving the L-argininamide nonexhangeable protons are labeled 1 to 13 and are assigned as follows: 1, I16(H2)-Arg(Hγ); 2, I16(H2)-Arg(Hβ); 3, I16(H2)-Arg(Hδ); 4, G19(H8)-Arg(Hγ); 5, G19(H8)-Arg(Hβ); 6, G19(H8)-Arg(Hδ); 7, A18(H8)-Arg(Hγ); 8, A18(H8)-Arg(Hβ); 9, A18(H8)-Arg(Hδ); 10, C10(H6)-Arg(Hγ); 11, C10(H6)-Arg(Hβ); 12, C10(H6)-Arg(Hδ); 13, G20(H8)-Arg(Hδ). Intramolecular NOEs involving the DNA aptamer nonexchangeable protons are labeled A to F and are assigned as follows: A, A12(H2)-A23(H4′); B, A23(H8)-A23(H4′); C, A12(H2)-A23(H5′, H5″); D, A23(H8)-A23(H5′, H5″); E, A23(H2)-A11(H2′); F, A12(H8)-A11(H2′).

Table 3

(a) NMR refinement statistics for the argininamide–DNA aptamer complex.

NMR distance restraints in the complex*	
Total number of DNA aptamer distance restraints	890
Exchangeable distance restraints	134
Nonexchangeable distance restraints	756
Hydrogen bond restraints†	27
Total number of intermolecular distance restraints	44
Exchangeable distance restraints	18
Nonexchangeable distance restraints	26
Total number of L-argininamide distance restraints	0
Structural statistics of the complex	
NOE violations	
Number > 0.2 Å	14.3 ± 4.5
Maximum violations (Å)	0.39 ± 0.22
rmsd of violations	0.076 ± 0.005
Deviations from the ideal covalent geometry	
Bond length (Å)	0.012 ± 0.001
Bond angle (°)	3.73 ± 0.06
Impropers (°)	0.36 ± 0.01
Pairwise rmsd (Å) among 9 distance-refined structures of complex	
L-argininamide plus DNA aptamer residues G1 to A30	1.18 ± 0.26
L-argininamide plus DNA aptamer residues G1 to G15 and A18 to G30	0.98 ± 0.22

(b) Pairwise individual rmsd values in the argininamide–DNA aptamer complex.



*NMR distance restraints are for the entire complex. †Three G•C and three A•T Watson–Crick base pairs from the stem region together with two Watson–Crick G•C (G15•C10 and G20•C14), one Watson–Crick A•T (A13•T21), one Hoogsteen G•G mismatch (G7•G24) and one sheared G•A mismatch (G22•A12) pairs from the zippered-up loop region.

between the H8 proton of G24 and its own H1′ proton [20] in short mixing time NOESY spectra of the complex. It is also supported by the expected break in the connectivity between the H8 proton of G24(*syn*) and the H1′ proton of A23 at the A23-G24 step in the complex. Further, the observation of NOEs between the imino and amino protons of G7 and the H8 proton (and H1′ proton by spin diffusion) of G24 is strongly indicative of a G7(*anti*)•G24(*syn*) base-mismatch alignment on complex formation.

A set of NOEs are observed between the base and sugar protons of A12 and G22 that are indicative of their relative alignment in the complex. An NOE is observed between the amino proton of G22 and the H8 proton of A12 which is diagnostic of a sheared G•A mismatch pair involving hydrogen-bonding alignment between the minor-groove edge of the guanine and the major-groove edge of the adenine [21,22]. The amino proton of A12 could not be identified and hence we could not monitor the potential NOE between the amino protons of A12 and the sugar protons of G22 which is also diagnostic of a sheared G•A pair. Replacement of A12 by nebularine, however, resulted in formation of a weak complex in which the hydrogen that replaced the amino group now showed strong NOEs to the sugar protons of G22 in the complex. The weak binding on replacement of A12 by nebularine

Figure 7

Superposed stereo view of stick representations of nine distance-refined solution structures of the argininamide–DNA aptamer complex. The Watson–Crick stem base pairs are colored white as are the poorly defined loop tip G16 and T17 residues. Loop residues involved in adaptive Watson–Crick base-pair formation are cyan whereas those involved in adaptive base mismatch and triple formation are magenta. Residue A11, which is positioned deep within the minor groove, is green. The DNA aptamer backbone is orange with the phosphate oxygens deleted for clarity in this and subsequent figures. The bound L-argininamide is yellow.

(disruption of the sheared G•A mismatch hydrogen bond involving the amino group of A12) and the lack of binding on replacement of A12 by deazaadenine (disruption of the sheared G•A mismatch hydrogen bond involving the N^7 of A12) also support formation of a sheared A12•G22 mismatch pair on complex formation.

Intermolecular restraints in the complex

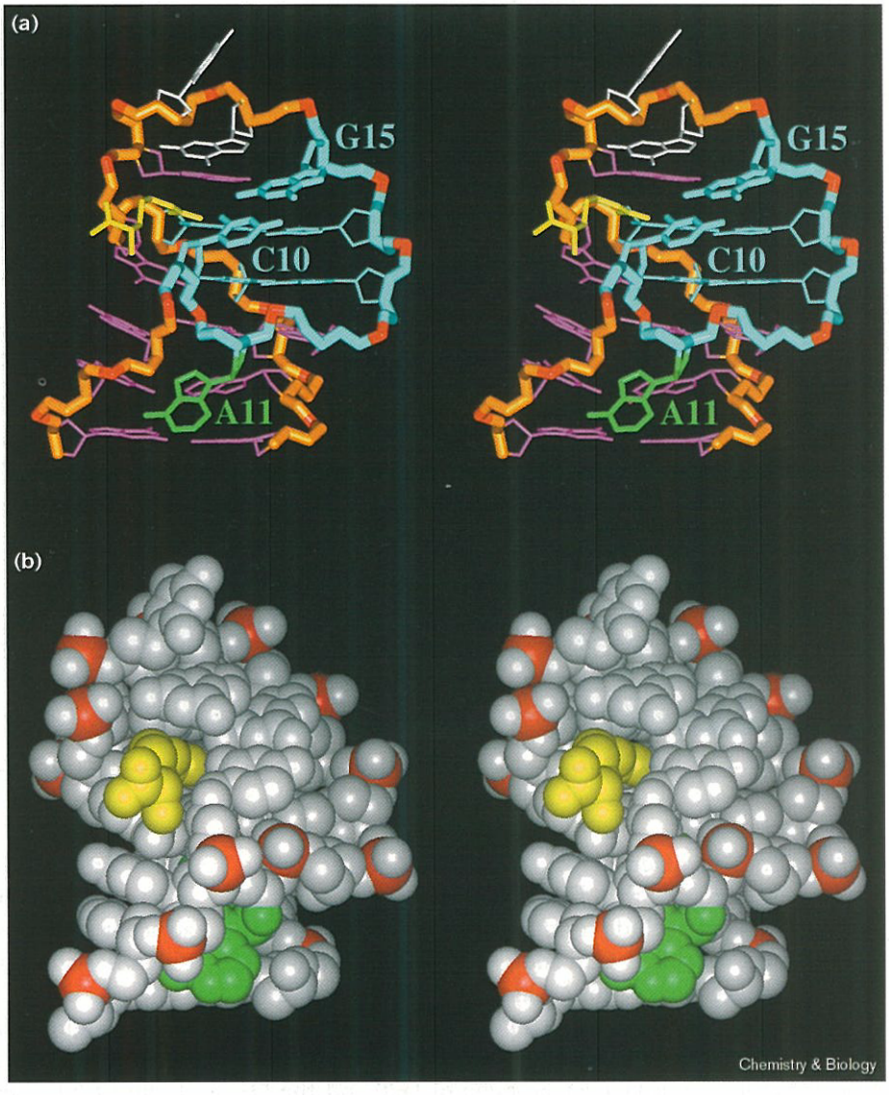
We have identified 44 intermolecular restraints between L-argininamide and the DNA aptamer in the complex. These intermolecular NOEs are listed in Table 2 with their intensities categorized as strong (s), medium (m), weak (w) and very weak (vw). We observe NOEs involving both the exchangeable (sidechain NHε and NH_2-η and backbone NH_2 protons) and nonexchangeable (Hα, Hβ, Hγ and Hδ) protons of L-argininamide to the base and sugar protons of C10, G16 (I16), T17, A18, G19 and G20 residues of the DNA aptamer in the complex. Examples of intermolecular NOEs involving the argininamide NHε protons are shown as boxed cross peaks (labeled 1 to 6) in Figure 6a and involving the argininamide Hβ, Hγ and Hδ protons are shown as boxed cross peaks (labeled 1 to 13) in Figure 6b. Additional intermolecular NOEs in the complex are shown in the Supplementary material (Figure S1). These intermolecular NOEs provide critical restraints for defining recognition events within the amino-acid-binding pocket in the argininamide–DNA aptamer complex.

Structure calculations

The input restraints statistics for the structure calculations on the argininamide–DNA aptamer complex are listed in Table 3a. The solution structure of the complex was solved starting from distance geometry followed by distance-restrained molecular-dynamics calculations guided by a total of 934 distance restraints which include 18 exchangeable proton and 26 nonexchangeable proton intermolecular restraints (Table 2) using the computational protocols outlined in the Materials and methods section. We incorporated hydrogen-bond restraints for the experimentally identified Watson–Crick pairs within the loop (G15•C10, G20•C14 and A13•T21) and stem segments of the complex throughout the computations. In addition, the low energy structures following initial distance-geometry runs consistently contained G7(*anti*)•G24(*syn*) and sheared A12•G22 mismatch alignments. Hydrogen-bond restraints defining these experimentally defined mismatch-pairing alignments were therefore also incorporated throughout the computations. The quality statistics of the nine distance refined structures of the complex are summarized in Table 3a and exhibit pairwise root mean square deviation (rmsd) values of 1.18 ± 0.26 for the entire complex. Residues G16 and T17 at the tip of the hairpin loop are least well defined and their deletion reduces the pairwise rmsd values to 0.98 ± 0.22. There were 14.3 ± 4.5 violations of magnitude > 0.2 Å out of a total of 934 restraints (1.5%).

Figure 8

(a) A stereo stick view of one representative refined solution structure of the G7 to G24 segment of the argininamide–DNA aptamer complex. The coloring code is the same as in Figure 7 except for the backbone from C10 to G15, which is cyan, and the phosphorus atoms, which are red. The L-argininamide, the A11 residue, the G15•C10 base pair and the backbone are highlighted with thicker bonds. **(b)** A stereo space-filling view of one representative refined solution structure of the G7 to G24 segment of the argininamide–DNA aptamer complex. The DNA is white except for phosphorus atoms (red) and A11 (green). The bound L-argininamide is yellow.

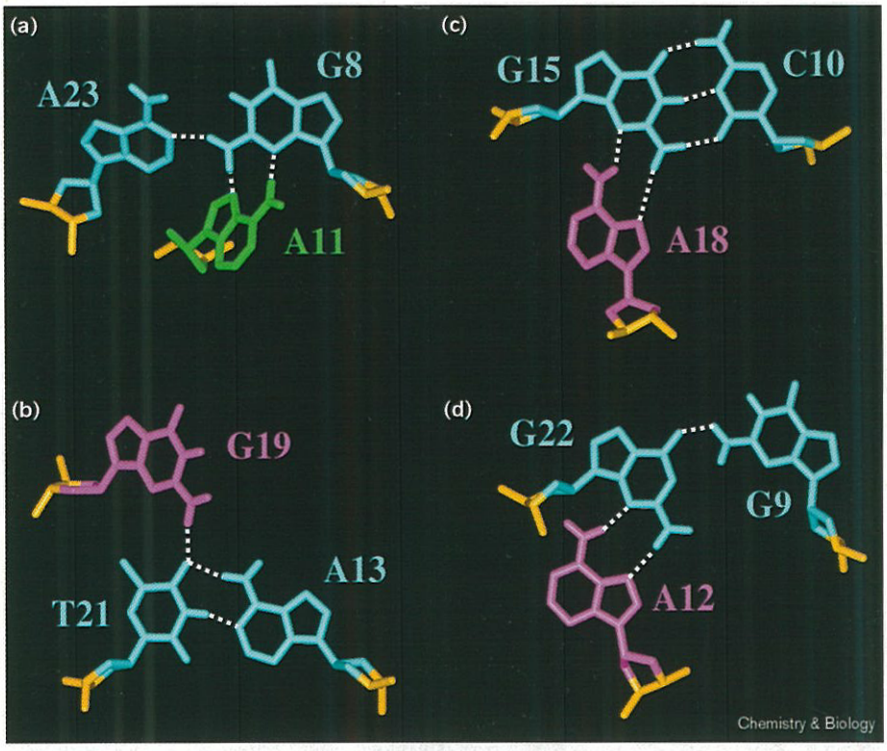


Structure analysis

The rmsd values at individual residue positions of the DNA aptamer between the distance-refined structures of the complex are plotted in Table 3b and establish G16, T17 and A18 as the least defined residues in the complex. A stereo view of the nine distance-refined structures of the quadruplex looking normal to the helix axis is shown in Figure 7. The bound argininamide is shown in yellow with the large hairpin loop zippering up through Watson–Crick (cyan) and mismatch and base triple (magenta) formation. Stereo views in stick and space-filling representations of the zippered-up G7 to G24 loop segment of a representative distance-refined solution structure of the DNA aptamer with bound L-argininamide are shown in Figure 8a and 8b, respectively. The C10 to G15 segment forms a nested loop closed by a Watson–Crick G15•C10 base pair (backbone and closing base pair of nested loop in cyan, Figure 8a) so that there are loops nested within loops in the solution structure of the complex. Strikingly, A11 (green) is positioned within the walls of the minor groove of the zippered up hairpin loop on complex formation (Figure 8a).

The structure of the complex contains a wealth of unusual alignments that are summarized below. We observe four base triples in the solution structure of the complex. Three of these triples ([A23•G8]•A11, Figure 9a; [C10•G15]•A18, Figure 9c; [G9•G22]•A12, Figure 9d) involve formation of sheared G•A pairs through alignment of the major groove of an adenine with the minor groove of a guanine involved in either Watson–Crick (Figure 9c) or mismatch (Figure 9a,d) pairs. The last triple ([A13•T21]•G19, Figure 9b) involves recognition by the minor groove of guanine of the major groove edge of a thymine of a A•T pair.

Figure 9



Pairing alignments of the four base triples in one representative refined solution structure of the argininamide–DNA aptamer complex. **(a)** Alignment of the A11 residue (green) within the minor groove of the propeller-twisted G8•A23 mismatch pair (cyan). G8 and A11 form a twisted sheared G•A pair aligned through two potential hydrogen bonds. **(b)** Alignment of the G19 residue (magenta) within the major groove of the Watson–Crick A13•T21 base pair (cyan). G19 and T21 align through one potential hydrogen bond. **(c)** Alignment of the A18 residue (magenta) within the minor groove of the Watson–Crick G15•C10 base pair (cyan). G15 and A18 form a slightly twisted sheared G•A pair aligned through two potential hydrogen bonds. **(d)** Alignment of the A12 residue (magenta) within the minor groove of the Watson–Crick G9•G22 mismatch pair (cyan). G22 and A12 form a sheared G•A pair aligned through two potential hydrogen bonds.

The residues involved in generating the L-argininamide binding pocket on the DNA aptamer are shown in Figure 10a. They include residues C10, A18, G20 and T21 (Figure 10a), as well as G16 and G19 (Figure 10c), which form a binding pocket into which fits the guanidinium group of the bound L-argininamide. The amino acid is anchored in position by hydrogen bonds from the guanidinium group to the Hoogsteen edge of G20 and the O^2 oxygen of C10 (Figure 10b), as well as interdigitative stacking of the guanidinium group between the purine rings of G16 and G19 (Figure 10d) in the complex.

The DNA aptamer adapts an unusual backbone conformation for the G9-C10-A11-A12 segment on complex formation, primarily due to Watson–Crick pairing of C10 with G15 and positioning of the A11 residue deep within the minor groove of the zippered-up helix (Figure 11a). There is no stacking between adjacent residues within the G9-C10-A11-A12 segment of the complex (Figure 11b). The unanticipated deep burial of the A11 residue (green) within the walls of the minor groove can be visualized in either space-filling (Figure 11c) or GRASP surface (Figure 11d) representations of the complex.

The solution structure of the complex contains two adjacent stacked triples [G9•G22]•A12 and [A13•T21]•G19 that form platforms from which emanate flanking stem segments (magenta and cyan) whose helical axis are displaced relative to each other (GRASP slab view in Figure 12a). There is extensive stacking between adjacent and non-adjacent residues within this domain of the complex (GRASP slab views in Figure 12a,b).

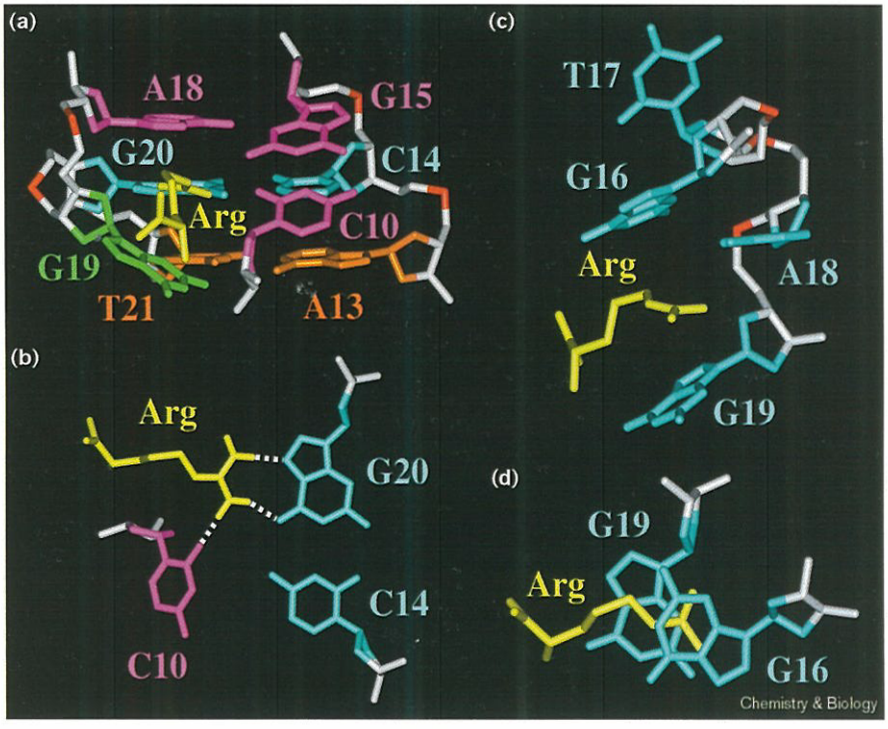
The L-argininamide-binding pocket within the DNA aptamer is highlighted in a GRASP view of the complex (Figure 12c) where the amino acid (yellow) is shown in a stick representation and the adapted DNA aptamer is shown in a surface representation.

Discussion

The Harada–Frankel model of the complex

The solution structure of the 18-residue hairpin loop of the 30-mer DNA aptamer formed through mismatch and triple alignment on L-argininamide complex formation is shown in Figure 8a. This structure of the complex contrasts with the model of the complex (Supplementary material Figure S2) proposed by Harada and Frankel [14] on the basis of an analysis of chemical interference and functional group deletion data. They proposed that the hairpin loop aligns through G•G, A•A and G•A mismatch formation, with one half of the mismatch-containing stem pivoting back on the other half to form a quadruplex stabilized by three mixed purine (G and A) tetrads. Our NMR–molecular dynamics structural studies reported in this paper do not support the Harada–Frankel model of the argininamide–DNA aptamer complex [14] and recommend

Figure 10



An expanded view of the ligand-binding site in one representative refined solution structure of the argininamide–DNA aptamer complex. **(a)** A stick view of the binding pocket composed of C10, A13-C14-G15 and A18-G19-G20-T21 segments and the bound L-argininamide. The A13•T21 Watson–Crick pair is orange, the G20•C14 Watson–Crick pair is cyan and the A18•(G15•C10) triple is magenta. The bound L-argininamide is yellow. **(b)** Intermolecular hydrogen-bonding alignments of the guanidium group of L-argininamide (yellow) and the O^2 of C10 (magenta) and the Hoogsteen edge of G20 of the G20•C14 Watson–Crick pair (cyan). **(c)** Interdigitation of the L-argininamide sidechain (yellow) between the purine rings of G16 and G19. The bases are cyan, whereas the backbone is white with phosphorus atoms in red. **(d)** Stacking of the L-argininamide between the purine rings of G16 and G19.

against proposing models of nucleic-acid-aptamer complexes based solely on footprinting and mutagenesis data.

Zippering-up of the 18-residue hairpin loop on complex formation

The generation of the L-argininamide-binding pocket within the 30-mer DNA aptamer is achieved through a series of novel pairing alignments that involve residues adjacent to the stem segment and those located further out into the hairpin loop (Figure 7). Three mismatches, G7•G24, G8•A23 and G9•G22, form a continuous stacked helix with the adjacent Watson–Crick stem segment on complex formation (mismatches designated by dashed lines in the schematic in Figure 2b and bases shown in magenta in GRASP slab view in Figure 12a,b). The A11 residue is positioned within the minor groove of this stacked mismatched segment and anchors the mismatched residues in place within the complex (Figure 8a).

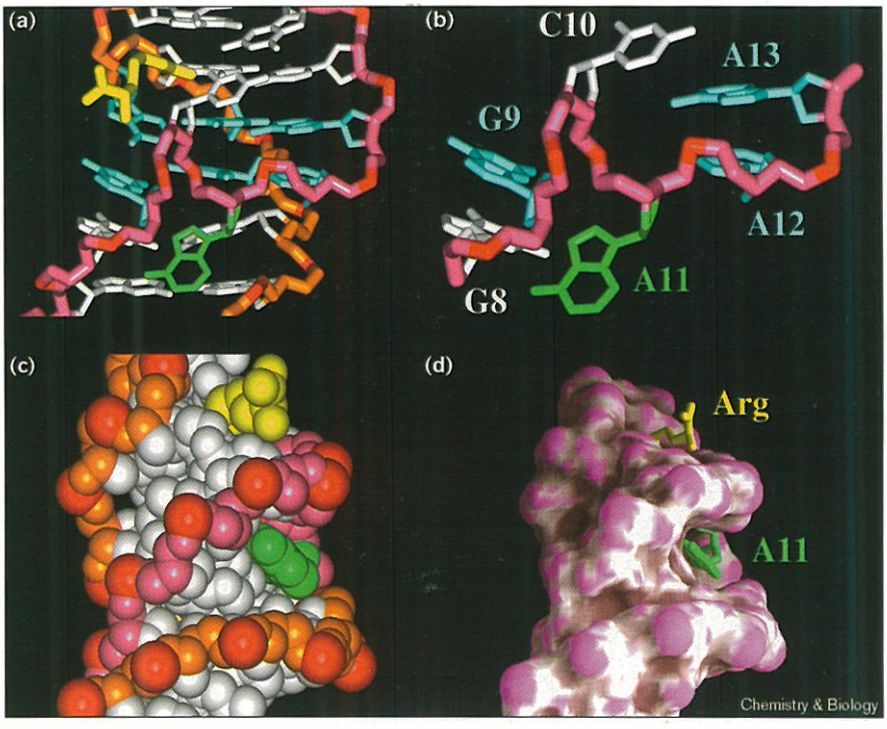
The sheared G22•A12 mismatch pair together with Watson–Crick A13•T21 and G20•C14 base pairs form a second continuous stacked helix on complex formation (schematic in Figure 2b and bases in cyan in the GRASP slab view in Figure 12a,b). The two helical segments are not collinear but have their helix axis displaced by ~5.4 Å (Figure 12a). Two adjacent base triples A12•(G22•G9) and (A13•T21)•G19 form stacked platforms centered about the junctional site between these two adjacent zippered-up helical segments (Figure 12a) and in the process extend the stacking alignments of the helical segments into the base triple platforms. Such a displacement of the axis of helical segments mediated by base triples and/or triads has also been reported previously following structure determination of the AMP–DNA aptamer complex [13]. The importance of stacking patterns within the zippered-up hairpin loop is reinforced by the alignment of the purine rings of G7, G8, G9 and G19 that are essentially stacked directly over each other (GRASP slab view in Figure 12a), as are the base rings of A18, G20, T21 and G22 (GRASP slab view in Figure 12b) in the solution structure of the complex.

The L-argininamide-binding site is formed by the A18-G19-G20-T21 and A13-C14-G15 segments anchored in place through formation of a critical Watson–Crick G15•C10 base pair (Figure 10a). These residues are located furthest (other than the poorly defined G16 and T17 residues) from the Watson–Crick stem segment and thus the DNA aptamer has to undergo a coordinated adaptive transition to generate its L-argininamide-binding pocket.

Watson–Crick pair alignments on complex formation

The adaptive structural transition of the DNA aptamer hairpin loop is achieved, in part, through formation of Watson–Crick G15•C10, A13•T21 and G20•C14 base pairs on complex formation. The Watson–Crick G15•C10 base pair is involved in the recognition of the bound amino acid because of the intermolecular hydrogen bond

Figure 11



The alignment of the A11 residue within the minor groove of the zippered-up helix in one representative refined solution structure of the argininamide–DNA aptamer complex. **(a)** A stick view emphasizing the sugar-phosphate backbone geometry in the vicinity of the A11 residue. The bases are in white except for A11 (green) and the (G9•G22)•A12 and (C10•G15)•A18 triples (cyan). The backbone of one half (from G7 to G15) of the DNA aptamer is pink whereas that of the other half (from G16 to G24) is orange with phosphorus atoms in red. The bound L-argininamide is yellow. The helix axis is displaced by ~5.4 Å on either side of the stacked base triples. **(b)** An expanded view of (a) focused on the G8-G9-C10-A11-A12-A13 segment of the complex. **(c)** A space filling view emphasizing the burial of the A11 residue in green within the walls of the minor groove. The color code is the same as in (a) except that the triples are also white. **(d)** A GRASP [37] surface representation of the complex emphasizing the binding cavity with the bound L-argininamide (in stick representation) in yellow and A11 residue (in stick representation) in green buried within the minor groove. The convex and concave surfaces are shown in pale purple and gray respectively.

between the O^2 carbonyl group of C10 and the guanidinium $NH_2\eta$ protons of the bound L-argininamide in the complex (Figure 10b). We were unable to form the complex following replacement of G15•C10 pair with the isoG15•isoC10 pair because replacement of the carbonyl group of C10 by the amino group of isoC10 would impact on intermolecular hydrogen-bond formation. Furthermore, all three hydrogen bonds across the G15•C10 pair appear to be important because replacement of G15 by inosine (one less hydrogen bond) prevented complex formation.

The formation of the Watson–Crick A13•T21 pair on complex formation is supported by base-analog substitution experiments. The importance of the adenine NH_2 group at position 6 for Watson–Crick A13•T21 pair formation was reflected in complex formation being disrupted following replacement by a hydrogen (nebularine analog). By contrast, both substitution by 2-aminoadenine (results in the formation of one additional hydrogen bond) or substitution by deazaadenine (modification along the non-hydrogen-binding N^7 position) had no effect on complex formation.

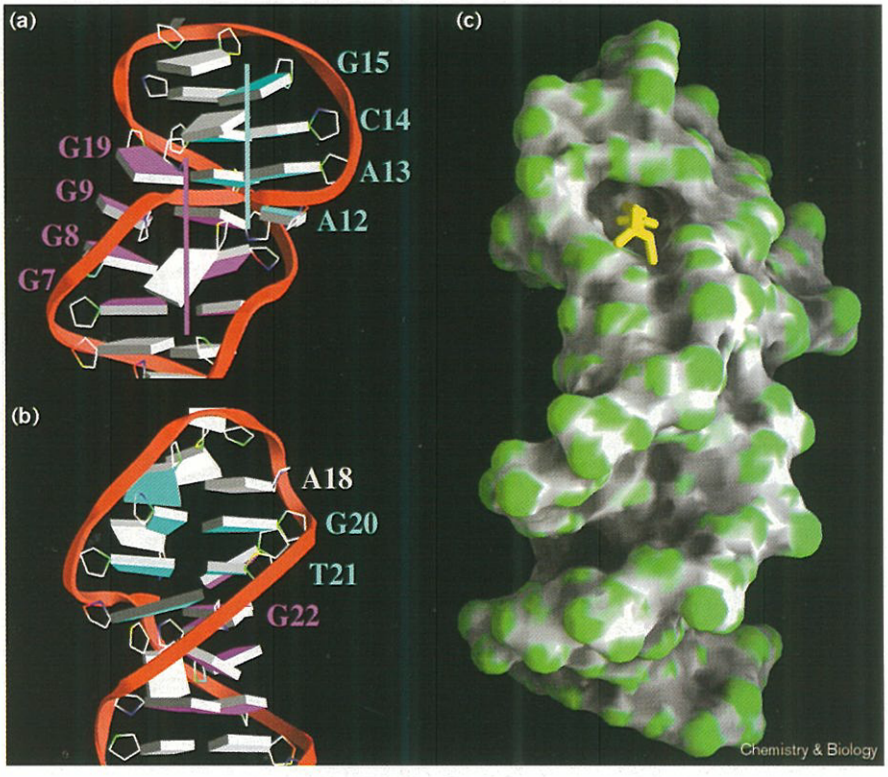
The requirement for Watson–Crick G20•C14 pair alignment on complex formation was reflected in complex formation being disrupted following substitution of G20 by inosine (one less hydrogen bond) or adenine (less stable A•C mismatch formation). Furthermore, G20 cannot be replaced by adenine because its Hoogsteen edge is the prime determinant in the intermolecular hydrogen-bond recognition of the bound L-argininamide on complex formation (Figure 10b).

Base-mismatch alignments on complex formation

The stem of the DNA aptamer helix is extended through formation of adjacent stacked G7•G24, G8•A23 and G9•G22 mismatches on complex formation. The NMR evidence supporting a G7(*anti*)•G24(*syn*) mismatch alignment was outlined in the Results section to justify the use of hydrogen-bonding restraints involving this alignment during the computations. The analog substitution experiments further support this G•G mismatch alignment, which involves hydrogen bonding between the Watson–Crick edge of G7(*anti*) and the Hoogsteen edge of G24(*syn*) in the complex. Thus, analogs (inosine and 2-aminopurine) that disrupt the N^1 and NH_2-2 positions along the Watson–Crick edge of G7 and analogs (O^6-methylguanine and deazaguanine) that disrupt the O^6 and N^7 positions along the Hoogsteen edge of G24 prevent complex formation.

The G9•G22 mismatch is of the sheared type with G22 moving towards the minor groove and G9 moving towards the major groove (Figure 9d). The guanine bases in the G•G mismatch are aligned through a single hydrogen bond between the O^6 carbonyl of G22 and the NH_2-2 proton of G9 (Figure 9d). The imino and amino protons of

Figure 12



Two views **(a,b)** of GRASP [37] slabs and ribbon representations of the argininamide–DNA aptamer complex with the L-argininamide removed in the interest of clarity. The helical segment encompassing the G7•G24, G8•A23 and G9•G22 mismatches are magenta with a magenta-colored vertical line defining the helix axis. The helical segment encompassing the A13•T21 and G20•C14 Watson–Crick pairs are cyan with a cyan-colored vertical line defining the helix axis. Both stems extend into adjacent stacked A12•(G22•G9) and (A13•T21)•G19 base triples with a pronounced displacement of their helix axis (~5.4 Å). A12 is cyan whereas G19 is magenta. G15 is also cyan. Note the stacking of G7, G8, G9 and G19 bases in (a) and the stacking of A18, G20, T21 and G22 bases in (b). **(c)** A GRASP [37] surface representation of the entire complex looking down into the binding cavity with the bound L-argininamide (in stick representation) in yellow. The convex and concave surfaces are shown in green and gray respectively.

G22 resonate at 11.54 ppm and 7.93, 8.01 ppm, respectively. The importance of the O^6 atom of G22 in this G•G mismatch alignment is supported by the disruption of complex formation following replacement of the carbonyl group by a hydrogen (2-aminopurine and adenine analogs). The role of the NH_2 proton of G9 in this G•G mismatch alignment is supported by the disruption of complex formation following replacement of the amino group by hydrogen (inosine analog). By contrast, replacement of G9 by 2-aminopurine has no effect on complex formation because this modification retains the 2-amino group but disrupts the rest of the Watson–Crick edge, which is not involved in mismatch formation (Figure 9d).

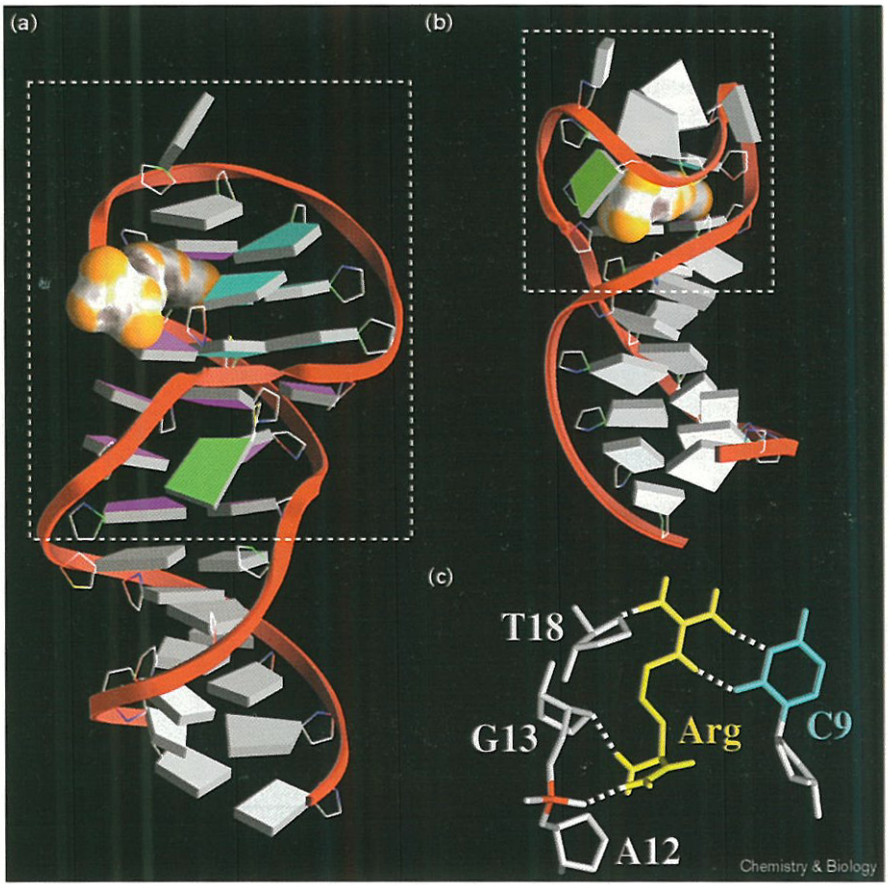
The G8•A23 mismatch involves the Watson–Crick edges of both bases and is stabilized by one hydrogen bond involving the NH_2-2 group of G8 and the N^1 of A23 on complex formation (Figure 9a). This propeller-twisted G(*anti*)•A(*anti*) alignment is supported by studies of base analogs at these residues in the complex. Replacement of G8 by inosine prevents complex formation because the NH_2-2 group cannot be replaced by hydrogen without disruption of the single hydrogen bond across the G•A mismatch pair in the complex (Figure 9a). By contrast, replacement of G8 by 2-aminopurine has no effect on complex formation because this modification retains the 2-amino group but disrupts the rest of the Watson–Crick edge, which is not involved in mismatch formation (Figure 9a). There is no effect on the binding following replacement of the NH_2-6 group of A23 by H (nebularine analog) because this amino group is not involved in G8•A23 mismatch alignment on complex formation (Figure 9a).

Base triple alignments on complex formation

We observe four base triples in the argininamide–DNA aptamer complex, three of which involve recognition of mismatches and Watson–Crick pairs through the minor groove and one of which involves recognition through the major groove. The G19 base is positioned in the major groove and forms one potential hydrogen bond through its NH_2-2 group with the O^4 carbonyl group of the A13•T21 Watson–Crick pair in the complex (Figure 9b). We have been unable to monitor the amino protons of G19 but their importance is confirmed by the loss of binding when replaced by hydrogen in the inosine analog. The G19•T21 alignment is propeller-twisted within the (A13•T21)•G19 triple (Figure 9b) in the complex.

The remaining three triples involve recognition by the major-groove edge of an adenine of the minor-groove edge of a guanine involved in either mismatch or base-pair alignment resulting in sheared G•A mismatch formation. This novel observation implies that base mismatches and base pairs can be targeted through the minor groove through sheared G•A mismatch formation. These results considerably expand our current knowledge of triple helix

Figure 13



GRASP [37] slabs and ribbons representations of **(a)** the argininamide–30-mer DNA aptamer complex (this study) and **(b)** the argininamide–24-mer DNA aptamer complex [15]. The adaptive structures adopted by the loop segments on complex formation are encompassed by the dashed boxes. The bound argininamide molecules are shown in surface representations with convex and concave surfaces shown in yellow and white, respectively. **(c)** Intermolecular hydrogen-bonding alignments in the binding pocket of the argininamide–24-mer DNA aptamer complex [15]. The protons of the guanidinium group of L-argininamide (yellow) are hydrogen bonded to the N^3 and O^2 atoms of C9 (blue) and the sugar ring oxygen of T18 (white). Additional hydrogen bonds can potentially anchor the peptide and amide functionalities of the bound argininamide in the complex.

architectures where all previous efforts had focused on targeting the DNA duplex through the major groove [23,24]. All three bases are propeller-twisted in the A11•(G8•A23) base triple (Figure 9a) with the A11 residue positioned deep within the walls of the minor groove. The imino proton of G8 (10.94 ppm) exhibits NOEs to the H8 proton of A11 and H2 proton of A23, defining the relative positions of the three bases in the triple (Figure 9a). Both G8 amino protons (7.47, 7.73 ppm) form potential hydrogen bonds (Figure 9a). The H2 proton of A23 shows an unusually strong NOE to the H8 proton of A11, as well as to the sugar (H1′, H2′,2″ and H4′) protons of the same residue. Both the NH_2-6 proton and N^7 atom of A11 are involved in sheared G•A pair formation because replacement of A11 by deazaadenine and nebularine analogs resulted in disruption of complex formation.

The A18•(G15•C10) base triple involves recognition through sheared G15•A18 mismatch formation of the Watson–Crick G15•C10 pair (Figure 9c). Both amino protons of G15 (8.08, 8.69 ppm) are downfield shifted and exhibit NOEs to the H8 proton of A18 characteristic of sheared G•A mismatch formation. The involvement of the NH_2-6 proton of A18 through hydrogen-bond formation in the sheared G•A mismatch alignment (Figure 9c) is supported by the disruption in binding when this adenine NH_2 group is replaced by its $NHCH_3$ analog and by H (nebularine).

The A12•(G22•G9) base triple involves recognition through sheared G22•A12 mismatch formation of a G22•G9 mismatch pair (Figure 9d). The NMR evidence supporting a sheared G22•A12 mismatch alignment was outlined in the Results section to justify use of hydrogen-bonding restraints involving this alignment during the computations. The G22 residue bridges the G9 and A12 residues within the A12•(G22•G9) base triple that forms on complex formation. The importance of the NH_2-2 protons of G22 in base triple formation are supported by the disruption of binding on replacement by hydrogen (inosine and adenine analogs).

Recognition of L-argininamide within its binding pocket

The major-groove edges of the zippered-up Watson–Crick A13•T21 and G20•C14 base pairs and sheared G15•A18 mismatch pair, which are stacked on each other, constitute one segment of the binding pocket targeted by the guanidium group of the bound L-argininamide (Figure 10a). The other segment is made up by the Watson–Crick G15•C10 pair which is positioned across

one side of this major groove and limits the size of the binding pocket (Figure 10a).

The guanidinium group of L-argininamide forms a pair of intermolecular hydrogen bonds to the Hoogsteen edge O^6 and N^7 atoms of G20 and is further buttressed in place by a intermolecular hydrogen bond to the O^2 of C10 in the complex (Figure 10b). The guanidinium group is anchored in its binding pocket by being sandwiched between the purine planes of G16 and G19 (Figure 10c,d) resulting in the interdigitation of the amino-acid sidechain with nucleotide purine bases in the complex. The backbone NH_2 protons of L-argininamide are within hydrogen-bonding distance of the O^6 atom of G19 in the refined structures of the complex.

There are both similarities and differences in the recognition of L-argininamide within the binding pocket (Figure 13b) of the 24-mer DNA aptamer (Figure 1a) reported previously [15] and the binding pocket (Figure 13a) of the 30-mer DNA aptamer (Figure 1b) reported in this study. The similarities include an extended L-argininamide sidechain inserted between base planes of the binding pocket with recognition associated with intermolecular hydrogen bond formation involving its guanidinium group and acceptor atoms along a base edge. The differences reflect the participants in the intermolecular recognition in the two complexes. The NHε and $NH_2\eta$ protons of the guanidinium group of the amino acid formed a pair of hydrogen bonds with N^3 and O^2 acceptor atoms along the Watson–Crick edge of a stacked cytosine (Figure 13c) in the argininamide-DNA 24-mer aptamer complex [15]. By contrast, the $NH_2\eta$ protons of the guanidinium group of the amino acid formed a pair of hydrogen bonds with the Hoogsteen edge of a guanine and one hydrogen bond with the O^2 of cytosine (Figure 10b) in the argininamide–DNA 30-mer aptamer complex reported in this study.

Comparison of DNA- and RNA-binding pockets for L-argininamide/L-arginine

The reported solution structure of L-arginine bound to a RNA aptamer showed the guanidinium group to be inserted into a purine-rich binding pocket with its NHε and $NH_2\eta$ protons intermolecularly hydrogen bonded to acceptor atoms along the Watson–Crick edge of a cytosine residue [17]. In addition, bent hydrogen bonds could also form between the guanidinium $NH_2\eta$ protons and the Hoogsteen edges of two guanine residues in the complex. There were no hydrogen bonds from the charged guanidinium group to the backbone phosphate groups of the RNA aptamer in the complex. Thus, there are similarities in the intermolecular hydrogen-bonding alignments in the arginine–RNA aptamer complex [17] and those that are observed in the two argininamide–DNA aptamer complexes reported from our laboratory ([15] and this study).

The L-arginine fork alignment for L-argininamide bound to the HIV-1 TAR RNA proposed anchoring the $NH_2\eta$ and NHε functionalities of the guanidinium group of the bound amino acid through a pair of intermolecular hydrogen bonds to the major-groove edge of a guanine residue and, in addition, through contacts between the charged guanidinium group and a pair of backbone phosphate groups [18,19]. These charged guanidinium-group–backbone-phosphate interactions proposed for the argininamide–TAR RNA complex [18,19] are not observed in the L-argininamide–DNA aptamer complex reported in the present study.

Adaptive alignment of the G9-C10-A11-A12 segment on complex formation

The G9-C10-A11-A12 segment of the complex plays a key role in the adaptive structural transition of the DNA aptamer on complex formation (Figure 11a). None of the adjacent bases in this tetramer segment stack on each other, primarily due to C10 and A11 being oriented in opposite directions (Figure 11b). The C10 residue plays a pivotal role through formation of a Watson–Crick pair with the G15 residue resulting in the C10-A11-A12-A13-C14-G15 segment forming a loop closed by a G15•C10 base pair in the complex (Figure 11a). Thus, G15 anchors, through hydrogen-bond formation, the upper half of the zippered-up 18-residue DNA aptamer loop which includes the amino-acid-binding site on complex formation.

Burial of the A11 residue within the minor groove

The A11 residue, which is oriented in the opposite direction from the C10 residue (Figure 11b), is buried within the walls of the minor groove of the zippered-up DNA helix (Figure 11c,d). To the best of our knowledge this is the first time a deoxy base has been shown to be sandwiched within the walls of the minor groove of DNA where it spans both strands of the helix. Such an envelopment of A11 within the walls of the minor groove has similarities with the binding of noncovalently bound antitumor drugs such as netropsin [25,26] and distamycin [27,28] to the minor groove of A•T rich segments and the Mg^{2+}-coordinated chromomycin dimer [29,30] to the minor groove of G•C rich segments of the DNA helix. The alignment of an adenine residue within the minor groove, however, has precedent in RNA recognition elements ranging from nucleoside triples in group I introns [31], GAAA tetraloops interacting with an RNA helix [32], GNRA tetraloops interacting with their internal loop receptors [33] and in an RNA pseudoknot [34].

The A11 residue fits snugly within the walls of the minor groove because the addition of functional groups to either the base (2-aminoadenine and 8-bromoadenine analogs) or the sugar ring (2′-O-methyladenine analogs) disrupts complex formation. The absence of stacking for the A11 residue together with its edge-on position within the

minor groove results in downfield-shifted positions for the base protons (H8, 8.62 ppm and H2, 8.34 ppm) of this residue in the complex. We also observe upfield-shifted positions for the H4′ proton of G9 (1.76 ppm) and the H8 proton of G24 (5.20 ppm), which are positioned above and below the purine ring of A11 in the complex.

Proton complexation shifts

Several additional base and sugar protons are shifted to high field in the argininamide–DNA aptamer complex because they are positioned over purine rings and experience upfield ring current contributions. These protons include the H6 (6.54 ppm) and CH_3 (0.93 ppm) protons of T25, which are positioned over the purine ring of adjacent G24 (*syn*), the H4′ proton of A23 (1.75 ppm), which is positioned over the purine ring of A12, the CH_3 protons of T21 (0.97 ppm), which are positioned over the purine ring of adjacent G22, and the imino proton of G19 (8.79 ppm), which is positioned over the purine ring of G9 in the complex.

Solution structure of complex explains footprinting and mutagenesis data

Harada and Frankel [14] have undertaken a set of chemical interference and mutagenesis experiments to identify those residues and their functionalities that are important for argininamide–DNA aptamer complex formation. The majority of these footprinting and mutagenesis data can be explained by our solution structure of the complex. Thus, Harada and Frankel [14] demonstrated that both the NH_2-2 and N^7 positions of G20 are critical for complex formation. This can be explained by our solution structure because the guanine NH_2-2 proton is involved in hydrogen bonding across the G20•C14 base pair, whereas the guanine N^7 atom forms an intermolecular hydrogen bond with the guanidinium group of the bound argininamide in the complex (Figure 10b). Harada and Frankel [14] similarly demonstrated that the NH_2-6 and N^7 positions of A11 are important for complex formation. This can be explained in our solution structure in which A11 is buried deep within the walls of the minor groove of the zippered-up mismatch helical segment and hence is inaccessible to chemical modification in the complex (Figure 11c,d). In addition, both NH_2-6 and N^7 positions of A11 participate in hydrogen-bond formation with the minor-groove edge of G8 through propeller-twisted sheared G8•A11 mismatch formation in the complex (Figure 9a).

New folds and alignments

The present solution structure of the argininamide–DNA aptamer complex makes unique and important contributions to our existing understanding of ligand–nucleic-acid structure and recognition. The structure of the complex contains several unique features that are summarized below. Previous studies of DNA triplex formation have invariably focused on recognition of base pairs or mismatches through the major groove [23,24]. Only one of four base triples in the complex involves recognition through the major groove, however. The other three involve recognition through the minor groove where an adenine uses its major-groove edge to target the minor groove edge of a guanine through a sheared G•A mismatch alignment. This new alignment code opens the possibility for the design and generation of a new family of triplexes in which the third strand is positioned in the minor groove. Second, our observation of an adenine residue positioned deep within the minor groove was most unexpected and adds a new dimension to what can be accommodated within the walls of this groove in duplex DNA. Past research has always focused on minor-groove recognition by antibiotics and dyes bound either covalently or noncovalently and accommodated, on the basis of their complementary shape, within the walls of this groove. Intramolecular base encapsulation within the minor groove blocks entry into this groove, whose width could be potentially modulated by the size and shape of the buried base. Third, the dimensions of the binding pocket in the argininamide–DNA aptamer complex are defined through formation of a Watson–Crick base pair that bridges across one side of the major-groove edge of a zippered-up helical segment targeted by the bound L-argininamide. This base pair is also involved in loop closure, highlighting the concept of loops nested within larger loops as a means for generating higher order nucleic acid folds. Fourth, pairs of zippered-up helical segments are found to align end to end but with a relative displacement of their helix axis in the complex. The present study on the argininamide–DNA aptamer complex and an earlier study on the AMP–DNA aptamer complex [13] establish that a pair of adjacent stacked triples and/or triads can mediate helix–helix junctional sites and, in so doing, extend stacking patterns despite the non-colinearity of end-to-end aligned helical segments.

Significance

Our laboratory has recently focused on the structural analysis of nucleic-acid–aptamer complexes for several reasons. First, nucleic-acid aptamers adopt novel folds on complex formation and add immeasurably to our current knowledge base of folded states of DNA and RNA and the motifs and scaffolds that stitch the higher order architectures together. Second, complex formation invariably involves adaptive structural transitions in which unstructured regions of the nucleic-acid aptamers zipper up to generate unique ligand-binding pockets. Third, the nucleic-acid aptamers target their ligands with high affinity and specificity and the structural analysis of ligand–nucleic-acid recognition can add significantly to our current knowledge of the relative contributions of hydrophobic, hydrogen bonding and electrostatic interactions to molecular recognition.

The structural transition by which the 18-residue hairpin loop of the DNA aptamer generates its L-argininamide-binding pocket presents a striking example of a propagated adaptive conformational change. A series of motifs that extend stacking beyond the stem segment are propagated towards the tip of the loop to generate the amino-acid-binding pocket. The integrity of this binding pocket must depend on the folding architecture of segments that bridge the binding site and the stem segment. Hence, different interconnected structural motifs can contribute either directly or indirectly to ligand binding.

It is striking that the primary determinant in the recognition of the charged guanidinium group of L-argininamide by both DNA aptamers involves intermolecular hydrogen bonding with acceptor heteroatoms along the base edge rather than contacts with the backbone phosphate groups. Furthermore, the sandwiching of the guanidinium group and a part of the attached methylene sidechain of L-argininamide between purine bases in the complex reinforces the importance of buttressing hydrophobic interactions. Such interdigitation of amino-acid sidechains and nucleic-acid bases might turn out to be a recurring theme in nucleic-acid recognition.

Materials and methods

Sample preparation

L-argininamide was purchased from the Sigma Chemical Co. and used without further purification. The 30-mer DNA aptamer (residues T1 to A30, Figure 1b), its 15-mer halves (residues T1 to G15 and G16 to A30) and base analogs were synthesized on a 10 μM scale on an Applied Biosystems 392 DNA synthesizer using solid phase β-cyanophosphoamidite chemistry and purified by reversed-phase HPLC. The L-argininamide was added gradually to the DNA aptamer and complex formation monitored by recording imino proton spectra at 4°C.

Site-specific $^{13}C,^{15}N$-labeled 15-mer DNAs

We have site-specifically incorporated 1,2,7-^{15}N-labeled guanine (designated *G), 2-^{13}C, 1,2,7-^{15}N-labeled guanine (designated **G) and 1,2,6-^{15}N labeled adenine (designated *A) into the two 15-mer halves (residues T1 to G15 and G16 to A30) of the 30-mer DNA aptamer (residues T1 to A30, Figure 1b). Four samples were prepared with site-specifically incorporated labels as follows to provide unambiguous exchangeable imino and amino and nonexchangeable H8 proton assignments: sample 1 contained **G7, *G9 and *A11 labels, whereas sample 2 contained *G8 and *A13 within the T1 to G15 fragment. Sample 3 contained *G19, **G22 and *A23, whereas sample 4 contained *G20, *G24 and *A18 within the G16 to A30 fragment. The details of the synthesis and purification of these labeled samples using previously published procedures [35] will be reported elsewhere.

NMR data collection and processing

NMR spectra of the exchangeable and nonexchangeable protons of the argininamide–DNA aptamer complex were collected in aqueous buffer (50 mM KCl, 10 mM phosphate, pH 6.4) at 4°C on Varian 600 MHz Unity *INOVA* NMR spectrometers. Two-dimensional data sets included NOESY, COSY, TOCSY and ROESY homonuclear experiments on unlabeled complexes and $^{1}H,^{15}N$-HSQC (with sensitivity enhancement) heteronuclear experiments on site-specifically labeled $^{13}C,^{15}N$-labeled argininamide–DNA aptamer (T1 to G15 plus G16 to A30) complexes. Data sets were processed using Varian VNMR software and analyzed using the FELIX program (Molecular Simulations Inc.).

Nonexchangeable interproton distance restraints were obtained from the buildup of NOE cross-peak volumes in NOESY data sets (50, 90, 150 and 200 ms mixing times) on the argininamide–DNA aptamer complex in D_2O buffer at 4°C and bounds were set between ± 10% and ± 20% of the calculated distances using the fixed cytidine H5–H6 reference distance of 2.45 Å. Nonstereospecific assignments were treated with r^{-6} averaging. Interproton distance restraints involving exchangeable protons of the same complex were obtained from NOESY spectra (90 and 135 ms mixing times) in H_2O buffer at 4°C with bounds set between ± 15% and ± 20% of the calculated distance using the thymine imino to adenine H2 reference distance of 2.91 Å across an A•T base pair.

Distance geometry and molecular dynamics calculation

A set of 300 initial structures of the argininamide–DNA aptamer complex were generated by the X-PLOR [36] based metric matrix distance geometry (DG) protocol guided by the available distance restraints. Hydrogen-bonding distance restraints were imposed to align the experimentally identified stem Watson–Crick pairs (± 0.10 Å), as well as the experimentally identified loop Watson–Crick G15•C10, G20•C14 and A13•T21 base pairs, Hoogsteen G7•G24 and sheared A12•G22 mismatch pairs (± 0.15 Å) during the DG (and subsequent molecular dynamics) calculations. We did not use planarity or angular restraints for the hydrogen bonds during the computations. These structures were quantitatively scored to select 20 structures with the least NOE violations, acceptable covalent geometry, and favorable van der Waals energy.

The X-PLOR [36] based restrained molecular dynamics (MD) calculations were carried out in two cycles using the simulated annealing protocol and the CHARMM force field with reduced phosphate charges. Each cycle of restrained MD simulations were initially carried out at 300°K with a force constant of 1 kcal $mol^{-1}Å^{-2}$ on all experimentally obtained distance restraints. The structure was subjected to 500 cycles of energy minimization and was slowly heated to 1000°K in 5 ps (0.5 ps per 50K increase). The force constants on the experimentally obtained distance restraints were slowly scaled up to 32 (nonexchangeable protons), 16 (exchangeable protons) and 64 (hydrogen bonds) kcal $mol^{-1}Å^{-2}$, over a period of 12 ps. The system was allowed to evolve for another 8 ps at 1000°K and next cooled gradually to 300°K over 5 ps (0.5 fs time step) with retention of the full scale of distance restraints and subsequently equilibrated for 12 ps at 300°K. The coordinates were averaged over the last 5 ps and the resulting coordinates subjected to 20 cycles (100 steps each cycle) of conjugate gradient energy minimization. Nine final distance refined structures were selected, based on the criteria of low restraints violation and low total energy.

Graphic programs

INSIGHT II (Molecular Simulations, Inc) and GRASP programs [37] were used to display structures.

Coordinates deposition

The coordinates of the argininamide–DNA aptamer complex have been checked for correct chirality and deposited (accession number: 2arg) in the Protein Data Bank, Brookhaven National Laboratory, New York.

Supplementary material

Supplementary material available with the online version of this paper includes: the proton chemical shifts of the argininamide–DNA aptamer complex and two figures outlining additional intermolecular NOEs in the argininamide–DNA aptamer complex in D_2O solution and the proposed Harada–Frankel model [14] of the argininamide–DNA aptamer complex.

Acknowledgements

This research was supported by NIH CA-46778 to D.J.P.

References

1. Joyce, G.F. (1994). *In vitro* evolution of nucleic acids. *Curr. Opin. Struct. Biol.* **4**, 331-336.
2. Gold, L., Polisky, B., Uhlenbeck, O.C. & Yarus, M. (1995). Diversity of oligonucleotide functions. *Ann. Rev. Biochem.* **64**, 763-797.

3. Lorsch, J.R. & Szostak, J.W. (1996). Chance and necessity in the selection of nucleic acid catalysts. *Accounts Chem. Res.* **29**, 103-110.
4. Osborne, S.E., Matsumura, I. & Ellington, A.D. (1997). Aptamers as therapeutic and diagnostic reagents: problems and prospects. *Curr. Opin. Chem. Biol.* **1**, 5-9.
5. Pan, T. (1997). Novel and variant ribozymes obtained through *in vitro* selection. *Curr. Opin. Chem. Biol.* **1**, 17-25.
6. Breaker, R.R. (1997). DNA aptamers and DNA enzymes. *Curr. Opin. Chem. Biol.* **1**, 26-31.
7. Patel, D.J. (1997). Structural analysis of nucleic acid aptamers. *Curr. Opin. Chem. Biol.* **1**, 32-46.
8. Santoro, S.W. & Joyce, G.F. (1997). A general purpose RNA-cleaving DNA enzyme. *Proc. Natl Acad. Sci. USA* **94**, 4262-4266.
9. Carmi, N., Shultz, L.A. & Breaker, R.R. (1996). *In vitro* selection of self-cleaving DNAs. *Chem. Biol.* **3**, 1039-1046.
10. Bock, L.C., Griffin, L.C., Latham, J.A., Vermaas, E.H. & Toole, J.J. (1992). Selection of single-stranded DNA molecules that bind and inhibit human thrombin. *Nature* **355**, 564-566.
11. Padmanabhan, K., Padmanabhan, K.P., Ferrara, J.D., Sadler, J.E. & Tulinsky, A. (1993). The structure of α-thrombin inhibited by a 15-mer single-stranded DNA aptamer. *J. Biol. Chem.* **268**, 17651-17654.
12. Huizenga, D.E. & Szostak, J.W. (1995). A DNA aptamer that binds adenosine and ATP. *Biochemistry* **34**, 656-665.
13. Lin, C.H. & Patel, D.J. (1997). Structural basis of DNA folding and recognition in an AMP-DNA aptamer complex: distinct architectures but common recognition motifs for DNA and RNA aptamers complexed to AMP. *Chem. Biol.* **4**, 817-832.
14. Harada, K. & Frankel, A.D. (1995). Identification of two novel arginine binding DNAs. *EMBO J.* **14**, 5798-5811.
15. Lin, C.H. & Patel, D.J. (1996). Encapsulating an amino acid in a DNA fold. *Nat. Struct. Biol.* **3**, 1046-1050.
16. Geiger, A., Burgstaller, P., von der Eltz, H., Roeder, A & Famulok, M. (1996). RNA aptamers that bind L-arginine with sub-micromolar dissociation constants and high enantioselectivity. *Nucleic Acids Res.* **24**, 1029-1036.
17. Yang, Y., Kochoyan, M., Burgstaller, P., Westhof, E. & Famulok, M. (1996). Structural basis of ligand discrimination by two related RNA aptamers resolved by NMR spectroscopy. *Science* **272**, 1343-1347.
18. Puglisi, J.D., Chen, L., Frankel, A.D. & Williamson, J.R. (1993). Role of RNA structure in arginine recognition of TAR RNA. *Proc. Natl Acad. Sci. USA* **90**, 3680-3684.
19. Puglisi, J.D., Tan, R., Calnan, B.J., Frankel, A.D. & Williamson, J.R. (1992). Conformation of the TAR RNA-arginine complex by NMR spectroscopy. *Science* **257**, 76-80.
20. Patel, D.J., Kozlowski, S., Nordheim, A. & Rich, A. (1982). Right-handed and left-handed DNA: studies of B-DNA and Z-DNA by using proton nuclear Overhauser effect and phosphorus NMR. *Proc. Natl Acad. Sci. USA* **79**, 1413-1417.
21. Li, Y., Zon, G. & Wilson, W.D. (1991). NMR and molecular evidence for a G•A mismatch base pair in a purine-rich DNA duplex. *Proc. Natl Acad. Sci. USA* **88**, 26-30.
22. Heus, H.A. & Pardi, A. (1991). Structural features that give rise to unusual stability of RNA hairpins containing GNRA loops. *Science* **253**, 191-194.
23. Sun, J-S. & Helene, C. (1993). Oligonucleotide-directed triple-helix formation. *Curr. Opin. Struct. Biol.* **3**, 345-356.
24. Radhakrishnan, I. & Patel, D.J. (1994). DNA triplexes: solution structures, hydration sites, energetics, interactions and function. *Biochemistry* **33**, 11405-11416.
25. Patel, D.J. (1982). Antibiotic-DNA interactions: intermolecular nuclear Overhauser effects in the netropsin-d(CGCGAATTCGCG) complex in solution. *Proc. Natl Acad. Sci. USA* **79**, 6424-6428.
26. Kopka, M.L., Yoon, C., Goodsell, D., Pjura, P. & Dickerson, R.E. (1985). The molecular origin of DNA-drug specificity in netropsin and distamycin. *Proc. Natl Acad. Sci. USA* **82**, 1376-1380.
27. Pelton, J.G. & Wemmer, D.E. (1990). Binding modes of distamycin A with d(CGCAAATTTGCG) determined by two dimensional NMR. *J. Am. Chem. Soc.* **112**, 1393-1399.
28. Coll, M., Frederick, C.A., Wang, A.H.J. & Rich, A. (1987). A bifurcated hydrogen-bonded conformation in the d(A•T) base pairs of the DNA dodecamer d(CGCAAATTTGCG) and its complex with distamycin. *Proc. Natl Acad. Sci. USA* **84**, 8385-8389.
29. Gao, X. & Patel, D.J. (1989). Solution structure of the chromomycin-DNA complex. *Biochemistry* **28**, 751-762.
30. Leroy, J.L., Gao, X., Gueron, M. & Patel, D.J. (1991). Proton exchange and internal motion of chromomycin dimer-DNA oligomer complex. *Biochemistry* **30**, 8009-8017.
31. Chastain, M. & Tinoco Jr., I. (1993). Nucleoside triples from the group I intron. *Biochemistry* **32**, 14220-14228.
32. Pley, H.W., Flaherty, K.M. & McKay, D.B. (1994). Model for an RNA tertiary interaction from the structure of an intermolecular complex between a GAAA tetraloop and an RNA helix. *Nature* **372**, 111-113.
33. 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.
34. Kolk, M.H., van der Graff, M., Wijmenga, S.S., Pleij, C.W., Heus, H.A. & Hilbers, C.W. (1998). NMR structure of a classical pseudknot: interplay of single- and double-stranded RNA. *Science* **280**, 434-438.
35. Zhao, H., Pagano, A.R., Wang, W., Shallop, A., Gaffney, B.L. & Jones, R.A. (1998). Use of a ^{13}C atom to differentiate two ^{15}N labeled nucleosides. Synthesis of [$^{15}NH_2$]-adenosine, [1, NH_2-^{15}N2]- and [2-^{13}C-1, NH_2-^{15}N2-]-guanosine, [1, 7, NH_2-^{15}N3]- and [2-^{13}C-1, 7, NH_2-^{15}N3]-2′-deoxyguanosine. *J. Org. Chem.* **22**, 7832-7835.
36. Brunger, A.T. (1992). *X-PLOR. A System for X-ray Crystallography and NMR.* Yale University Press, New Haven, CT.
37. Nicholls, A., Sharp, K.A. & Honig, B.H. (1991). Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins* **11**, 281-296.

Because *Chemistry & Biology* operates a 'Continuous Publication System' for Research Papers, this paper has been published via the internet before being printed. The paper can be accessed from http://biomednet.com/cbiology/cmb – for further information, see the explanation on the contents pages.