Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Crystal structures of r(GGUCACAGCCC)$_2$

Viktor Kacer,[a,b] Stephen A. Scaringe,[c] J. Neel Scarsdale[b,d] and Jason P. Rife[a,b]*

[a]Department of Medicinal Chemistry, Virginia Commonwealth University, Richmond, VA 23298-0133, USA, [b]Institute for Structural Biology, Virginia Commonwealth University, Richmond, VA 23298-0133, USA, [c]Dharmacon Research, Boulder, CO 80301, USA, and [d]Department of Biochemistry and Molecular Biology, Virginia Commonwealth University, Richmond, VA 2398-0133, USA

Correspondence e-mail: jason.rife@vcu.edu

Received 13 February 2002
Accepted 26 November 2002

PDB References: r(GGUCACAGCCC)$_2$, 1kd3, r1kd3sf; 1kd4, r1kd4sf; 1kd5, r1kd5sf.

Crystals of small RNAs, which regularly diffract to very high resolution, can often be readily obtained. Unfortunately, for some RNAs the conformations adopted in the crystalline form are different from those found in solution. For example, short RNAs that form hairpins in solution virtually never crystallize thus; rather, they form duplexes. Nevertheless, these unintended structures have contributed greatly to the understanding of RNA structure. In a similar occurrence, the homodimer r(GGUCACAGCCC)$_2$ has been crystallized from an 11-mer/12-mer heteroduplex, r(GGCUGAAGUCCG)/r(GGUCACAGCCC). This surprising phenomenon was observed under a variety of crystallization conditions. The structure of the homoduplex was determined from crystals that differed in the precipitant used and the type of metal present. In all cases, the resulting homoduplexes contain ten base pairings, of which the central six are non-canonical pairings. In two of the variants, ordered metal-binding sites were observed: two equivalent octacoordinate Tl^+ sites in one and two equivalent nanocoordinate Ba^{2+} sites in another.

1. Introduction

The ability to prepare large quantities of small RNA molecules (about 30 nucleotides or less) by solid-phase chemistry or by using RNA polymerases allows researchers to study in fine detail the structures of smaller motifs that are the building modules of larger naturally found RNA molecules that are more difficult to handle experimentally (Scaringe, 2001; Milligan *et al.*, 1987). Detailed structural work on these stable 'fragments' has traditionally been the domain of nuclear magnetic spectroscopists, but in recent years X-ray crystallographers have made numerous contributions (Holbrook & Kim, 1997; Correll *et al.*, 1997, 1999; Ippolito & Steitz, 1998; Moore, 1995).

Small RNAs (<30 nucleotides) are readily crystallized, but not always in the form expected. For example, essentially all attempts to crystallize hairpins of fewer than about 20 nucleotides in length have failed: they either fail to crystallize or they crystallize as homodimers (Holbrook *et al.*, 1991; Cruse *et al.*, 1994; Baeyens *et al.*, 1996; Anderson *et al.*, 1999; Tanaka *et al.*, 1999). In contrast, short duplexes containing mismatches, bulged nucleotides or internal loops have fared better (Correll *et al.*, 1997; Ippolito & Steitz, 1998, 2000; Deng & Sundaralingam, 2000; Hung *et al.*, 2000). To date, crystals obtained from solutions containing duplexes have yielded duplexes. For example, conversion of a duplex with an internal loop, made from self-complementary RNA, to a hairpin has not been observed. For reasons that are not entirely clear, duplex RNA crystallizes more readily than does hairpin RNA.

We attempted to crystallize RNA duplexes and hairpins that contained the aminoglycoside (*e.g.* gentamycin, paromomycin, tobramycin) binding site that is found in bacterial ribosomes. For one of our constructs (termed vkIX), a duplex composed of an 11-mer and a 12-mer, we obtained crystals that were grown in both the absence and presence of several aminoglycoside antibiotics. Depending on the antibiotics used, those crystals diffracted X-rays to resolution limits of between 2.7 and 1.6 Å.

Upon solving the structures of some of those crystals, we observed that instead of the expected heteroduplex, the 11-mer crystallized as a homodimer (Fig. 1). No drug was observed in any of the structures, which in itself is not surprising because the homodimer no longer contains the aminoglycoside-binding site found in the ribosome. Nevertheless, several interesting features are found in the structure of the resulting homodimer. Because the homodimer, which we termed vkIX-AA, as a crystal diffracted to 1.6 Å, is able to bind two different metal ions (Ba^{2+} and Tl^{+}), is composed mainly of mismatched base pairs and illustrates an unusual crystallization phenomenon, we feel this structure is of general importance to the RNA community.

2. Materials and methods

2.1. RNA preparation and crystallization

The RNA described in this report were synthesized at Dharmacon Research using 5′-silyl-2′-orthoester chemistry (Scaringe, 2001). Each RNA was deprotected by adding 0.4 ml of the manufacturer's deprotection buffer to 0.25 µmol of either vkIX-A or vkIX-B and incubating the mixture at 333 K for 30 min. From this point on, the oligomers were treated in one of two ways, yielding RNA of indistinguishable quality. Some of the time we simply lyophilized the RNA to dryness, removing water and the other volatile components. Oligomers were then twice redissolved in water and lyophilized to dryness. Alternatively, we passed the oligomers over desalting columns (Waters, Inc.); the RNA was eluted with an aqueous/organic buffer (35% acetonitrile, 35% methanol, 30% 10 m*M* triethylammonium bicarbonate pH 7–8). Repeated lyophilizations removed the components of the elution buffer. In all cases, the purity of the oligomers was judged to be greater than 95% by denaturing (8 *M* urea) polyacrylamide gel electrophoresis. Therefore, no additional purification was performed. The stocks of RNA were redissolved to 10 mg ml^{-1} into a working buffer (10 m*M* Tris–HCl pH 7.0 and 5 m*M* $MgCl_2$). Duplexes of vkIX were obtained by mixing the two strands, adjusting the volume with working buffer to give an RNA concentration of 4 mg ml^{-1} and finally annealing by incubating the mixture at 333 K for 10 min and then cooling it to room temperature.

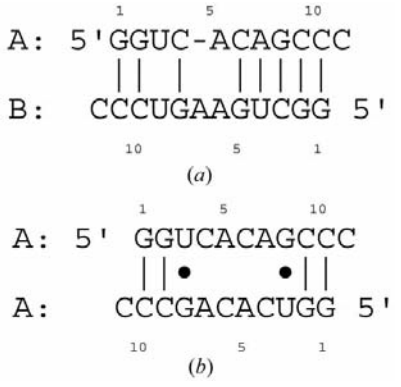

Figure 1
(*a*) Sequence of the heterodimer vkIX with strands A and B. (*b*) Sequence of the homodimer vkIX-AA composed of strand A only.

Although high-quality crystals were obtained under several crystallization conditions, only crystals from two of those conditions plus Tl^{+}-soaked crystals were used here to address the consequences of metal binding and precipitant type on RNA structure. In all cases, crystals were grown at room temperature by the hanging-drop vapor-diffusion method.

2.1.1. Crystal form 1 (MPD/Ba^{2+}). In growth condition 1, 2 µl of vkIX (4 mg ml^{-1}) was added to 2 µl of Miniscreen No. 12 [10%(*v*/*v*) MPD, 40 m*M* sodium cacodylate pH 6.0, 12 m*M* spermine.4HCl, 80 m*M* KCl and 20 m*M* $BaCl_2$; Hampton Research]. The resulting drop was stored over an aqueous solution containing 30% MPD. Crystals appeared within 1–3 d and belonged to space group $P4_3$, with unit-cell parameters $a = b = 39.99$, $c = 52.62$ Å.

2.1.2. Crystal form 2 (AS/no metal). In growth condition 2, 1 µl of vkIX was added to 2 µl of a crystallization solution containing 50 m*M* Tris–HCl pH 6.4, 1.4 *M* ammonium sulfate, 20 m*M* $MgCl_2$, 18 m*M* spermine and two molar equivalents of neomycin B (relative to vkIX-AB). The resulting solution was stored over the same crystallization solution. Crystals usually appeared after several weeks, but on occasion grew in a matter of days. They belonged to space group $P4_3$, with unit-cell parameters $a = b = 44.36$, $c = 49.74$ Å, and contained a single duplex per asymmetric unit.

2.1.3. Crystal form 3 (AS/Tl^{+}). Crystals grown in condition 2 were also used to generate a thallium heavy-atom derivative. A chloride-depleted version of the crystallization solution (40 m*M* cacodylate pH 6.5, 1.4 *M* ammonium sulfate, 12 m*M* spermine and 80 m*M* thallium acetate) was used for crystal soaking; this modified solution was required because TlCl readily precipitates out.

When ammonium sulfate was used as a precipitant, crystals were soaked in successive cryosolvents containing precipitant buffer and up to 20% xylitol before freezing in a cold N_2 stream (98 K). When crystals were grown in MPD, no additional cryosoaking step was performed prior to freezing.

2.2. Data collection and structure determination

Data were collected from single crystals at 98 K on an R-AXIS IIc imaging plate and a 50 kV/100 mA graphite-monochromated Cu *Kα* (1.5418 Å) X-ray beam focused with confocal optics (MSC, Inc.). A single data set was collected on a vkIX-AA crystal grown in ammonium sulfate (crystal form 2) at Brookhaven National Laboratory using the X8C synchrotron beamline tuned to a wavelength of 1.0 Å. All data

were integrated and reduced with the *HKL* suite of programs (Otwinowski & Minor, 1997).

The first structure of vkIX-AA solved used crystals grown in ammonium sulfate in the presence of neomycin-B (crystal form 2). Initial phasing information was obtained from a single isomorphous replacement with anomalous scattering (SIRAS) experiment using crystals soaked with Tl^+ (crystal form 3). Heavy-atom positions were refined and initial phases were determined using *SHARP* (de La Fortelle & Bricogne, 1997) followed by density modification using *SOLOMON* (Abrahams, 1997) as implemented in the *SHARP* suite. Initial maps proved sufficient to identify the position of 22 nucleotides. Once we realised that the crystals were of the homoduplex vkIX-AA and not of the heteroduplex vkIX-AB, model building proceeded smoothly. The initial model was refined using high-temperature torsion-angle molecular dynamics (TAMD) as implemented in *CNS* v. 1.0 (Brünger *et al.*, 1998) to R_{work} and R_{free} values of 29.8 and 31.5%, respectively. This model of vkIX-AA (from data collected at our home source) was used as a starting model to refine data collected at Brookhaven National Laboratory on similar crystals. After several rounds of automated and manual water fitting, followed by restrained isotropic *B*-factor refinement and the inclusion of waters, these values were reduced to 21.6 and 24.6%, respectively. Final refinement was carried out using *SHELX*97 (Sheldrick & Schneider, 1997) using least-squares refinement of atomic positions and anisotropic *B* factors, which led to terminal values of 0.171 and 0.228 for R_{work} and R_{free}, respectively.

The refined structure of vkIX-AA was used to obtain the structures of other vkIX-AA variants, including crystal form 1, *via* molecular replacement and refinement strategies as above, except that anisotropic *B*-factor refinements were not performed. For all of the variants studied, the map of electron density was poorest within the backbone of A7 and suggested that this region could best be modelled by two conformers. However, models refined this way always resulted in higher R_{work} and R_{free} values, which prompted us to only refine single conformations. At no point during our refinements were restraints included to force the two strands of vkIX-AA to be alike. Despite the freedom to refine independently, the two strands that make up a given homoduplex remained nearly identical. The heavy-atom root-mean-square difference (ha-RMSD) between the two strands from the three variants ranged from between 0.03 and 0.1 Å. Data-collection and refinement statistics for crystal forms 1, 2 and 3 are presented in Table 1.

Table 1
Data collection and refinement statistics for RNA oligomers.

Values in parentheses are for the highest resolution shell.

	Free	Ba^{2+}	Tl^+
Data collection			
Resolution (Å)	1.58	1.85	1.80
R_{merge}† (%)	5.8	9.2	6.7
$\langle I\rangle/\langle\sigma(I)\rangle$	22.4 (1.7)	19.9 (1.8)	28.7 (3.5)
Completeness	95.6 (91.7)	99.7 (98.4)	99.6 (99.1)
Reflections (unique)	13635	7128	9195
Multiplicity	12.1	5.4	4.33
Wilson plot *B* (Å^2)	21.9	25.8	30.2
Optical resolution‡ (Å)	1.37	1.52	1.49
Refinement			
R_{work}§	0.171	0.202 (0.244)	0.227 (0.266)
R_{free}§	0.228	0.222 (0.277)	0.241 (0.300)
Completeness¶	95.2	97.9 (95.1)	97.9 (96.0)
$\langle B\rangle$ (Å^2)	32.0	29.1	24.9
N_{work}	12005	6236	7789
N_{free}	1324	718	911
N_{atoms}	587	531	536
$N_{parameters\ refined}$	5280	824	834
RMSD from ideal			
Bond length (Å)	0.013¶	0.004	0.004
Bond angles (°)	1.9¶	0.9	0.8
Dihedral angles (°)		10.3	11.1
Improper angles (°)		1.34	1.18
DPI†† (Å)	0.08	0.14	0.11

† $R_{merge} = \sum_h \sum_i |I_h - I_{hi}| / \sum_h \sum_i I_h$, where I_h is the mean intensity of reflection *h*. All data with $I > -3\sigma$ are included. ‡ $W = (\sigma^2_{Patterson} + \sigma^2_{sph})^{1/2}$, the expected minimum distance between two resolved atom peaks (Vaguine *et al.*, 1999). § Calculated with reflections with $F > 2.0\sigma$. ¶ Structure was refined with *SHELX*97 with distance restraints corresponding to bond lengths and angles. Values reported were calculated using *CNS* v. 1.0. †† Diffraction data-precision indicator, $\sigma(x) = (N_{atoms}/N_{obs})^{1/2} c^{-1/3} \times d_{min} R_{free}$.

3. Results

3.1. Crystallization

Strands A and B of vkIX were annealed together prior to setting up hanging-drop vapor-diffusion trays. Crystals typically grew within 1–2 d, but often required several weeks to grow. vkIX-AA only crystallized at or near room temperature, but did so under many different conditions. (i) In the absence of drug, Ba^{2+} was required when the precipitant was MPD. However, a similar dependence on Ba^{2+} was not observed when crystals were grown using ammonium sulfate as a precipitant (crystals of vkIX-AA could also be obtained from ammonium sulfate without drug being present). In appearance, the morphology of these crystals was identical to that of crystals grown in the presence of neomycin B. (ii) vkIX-AA readily crystallized in the presence of a variety of aminoglycosides (neomycin B, paromomycin, kanamycin, amikacin and streptomycin) when ammonium sulfate was the precipitant; the unit-cell parameters and resolution limits varied with the addition of each drug. (iii) A single crystal was obtained of vkIX-AA (no drug present) using NaCl as a precipitant, but it required months to crystallize. (The attempt to soak this single example with a cryoprotectant dissolved the crystal, preventing us from collecting data on this crystal. Since its morphology was identical to the crystals obtained under other conditions, we assume it was also a crystal of vkIX-AA.)

3.2. Structure determination

The homoduplex vkIX-AA was initially solved by SIRAS and required data collected from a native crystal and a crystal into which Tl^+ was soaked. Our maps were readily interpretable and it was soon apparent that there was only density for 22 nucleotides and not the anticipated 23, as would be required for the heteroduplex. This observation, coupled with the recognition that for some nucleotides the base type and

corresponding density were poorly matched, led us to consider the possibility that we had obtained crystals of the homoduplex vkIX-AA and not of the intended heteroduplex.

Our suspicions were confirmed by polyacrylamide gel electrophoresis of redissolved crystals (Fig. 2). Redissolved crystals were labeled at the 5′ end with a ^{32}P phosphate tag. RNA found within the crystals migrated as a single band at the same rate as vkIX-A and faster than vkIX-B. As a final test, we crystallized vkIX-A alone; no vkIX-B was present. Although the crystals were smaller than crystals grown with both strands, their appearance was the same.

The homoduplex vkIX-AA fitted the density well. Within a few rounds of fitting and simulated annealing, we arrived at R_{work} and R_{free} values of 29.8 and 31.5, respectively. Water molecules were fitted automatically, then refined manually by adding, deleting and moving waters as suggested by composite omit density maps. Subsequent rounds of simulated annealing, individual *B*-factor refinement and occupancy refinement of the water molecules resulted in final R_{work} and R_{free} values of 21.6 and 24.6, respectively, at 1.8 Å.

Crystals used to solve this initial structure were grown from solutions containing neomycin B, but no remaining density could be interpreted as arising from the drug, which is not surprising as vkIX-AA no longer retains the aminoglycoside-binding site. Although visual inspection of the packed duplexes suggests that ample room is available to accommodate a molecule the size of neomycin B, the relatively low *R* values support the conclusion that no drug is in complex with this homoduplex at significant occupancy.

The above structure was used as a molecular-replacement search model to solve the structures of vkIX-AA prepared in other ways: vkIX-AA complexed with Ba^{2+} cations, vkIX-AA complexed with Tl^{+} cations and vkIX-AA collected to a resolution of 1.6 Å. We have refined all of these variants to reasonably low *R* values (Table 1). The structures of the variants of vkIX-AA agree with one another qualitatively, but variations exist.

3.3. The structure of vkIX-AA

For all of the variants studied, vkIX-AA forms ten base pairs, leaving single cytosines to overhang the 3′ ends of both strands. Four G·C base pairs make up the two closing base pairs on both ends of the helix. Between the Watson–Crick base-pair regions are six consecutive mismatch base pairs: two identical G·U wobble base pairs and four A/C pairings: two AC *trans* Watson–Crick/Hoogsteen pairs and two AC *trans* Hoogsteen/Sugar edge pairs (the nomenclature of AC mismatches is according to Leontis *et al.*, 2002) (Fig. 3). Both AC mismatches were recently reported in other RNA crystal

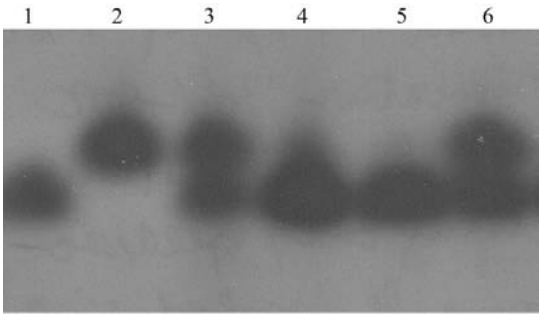

Figure 2
Polyacrylamide gel electrophoresis of component strands of vkIX and dissolved crystal. Individual strands or redissolved crystal were labeled with ^{32}P at the 5′ end. Lane 1, strand A only; lane 2, strand B only; lane 3, strand A + strand B; lane 4, redissolved crystal; lane 5, redissolved crystal + strand A; lane 6, redissolved crystal + strand B.

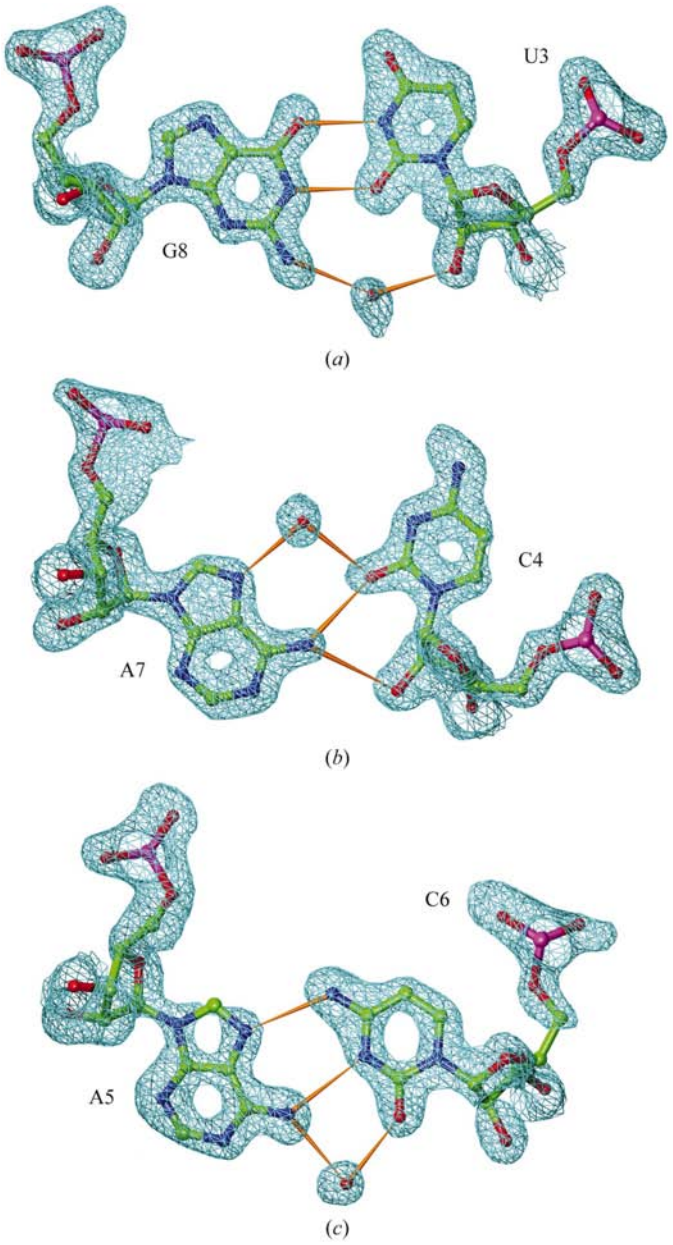

Figure 3
Mismatched pairs observed in vkIX-AA. (*a*) One of two G·U wobble pairs from vkIX-AA. A water molecule extends the hydrogen-bonding network to include N2 of G8 and O2′ of U3. (*b*) One of two AC *trans* Hoogsteen/sugar-edge base pairs. N6 of A7 forms a bifurcated hydrogen bond to O2 and O2′ of C4. A water molecule bridges O2 of C4 and N7 of A7. (*c*) One of two AC Watson–Crick/Hoogsteen base pairs. In addition to direct base-to-base hydrogen bonds, a water molecule bridges O2 of C6 and N6 of A5. (Hydrogen bonds are shown in yellow. In all panels the electron density is contoured at 1.5σ.)

structures (Fukai *et al.*, 2000; Ban *et al.*, 2000). In all three types of mismatches a water molecule forms part of the hydrogen-bonding network. However, only the mediating waters found in the AC *trans* Watson–Crick/Hoogsteen pairs are observed in all three variants. Although there is only a single hydrogen bond formed directly between the bases A7 and C4, the amino group of A7 also forms a hydrogen bond to the 2′-OH of C4 of the opposite strand. The interatomic distances for the heavy atoms that contribute to the hydrogen-bonding network for the mismatch pairs are shown in Table 2; only minor deviations are observed between the three variants.

Overall, the homoduplex has twofold symmetry, as would be expected. The gross consequence of the mismatches is an extraordinarily narrowed major groove. In particular, the major groove of the Ba^{2+}/MPD variant is so narrow that a phosphate O atom of A5 is within hydrogen-bonding distance of N4 of C4. With few exceptions, for a given backbone or sugar dihedral angle the rotamer class or sugar pucker is in an A-form RNA conformation and is preserved for all three duplex variants. Those exceptions are restricted to A5 and A7. For A5, the backbone angle α lies between the low-energy rotamers *gauche*$^-$ and *trans* and the angles β and γ are in *gauche*$^+$ and *trans* rotamers, respectively. The sugar pucker for A5 is 2′-*exo*. The backbone angles α and γ for A7 are both in the *trans* rotamer class. A-form values for rotamer class and sugar pucker are *gauche*$^-$ for α, *trans* for β, *gauche*$^+$ for γ and 3′-*endo* for sugar pucker. For one vkIX-AA variant, MPD/Ba^{2+}, A7 from one strand adopts normal A-form RNA torsion angles throughout its backbone, while A7 from the other strand has torsion angles analogous to those found in the other variants.

vkIX-AA in complex with Tl^+ and vkIX-AA in complex with Ba^{2+} are similar at the local level (sugar pucker and dihedral backbone angles) and with regard to helical parameters (data not shown), but the heavy-atom root-mean-square difference (ha-RMSD) between these two structures is 1.88 Å, whereas the ha-RMSD between the two variants for which ammonium sulfate was used as the precipitant vary by only 0.16 Å. The large global variation that exists between the structures of vkIX-AA grown in the presence of MPD and Ba^{2+} and the other two variants is the sum of many small variations in local geometry. Moreover, those differences observed between the physical constants of vkIX-AA grown in ammonium sulfate and vkIX-AA grown in MPD in complex with Ba^{2+} are not localized to those nucleotides that interact directly with Ba^{2+}; rather, such differences are spread somewhat evenly throughout the duplex.

Since the structures of the two variants of vkIX-AA crystallized in ammonium sulfate (with and without Tl^+ soaked in) are virtually identical with regard to ha-RMSD, either the binding of Ba^{2+} or the type of precipitant (MPD *versus* ammonium sulfate) is responsible for the differences observed here. However, without additional data it is impossible to distinguish between these two possibilities.

Table 2
Atom–atom distances for the hydrogen bonds in the observed mismatches.

Distances are the average of the values from both examples of a given mismatch type.

		Distance (Å)		
Base pair	Atoms	No metal/AS	Ba^{2+}/MPD	Tl^+/AS
G·U	N3(U3)–O6(G8)	2.74	2.71	2.70
	O2(U3)–N1(G8)	2.75	2.68	2.68
	O2′–water	2.83	—	—
	Water–N2(G8)	3.02	—	—
AC *trans*	N6(A7)–O2(C4)	3.00	2.88	2.88
Hoogsteen/	N6(A7)–O2′(C4)	3.32	3.31	3.31
sugar face	N7(A7)–water	2.82	—	—
	Water–O2(C4)	2.93	—	—
AC *trans*	N4(C6)–N7(A5)	3.04	2.96	2.96
Watson–Crick/	N3(C6)–N6(A5)	3.21	3.06	3.06
Hoogsteen	O2(C6)–water	2.84	2.59	2.59
	Water–N6(A5)	2.81	2.89	2.89

vkIX-AA does not form pseudo-continuous helices as do many other short RNA duplexes. Rather, each duplex is in contact with four other duplexes: the overhanging nucleotides at each end are buttressed into the minor grooves of two other duplexes (Fig. 4). Several RNA–RNA and water-mediated RNA–RNA interactions occur at the packing interface to stabilize the duplex associations. The contact surface between any two duplexes is about 485 Å^2. Therefore, roughly 50% of any given duplex is in contact with other duplexes.

3.4. Metal binding

Both Ba^{2+} and Tl^+ cations bind vkIX-AA within the major groove of vkIX-AA, but at slightly different positions. The

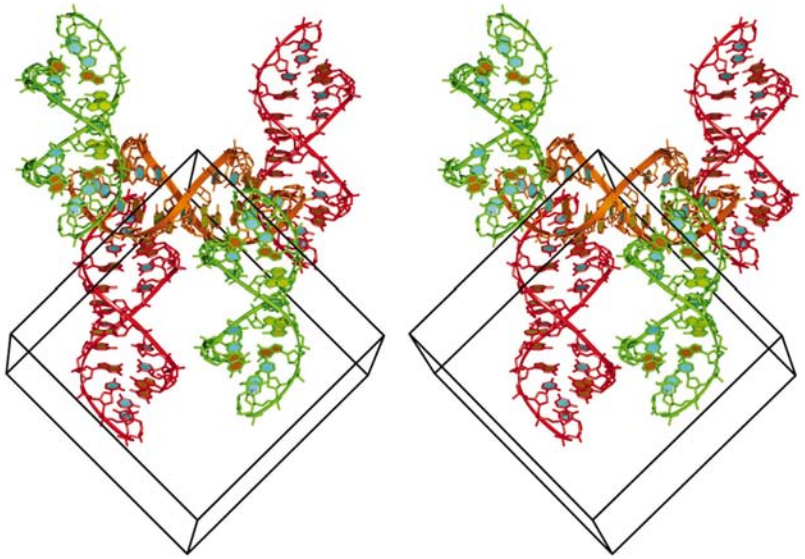

Figure 4
Packing of vkIX-AA in stereo. Duplexes are rendered in various colors for clarity. The image was created using the program *RIBBONS* (Carson, 1991) and rendered using the program *POV-Ray* (http://www.povray.org).

Ba^{2+} cations bind in the plane of the G·U pairs and form nanocoordinate complexes in which three carbonyl O atoms (O6 of G8, O4 of U3 and O6 of G2) and six waters serve as ligands (Figs. 5*a* and 5*b*). In contrast, the Tl^{+} cations bind in the planes of the reverse AC *trans* Watson–Crick/Hoogsteen base pairs and coordinate N7 of A7, O^2 and N3 of C4, O4 of U3 and three water O atoms (Figs. 5*c* and 5*d*).

3.5. Effect of aminoglycosides on crystallization

The heteroduplex vkIX-AB contains an aminoglycoside-binding site. Several attempts were made to crystallize this RNA construct in complex with one of the common aminoglycosides. We were able to crystallize the RNA (later determined to be vkIX-AA) in the presence of numerous aminoglycosides under very standard conditions containing 50 m*M* Tris–HCl pH 6.4, 1.4 *M* ammonium sulfate, 20 m*M* $MgCl_2$, 18 m*M* spermine and two molar equivalents of antibiotic. All samples crystallized in the $P4_3$ space group and all with unique unit-cell parameters and resolution. Crystals of vkIX-AA cyrstallized in the presence of kanamycin had the most unusual unit-cell parameters ($a = b = 41.2$, $c = 54.7$ Å) and a diffraction resolution limit of only 2.7 Å. The structure obtained from crystals of vkIX-AA grown in the presence of neomycin B was fully refined (described above) and this structure was used as an initial model to partially refine the structure of vkIX-AA from crystals grown in the presence of kanamycin to an *R* value of about 33% (data not shown). In neither example is there any indication of electron density that could be interpreted as a relatively large-molecular-weight ligand. In both cases, since the solvent fraction is about 60%, there should be sufficient volume to accommodate a ligand the size of either neomycin B or kanamycin.

4. Discussion

4.1. Crystallization

We attempted to crystallize the heteroduplex r(GGUCACAGCCC)/r(GGCUGAAGUCCG) under a variety of conditions. However, when crystals did appear, they were always of the homodimer composed of two 11-mer strands, which we termed vkIX-AA, and not of the expected heteroduplex. Several variants of the homodimer were obtained, all displaying the same general features: ten base pairs, of which six are mismatches, and overhanging cytosines on the 3′ ends. This duplex, vkIX-AA, is unusual in that the majority of the helix is composed of non-Watson–Crick base pairs. Two of the vkIX-AA variants have metal ions bound, Ba^{2+} ions in one structure and Tl^{+} ions in the other structure.

Since the majority of the base pairings in vkIX-AA are mismatched pairs, the duplex is not expected to be very stable. Nonetheless, this duplex does exist in the crystalline state. Partly, this is because few non-canonical A-form RNA conformations were adopted by the nucleotides and in part this is because each nucleotide remains stacked and paired to some degree with another nucleotide. However, these two factors alone are not enough to explain why the crystalline state of the homodimer of A (the state of strand B, or the 12-mer strand, is unknown) is preferred to the crystalline state of the heterodimer. Using current thermodynamic data of nearest-neighbor contributions of base-pair formation, the heterodimer is expected to be about 12–16 kJ mol^{-1} more stable than the homodimer (Mathews *et al.*, 1999). (This calculation does not take strand B into account. It could be that the BB homodimer, for example, is so stable as to drive the reaction from the heteroduplex.) We believe that the reaction pathway from vkIX to vkIX-AA involves a small population of vkIX-AA, a structure that is

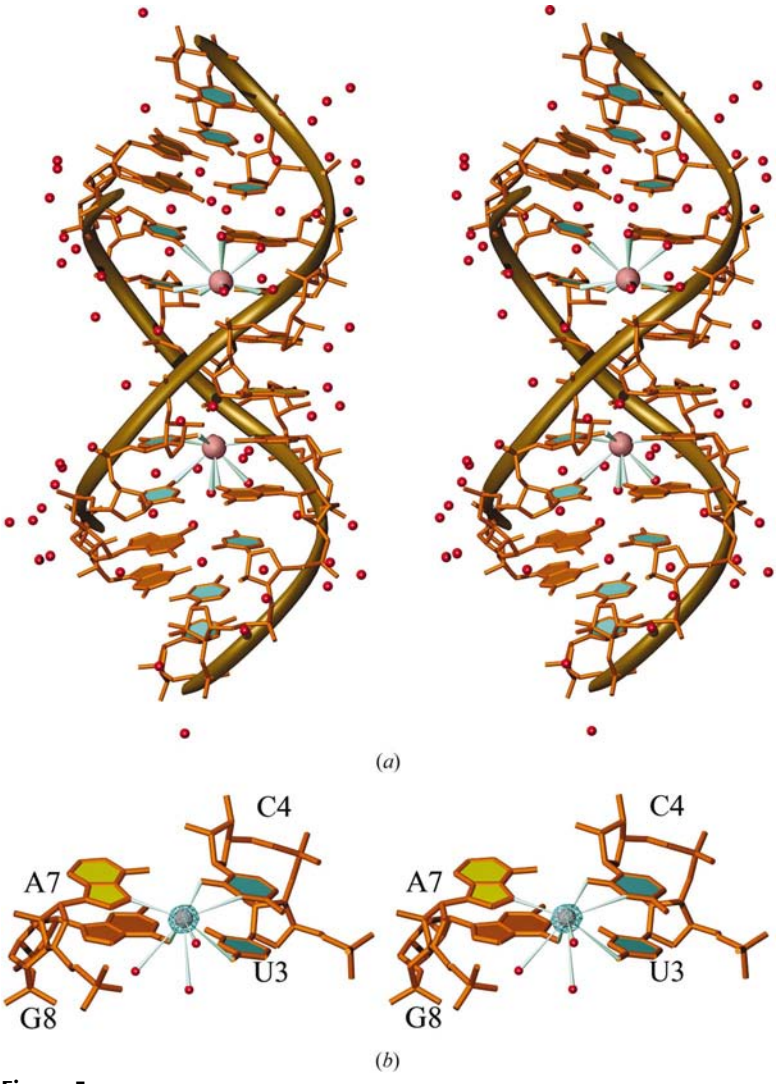

Figure 5
Metal coordination to vkIX-AA. Coordination bonds are shown in blue and waters are shown as red spheres. (*a*) The two Tl atoms (magenta) coordinated to vkIX-AA. (*b*) Detail of one of the two Tl^{+} sites. Tl^{+} coordinates to N7 of A7, O2 and N3 of C4, O4 of U3 and three water molecule O atoms. The image was created using the program *RIBBONS* (Carson, 1991) and rendered using the program *POV-Ray* (http://www.povray.org).

exceptionally conducive to crystallization, nucleating into crystals, which then grow by drawing additional strand A from the pool of vkIX. The crystallization of vkIX-AA appears to be an unusual case of how crystal packing forces can dominate an RNA into forms not appreciably observed in solution.

It has been apparent since about 1990 that small (<20 nucleotides) RNA hairpins prefer to crystallize as homodimers rather than as hairpin monomers. Although the reasons for this have not been determined in any detail, several factors regularly encountered in RNA crystallization (such as high RNA and salt concentrations) are assumed to promote dimerization. The conditions necessary for the observation of homodimers in solution are not extreme: examples exist in which dimers are observed by NMR at salt concentrations as low as 100 m*M* and RNA concentrations of 1 m*M* or lower (for an example, see Butcher *et al.*, 1997). However, there are no examples of small RNA hairpins that have been crystallized. Several medium-size RNAs absent of protein have been crystallized as hairpins, but optimism that RNA haipins above a certain length (~30 nucleotides) naturally escape the controlling forces of dimerization must be tempered with the recent observation that a 45-nucleotide hairpin crystallized as a dimer (Correll *et al.*, 1998, 1999; Luca *et al.*, 2000). So, whatever the contributing factors are, they appear to be profound and to extend beyond the contributing parameters of salt and RNA concentration, falling under the general term of crystal packing forces.

The dimerization of hairpins and the homodimerization of vkIX-AA are obvious examples of the crystallization process altering the structure of the RNA as found in solution; conclusions drawn from the crystal structures, normally applied to the corresponding RNAs in solution, are consequently tempered. However, most often there are no obvious pathologies. In fact, we know that the structures of small RNA duplexes and somewhat larger hairpins can retain their solution conformations, which strongly argues for the continued crystallographic study of small RNA alone and in complex with drugs and drug-like ligands (Rife *et al.*, 1999; Rife, 2003). (The original goal of this work, to crystallize a small RNA–aminoglycoside complex, has recently been reported; Vicens & Westhof, 2001.) However, there are other examples where subtle differences exist between the crystal and solution structures. The UCU bulge found in the Tat-binding site of TAR of HIV is one such case. In the crystal structure, the three pyrimidines fold around a single Ca^{2+} atom, forming an extrahelical bulge in a manner that is not supported by NMR data, suggesting that some features of the crystal structure are artifacts (Ippolito & Steitz, 1998; Puglisi *et al.*, 1992; Aboul-ela *et al.*, 1995).

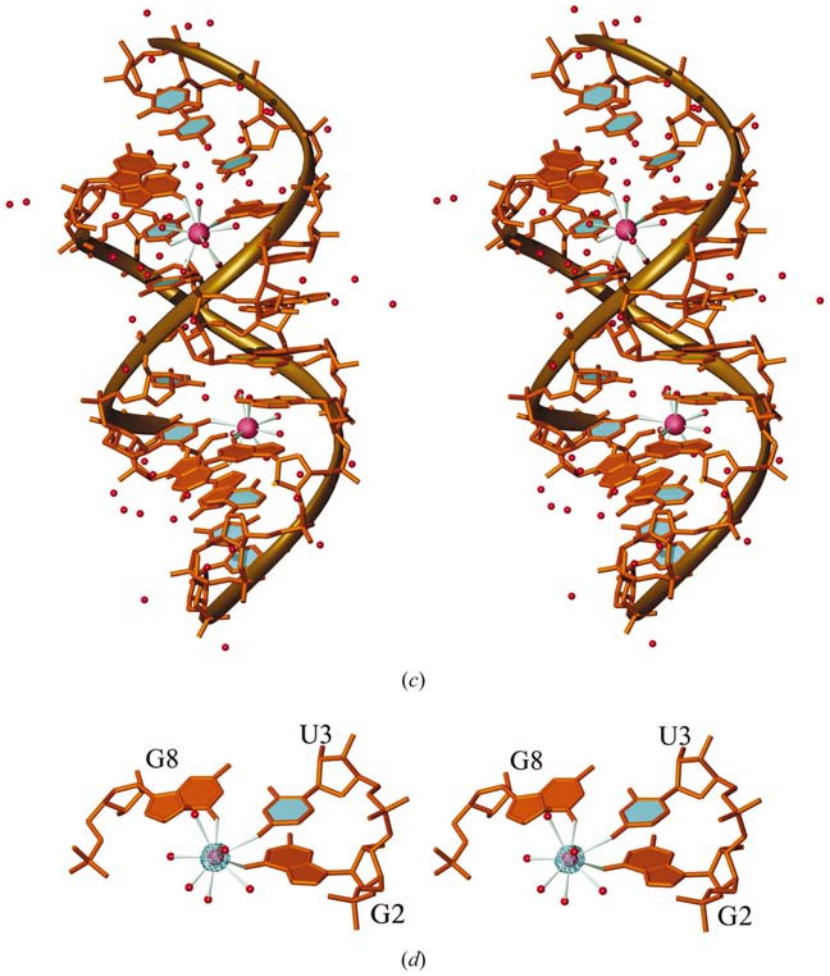

Figure 5 (continued)
Metal coordination to vkIX-AA. (*c*) The two Ba atoms (magenta) coordinated to vkIX-AA. (*d*) Detail of one of the two Ba^{2+} sites. Ba^{2+} coordinates to O6 of G8, O4 of U3, O6 of G2 and six water molecule O atoms.

4.2. Metal binding

The Ba^{2+} and Tl^{+} ions observed here provide additional examples of the tendency for metal ions to bind within the major grooves of RNA (for examples, see Correll *et al.*, 1997; Cate & Doudna, 1996; Wedekind & McKay, 1999). RNA naturally displays the right stereochemistry for metal binding within the major grooves: a favorable distance between adjacent base pairs and the availability of appropriate ligands (*i.e.* O6 of guanosines). In the context of secondary structure, these sites appear to be preformed (Williams & Maher, 2000). Nevertheless, such preformed sites are not universal for all metals. As observed here, vkIX-AA binds Tl^{+} and Ba^{2+} differently and Mg^{2+} not at all. The positions of bound Tl^{+} and Ba^{2+} atoms are close but not identical and coordinate the RNA through different functional groups. Nine is the expected coordination number for Ba^{2+} and that is the number observed here: three carbonyl O atoms from vkIX-AA and six O atoms from water molecules coordinate to Ba^{2+}. The crystal structure of a lead-dependent ribozyme was solved in complex with two un-

related Ba^{2+} atoms (Wedekind & McKay, 1999). The major Ba^{2+} site coordinated the ribozyme through four first-sphere contacts (N1 of an adenosine, C2 of a cytidine, an O2′ and a phosphate O atom). No information regarding the coordination of water molecules was provided.

Several sites within the P4–P6 domain from the group 1 intron can bind Tl^{+} (Basu *et al.*, 1998). In each of these sites, the Tl^{+} binds at an AA-platform, a motif that supports potassium binding. In these cases, Tl^{+} fulfils its common role as a substitute for K^{+}. In binding to AA-platforms, Tl^{+} or K^{+} resides in the major groove and coordinates to four O atoms (two carbonyl O atoms and two phosphate O atoms) and a single base N atom. Presumably, the remaining coordination sites are filled with water O atoms. In the structure of vkIX-AA, Tl^{+} also binds in the major groove, which is accomplished by coordinating two base O atoms and two N atoms. Water O atoms fill the remaining four coordination sites.

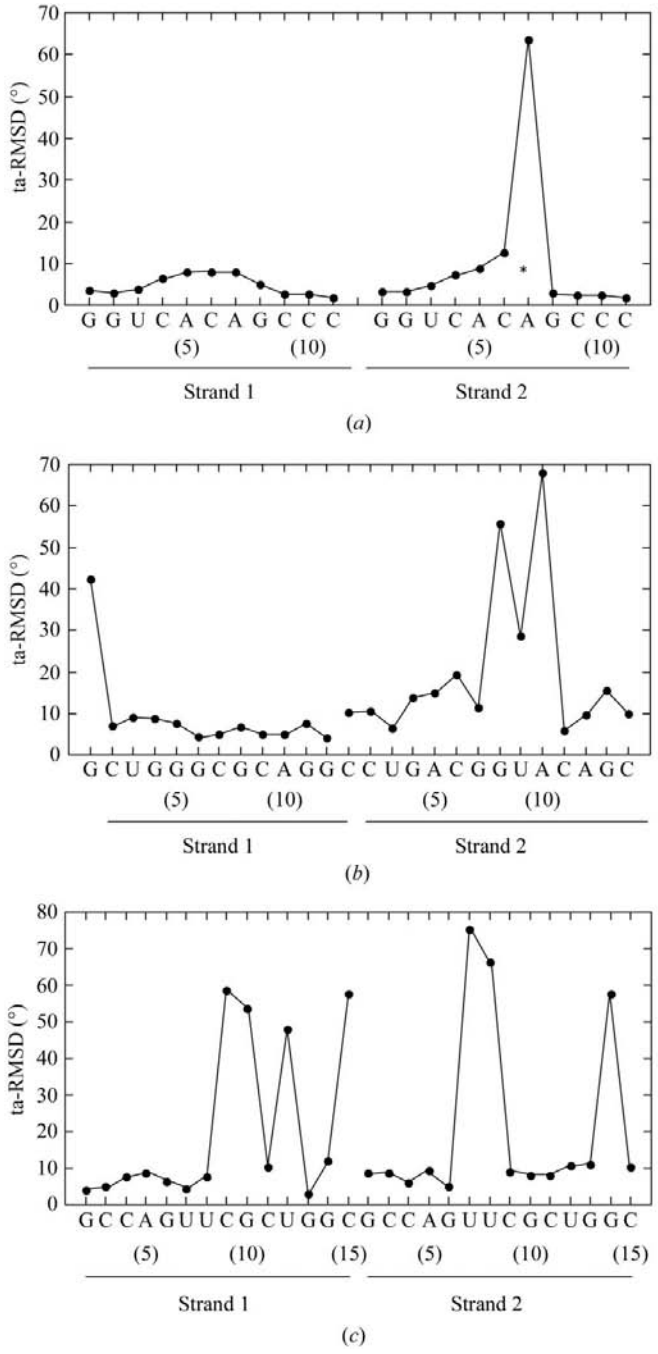

Figure 6
Per-residue plots of ta-RMSD values between two conformations of three sets of RNA duplexes. (*a*) Comparison between strands 1 and strands 2 of vkIX-AA (AS/no metal) and vkIX-AA (MPD/Ba^{2+}). (*b*) Comparison between the two most dissimilar duplexes within the asymmetric unit of the RBE duplex reported by Hung *et al.* (2000). (*c*) Comparison between the two most dissimilar duplexes within the asymmetric unit of the UUCG duplex reported by Anderson *et al.* (1999).

Although crystals of vkIX-AA were also grown in the absence of Ba^{2+} and Tl^{+} and in the presence of Mg^{2+}, no Mg^{2+} binding was observed. We have no conclusive explanation for this. We do note that Mg^{2+} has been observed to bind within the major groove at select sites in RNA, but usually as the hexahydrate species (for a recent discussion, refer to Misra & Draper, 2001). Tl^{+} and Ba^{2+} ions both coordinate directly to vkIX-AA through nucleobase O and N atoms, which requires that they shed several inner shell waters before binding, something that Mg atoms do with greater reluctance (Peschke *et al.*, 1998). Ammonium ion binding sites have been identified in RNA (Klosterman *et al*, 1999; Lu & Draper, 1994), but the recognition of such a site in an electron-density map would be problematic. Only under very restrictive hydrogen-bonding networks could one differentiate between a water molecule and an ammonium ion. Therefore, only in the most favorable circumstances would it be possible to identify an ammonium ion in an electron-density map.

4.3. Conformational variability

The global conformation of the Ba^{2+}-containing variant differs from those of the Tl^{+}-containing variant and the variant containing no bound metal by about 1.9 Å heavy-atom root-mean-square difference (ha-RMSD), which is a significant difference. Without superimposing the various duplexes and relying only upon the comparisons of torsion angles and local helical parameters, one would probably conclude that the structures of the variants were essentially identical by virtue of those values being so similar. Only when the structures are analyzed by methods that probe both local changes and global changes is it apparent that the large global changes observed in the structures of vkIX-AA arise from numerous small local deviations. Where larger than average variations do exist, they are not confined to the site of Ba^{2+} binding, as might be expected. Rather, small deviations in local geometry exist throughout the duplexes. We cannot identify the root cause of the conformational variability, but offer several possibilities. (i) In the case of the VKIX-AA/Ba^{2+} structure, MPD was used as the precipitant, while ammonium sulfate was used to crystallize the other two forms of VKIX-AA. The different ionic strengths of the two precipitants may account for the variation in the vkIX-AA structures. (ii) The greater positive charge of the bound Ba^{2+} atoms better neutralizes the negative phosphate charges on one side of the helix, which may result in helix bending (Williams & Maher, 2000). (iii) The

Ba^{2+}-binding site is only partially pre-formed, whereas the Tl^{+}-binding site is wholly formed, thus requiring additional conformational changes to accommodate the Ba^{2+} atoms.

Two other groups have recently reported on the conformational variability of RNA duplexes similar in size to vkIX-AA. Hung *et al.* (2000) reported on the crystal structures of four unique duplexes of the Rev binding element of HIV-1 present in a crystal that comprised two distinct structural variants. The ha-RMSD between the two variants of the 12-mer/14-mer heteroduplex (referred to here as the RBE duplex; PDB code 1duq, duplexes 3 and 4) was about 1.6 Å. The authors could identify specific regions where variation between the two structures existed; these regions were reported to be at the bulged nucleotides. Anderson *et al.* (1999), on the other hand, compared the conformations of three unique duplexes that included an internal loop composed of four mismatches (referred to here as the UUCG duplex; PDB code 1qbp, duplexes 1 and 3). The ha-RMSD between the two most dissimilar duplexes was about 0.9 Å. The authors identified the sites of greatest variation as the mismatched region and the helical ends. In both cases, visual inspection of the overlaid structures was sufficient to identify the regions of greatest conformational variability. Although the global variation is greater for the two most dissimilar structures of vkIX-AA than it is for either of the above duplexes, we were unable to ascribe any specific region(s) as the genesis of the overall conformational variability.

Rife *et al.* (1999) relied upon torsional-angle root-mean-square difference (ta-RMSD) as one of the methods used to compare conformational variability between NMR and X-ray diffraction RNA structures. Here, we used the same analysis on a per-nucleotide basis to highlight specific regions of conformational variability between dissimilar variants of vkIX-AA (Fig. 6*a*). For comparison, the same analysis was performed on the most dissimilar duplex pairs reported by Hung *et al.* (2000) and Anderson *et al.* (1999) (Figs. 6*b* and 6*c*). We also calculated the overall ta-RMSD for each of the duplexes as 14.8, 22 and 30.9° for vkIX-AA, RBE duplex and UUCG duplex, respectively. The first observation is that this analysis supports our claim that the large global variation observed between vkIX-AA variants is indeed the result of many small local variations; the per-residue ta-RMSD value is below 10° for all of the nucleotides, with the exception of A7 in strand 2, for which it is about 65°. However, the large ta-RMSD value found for A7 in strand 2 arises entirely from the 'crankshaft' rotation of the backbone angles α and γ of this nucleotide in the Ba^{2+}/MPD variant of vkIX-AA, a change which has little consequence on the overall structure; the ha-RMSD between strand 1 and strand 2 of the Ba^{2+}/MPD variant is only 0.3 Å. (Removal of those two angles from the overall ta-RMSD calculation lowers the ta-RMSD value to only 5.9°.) The second observation is that this analysis correctly identifies the same hypervariable sites in the UUCG duplex and the RBE duplex as reported by Anderson *et al.* (1999) and Hung *et al.* (2000), respectively.

A final surprising observation is that there is a lack of correlation between global ha-RMSD and ta-RMSD values. For example, the two vkIX-AA structures have the highest ha-RMSD value but the lowest ta-RMSD value, whereas the UUCG duplex structures have the lowest ha-RMSD value but by far the highest ta-RMSD value. In the three comparisons, the nucleotides with the highest correspondence in structure has ta-RMSD values that hovered around 5–10°, which means that if this trend holds for other RNAs the existence of high local variation between two RNA structures could be predicted solely on the basis of the magnitude of the overall ta-RMSD. (The expected overall ta-RMSD value for two structures similar in local geometry would be around 5–10°.)

4.4. Aminoglycosides

Crystals of vkIX-AA were grown under several conditions, many of which only varied in the aminoglycoside added. Although no drug was found in complex with the RNA, the addition of an aminoglycoside did alter the quality and physical properties of the crystals grown. For example, crystals of vkIX-AA grown in the presence of neomycin B diffracted to 1.6 Å, while crystals obtained in presence of kanamycin diffracted to only 2.7 Å, with different unit-cell parameters. From these results, we conclude that the drugs did not bind tightly to vkIX-AA, but nevertheless had noticeable effects on crystal growth. We believe that aminoglycosides may be useful crystallization additives for RNA crystal growth. Therefore, they can be considered in the same way as other polyamine additives such as spermidine and spermine, which are often added to crystallization reagents yet regularly do not tightly bind the RNA in the crystalline form and are therefore often invisible to X-ray diffraction analysis.

This research was carried out in part at the National Synchrotron Light Source, Brookhaven National Laboratory, which is supported by the US Department of Energy, Division of Materials Sciences and Division of Chemical Sciences. The authors wish to thank Drs Leon Flaks, Joel Berendzen and Robert Sweet for their help at BNL. We also wish to thank Professor H. T. Wright for his advice on all aspects of this project. This work was supported by a research grant to JPR from the Deafness Research Foundation.

References

Aboul-ela, F., Karn, J. & Varani, G. (1995). *J. Mol. Biol.* **253**, 313–332.
Abrahams, J. P. (1997). *Acta Cryst.* D**53**, 371–376.
Anderson, A. C., O'Neil, R. H., Filman, D. J. & Frederick, C. A. (1999). *Biochemistry*, **38**, 12577–12585.
Baeyens, K. J., DeBondt, H. L., Pardi, A. & Holbrook, S. R. (1996). *Proc. Natl Acad. Sci. USA*, **93**, 12851–12855.
Ban, N., Nissen, P., Hansen, J., Moore, P. B. & Steitz, T. A. (2000). *Science*, **289**, 905–920.
Basu, S., Rambo, R. P., Strauss-Soukup, J., Cate, J. H., Ferré-D'Amaré, A. R., Strobel, S. A. & Doudna, J. A. (1998). *Nature Struct. Biol.* **5**, 986–992.
Brünger, A. T., Adams, P. D., Clore, G. M., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst.* D**54**, 905–921.

Butcher, S. E., Dieckmann, T. & Feigon, J. (1997). *J. Mol. Biol.* **268**, 348–358.
Carson, M. (1991). *J. Appl. Cryst.* **24**, 958–961.
Cate, J. H. & Doudna, J. A. (1996). *Structure*, **4**, 1221–1229.
Correll, C. C, Freeborn, B., Moore, P. B. & Steitz, T. A. (1997). *Cell*, **91**, 705–712.
Correll, C. C., Munishkin, A., Chan, Y.-L., Ren, Z., Wool, I. G. & Steitz, T. A. (1998). *Proc. Natl Acad. Sci. USA*, **95**, 13436–13441.
Correll, C. C., Wool, I. G. & Munishkin, A. (1999). *J. Mol. Biol.* **292**, 275–287.
Cruse, W. B. T., Saludjian, P., Biala, E., Strazewski, P., Prange, T. & Kennard, O. (1994). *Proc. Natl Acad. Sci. USA*, **91**, 4160–4164.
Deng, J. & Sundaralingam, M. (2000). *Nucleic Acids Res.* **28**, 4376–4381.
Fukai, S., Nureki, O., Sekine, S., Shimada, A., Tao, J., Vassylyev, D. G. & Yokoyama, S. (2000). *Cell*. **103**, 793–803.
Holbrook, S. R., Chaejoon, C., Tinoco, I. Jr & Kim, S.-H. (1991). *Nature (London)*, **353**, 579–851.
Holbrook, S. R. & Kim, S.-H. (1997). *Biopolymers*, **44**, 3–21.
Hung, L. W., Holbrook, E. L. & Holbrook, S. R. (2000). *Proc. Natl Acad. Sci. USA*, **97**, 5107–5112.
Ippolito, J. A. & Steitz, T. A. (1998). *Proc. Natl. Acad. Sci. USA*, **95**, 9819–9824.
Ippolito, J. A. & Steitz, T. A. (2000). *J. Mol. Biol.* **295**, 711–717.
Klosterman, P. S., Shah, S. A. & Steitz, T. A. (1999). *Biochemistry*, **38**, 14784–14792.
La Fortelle, E. de & Bricogne, G. (1997). *Methods Enzymol.* **276**, 472–493.
Leontis, N. B., Stombaugh, J. & Westhof, E. (2002). *Nucleic Acids Res.* **30**>, 3497–3531.
Lu & Draper (1994). *J. Mol. Biol.* **244**, 572–585.
Luca, J., Hainzl, T., Oubridge, C., Scott, W. G., Li, J., Sixma, T. K., Wonacott, A., Skarzynski, T. & Nagai, K. (2000). *Structure*, **8**, 527–540.
Mathews, D. H., Sabina, J., Zuker, M. & Turner, D. H. (1999). *J. Mol. Biol.* **288**, 911–940.
Milligan, J. F., Groebe, D. R., Witherell, G. W. & Uhlenbeck, O. C. (1987). *Nucleic Acids Res.* **21**, 8783–8798.
Misra, V. K. & Draper, D. E. (2001). *Proc. Natl Acad. Sci. USA*, **98**, 12456–12461.
Moore, P. B. (1995). *Acc. Chem. Res.* **28**, 251–256.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Peschke, M., Blades, A. T. & Kebarle, P. (1998). *J. Phys. Chem. A*, **102**, 9978–9985.
Puglisi, J. D., Tan, R., Calnan, B. J., Frankel, A. D. & Willamson, J. R. (1992). *Science*, **257**, 76–80.
Rife, J. P. (2003). *Burger's Medicinal Chemistry*, 6th ed., Vol. 2, edited by D. J. Abraham, ch. 7. New York: Wiley-Interscience. In the press.
Rife, J. P., Stallings, S. A., Correll, C. C., Dallas, A., Steitz, T. A. & Moore, P. B. (1999). *Biophys. J.* **76**, 65–75.
Scaringe, S. A. (2001). *Methods*, **23**, 206–217.
Sheldrick, G. M. & Schneider, T. R. (1997). *Methods Enzymol.* **277**, 319–343.
Tanaka, Y., Fujii, S., Hiroaki, H., Sakata, T., Tanaka, T., Uesugi, S., Tomita, K. & Kyogoku, Y. (1999). *Nucleic Acids Res.* **27**, 949–955.
Vaguine, A. A., Richelle, J. & Wodak, S. (1999). *Acta Cryst.* D**55**, 191–205.
Vicens, Q. & Westhof, E. (2001). *Structure*, **9**, 647–658.
Wedekind, J. E. & McKay, D. B. (1999). *Nature Struct. Biol.* **6**, 261–268.
Williams, L. D. & Maher, L. J. III (2000). *Annu. Rev. Biophys. Biomol. Struct.* **29**, 497–521.