Crystal structure of an RNA·DNA hybrid reveals intermolecular intercalation: Dimer formation by base-pair swapping

Gye Won Han[†], Mary L. Kopka[†], David Langs[‡], Michael R. Sawaya[†], and Richard E. Dickerson[†§]

[†]Molecular Biology Institute, University of California, Los Angeles, CA 90095-1570; and [‡]Hauptman–Woodward Medical Research Institute, Inc., 73 High Street, Buffalo, NY 14203

Contributed by Richard E. Dickerson, June 2, 2003

An intermolecular intercalation of base pairs was found at the CA step in the I222 crystal structure of the RNA·DNA hybrid, r(CAAAGAAAAG)·d(CTTTTCTTTG), which contains two-thirds of the polypurine tract sequence of HIV-1 with a substitution of cytosine for the initial adenine. This sequence crystallized in both $P2_12_12_1$ and I222 space groups, with an rms difference of only 0.63 Å between residues 3 to 18 of the two forms. $P2_12_12_1$ and I222 helices are both A-like, but intercalation occurs only in the I222 crystal form. The present structure shows bases stacked in parallel rather than perpendicular as in intercalated DNA (I-DNA). The base intercalation is also different from zipper-like meshing of bases seen in the center of the crystal structure of d(GCGAAAGCT), which does not have Watson–Crick base pairing. The base-step intercalation seen here is reminiscent of domain swapping in proteins; therefore, we call this phenomenon "base-pair swapping." It involves a highly mobile CA step and seems to be sequence-specific and electrostatically stable without disrupting Watson–Crick interactions. It also exhibits a large rise concurrent with unwinding of the helix (low twist). We present a base-pair swapping dimer in nucleic acids.

RNA·DNA polypurine tract | DNA intercalation

Nucleic acids have been known to adopt various deformations besides the established B, A, and Z helices. DNA can be intermediate between two types of helices (1, 2) and can form triplexes (3) and also structures such as the G-DNA tetraplexes formed by guanine-rich sequences (4–6). Four-stranded tetraplexes are also found in the guanine-rich repeats of RNA (7, 8).

Intercalation by insertion of small aromatic rings between base pairs has been known since the early 1960s (9). Most intercalators such as ethidium and acridine are nonspecific DNA intercalators (10), but some metallo-intercalators such as the rhodium intercalator, Δ-α-[Rh[(*R*,*R*)-Me_2trien]phi]$^{3+}$ [where (*R*,*R*)-Me_2trien is 2*R*,9*R*-diamino-4,7-diazadecane] (11) are known to be sequence-specific.

Here we present an RNA·DNA hybrid intermolecular intercalation occurring between two duplexes (Fig. 1*a*). This type of intercalation resembles domain swapping in proteins (Fig. 1*b*), which was first introduced to describe a dimer of diphtheria toxin in which two or more protein chains exchange identical domains to form an oligomer (12).

A different type of DNA intercalation, known as I-DNA or the I-motif structure, has been found in a single strand containing four cytidine repeats (13), by association of two strands containing two cytidine repeats (14), or four strands containing a cytidine-rich stretch (15–18) found especially in centromeric and telomeric regions (19). I-DNA involves a systematic base intercalation (20) formed by two parallel strands holding together and interdigitating into each other (Fig. 1*c*).

DNA can self-intercalate via its base pairs. A zipper-like DNA structure with intercalated bases stacking on each other instead of forming Watson–Crick G·C and A·T base pairs was first modeled in 1983 (21). Detailed structural information on this structure has recently been observed crystallographically (Fig. 1*d*) showing the unusual zipper-like structure of d(GCGAAAGCT)·d(T*C*G*A*-A*A*G*C*G*) (with one strand per asymmetric unit; asterisks denote a crystallographically related strand). In a central segment of four adenine residues flanked by two sheared G·A mismatches (G3·A6* and A6·G3*) the adenines were not base-paired but instead were intercalated (ref. 22 and Fig. 1*d*). Such zipper-like motifs have also been demonstrated by unconstrained nanosecond molecular dynamics (23).

Base stacking by self-intercalation has also been found in RNA. In yeast $tRNA^{Phe}$, the UA base step of 8–9 close to the D stem is spread ≈6.5 Å apart in the vertical direction rather than the normal 3.9 Å. In between the two bases U8 and A9, C13 stacks on both bases and ties residues of 8 and 9 to the D stem by hydrogen bonding between C13-N4 and the phosphate oxygen of A9 (24).

In the RNA·DNA hybrid structure reported here the dimer formation occurs at the outermost CA step of the I222 crystal form of an RNA·DNA hybrid designed as the first two-thirds of the HIV-1 polypurine tract: r(CAAAGAAAAG)·d(CTTTTCT-TTG). Cytosine was substituted for adenine at the 5′ end to prevent fraying of the helix. Although this sequence crystallized in both $P2_12_12_1$ and I222 space groups, only I222 intercalates (ref. 25 and M.L.K., L. Lavelle, G.W.H., H.-L. Ng, and R.E.D., unpublished data).

Materials and Methods

Oligonucleotide Synthesis and Purification. The RNA decanucleotide CAAAGAAAAG was purchased from Yale University (Keck Oligonucleotide Synthesis Facility, New Haven, CT) and purified by gel electrophoresis with gels 50 × 30 × 3 cm thick. This purification procedure was crucial in obtaining a very pure RNA strand for crystallization. The DNA strand d(CTTTTCTTTG) was synthesized by the solid-phase phosphoramidate method on an Eppendorf ECOSYN D300 synthesizer. After detritylating the 5′-dimethoxy group, various single-stranded oligonucleotides were separated from the decamer sequence on a DE52 (Whatman) column as reported by Han *et al.* (26).

Crystallization and Data Collection. The I222 crystal of the RNA·DNA hybrid was grown by the sitting-drop vapor diffusion method at room temperature from $Mg(OAc)_2$, spermidine, *n*-octyl-β-D-glucoside, 2,4-methyl-pentanediol, and sodium cacodylate at pH 5.8 (M.L.K., L. Lavelle, G.W.H., H.-L. Ng, and R.E.D., unpublished data). The droplet containing low $Mg(OAc)_2$ (5.14 mM) produced I222 crystals, whereas $P2_12_12_1$ crystals were formed with 11.94 mM of either $Mg(OAc)_2$ or $Ca(OAc)_2$. However, the

Abbreviations: I-DNA, intercalated DNA; MR, molecular replacement; cc, correlation coefficient.

Data deposition: The atomic coordinates and structure factors have been deposited in the Protein Data Bank, www.rcsb.org (PDB ID code 1JB8), and the atomic coordinates have been deposited in the Nucleic Acid Database, Department of Chemistry, Rutgers, The State University of New Jersey, Piscataway (NDB structure ID code AH0019).

§To whom correspondence should be addressed. E-mail: red@mbi.ucla.edu.

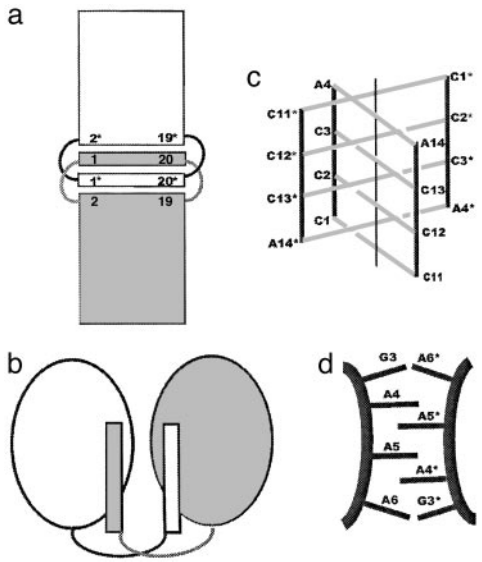

Fig. 1. Intercalation and domain swapping in nucleic acids and proteins. (*a*) Schematic diagram of this intercalated RNA·DNA structure. Helices stack head to head. Base pair 1*·20* of one helix intercalates between base pairs 1·20 and 2·19 of the other. Similarly, base pair 1·20 intercalates between base pairs 1*·20* and 2*·19* (* indicates symmetry-related molecule). (*b*) Schematic of 3D domain swapping as seen in proteins. (*c*) The central I-DNA motif region of the crystal structure of d(CCCAAT). Two strands are in the asymmetric unit, and an additional duplex (marked with asterisks) completes the four-stranded intercalated motif of the molecule. C·C base pairing is extended to A·A base pairs in this intercalation. (*d*) The schematic diagram of the zipper-like motif found in the middle of the crystal structure of d(GCGAAAGCT) with alternate intercalations of unpaired adenines and doubling of the interspacing found in this region.

lower pH seems to be more critical for I222 crystals, which formed in the $Mg(OAc)_2$ only at pH 5.8.

Unlike $P2_12_12_1$ crystals, which diffracted to high resolution (ref. 25 and M.L.K., L. Lavelle, G.W.H., H.-L. Ng, and R.E.D., unpublished data), the I222 crystal only diffracted to 2.38 Å on a R-AXIS IV imaging plate at −180°C by using CuKα radiation. X-ray data were processed by using the programs DENZO and SCALEPACK (27). Crystal data are shown in Table 1.

Structure Solution and Refinement. The structure was solved by molecular replacement (MR) with the program EPMR (28). Initially, the high-resolution crystal structure of the $P2_12_12_1$ RNA·DNA hybrid of the same sequence (ref. 25 and M.L.K., L. Lavelle, G.W.H., H.-L. Ng, and R.E.D., unpublished data) was used as a search molecule for the MR method, testing both $I2_12_12_1$ and I222

Table 1. Crystal data and refinement statistics for r(CAAAGAAAAG)/d(CTTTTCTTTG)

Crystal data	
Space group	I222
Unit cell parameters, Å	a = 45.234; b = 46.779; c = 56.560
Data collection	
Resolution, Å	2.38
No. of unique reflections	2,814
Completeness, %	98.9
R_{sym},* %	5.3
Refinement	
R_{cryst} (R_{free})	0.223 (0.251) [$F \geq 2\sigma(F)$]
	0.238 (0.264) [$F \geq 0\sigma(F)$]
rms deviations	
Bonds, Å	0.004
Angles, °	0.82

*$R_{sym} = \Sigma|I - \langle I\rangle|/\Sigma I$.

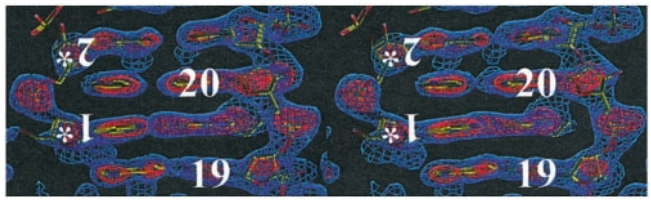

Fig. 2. 2 $F_o - F_c$ and $F_o - F_c$ omit maps calculated by omitting the problem region at base pairs 1·20 and 2·19. Red represents the 3σ level in the $F_o - F_c$ map, and blue represents the 1σ level in the 2 $F_o - F_c$ map. The intercalation bases 1* and 2* are symmetry mates of residues 1 and 2.

space groups. The best MR solutions were at $R = 0.432$ with I222 [correlation coefficient (cc) = 0.580] and $R = 0.503$ with $I2_12_12_1$ (cc = 0.407). I222 with the lower R factor was chosen as the correct space group.

The best solution ($R = 0.432$) for I222 was significantly stronger than the second solution at $R = 0.529$ (cc = 0.400), and refinement was begun with the former solution by using the program X-PLOR (29) followed by CNS (30). After rigid-body and positional refinement in X-PLOR, the R factor dropped to 0.350 ($R_{free} = 0.378$). At this stage, 2 $F_o - F_c$ and $F_o - F_c$ maps were calculated. The first two sugars and their phosphate backbone then were adjusted based on the 2 $F_o - F_c$ and $F_o - F_c$ maps, but with difficulty, because the phosphate backbone and sugar density were poor in this region. After simulated annealing and conjugate gradient least-square refinement, the R factor was 0.317 ($R_{free} = 0.357$). However, the R factor did not improve after further refinement.

Hence, it was necessary to check the accuracy of the MR solution. EPMR was re-run by using the same model but omitting the problem area: bases 1 and 2 and 19 and 20. This run provided exactly the same solution as before at $R = 0.432$ (cc = 0.580), implying that the solution was correct.

At this stage, $F_o - F_c$ and 2 $F_o - F_c$ omit maps were recalculated, and residues 1 and 2 and 19 and 20 were remodeled into the density. Surprisingly, the 2 $F_o - F_c$ map without phase bias showed that the

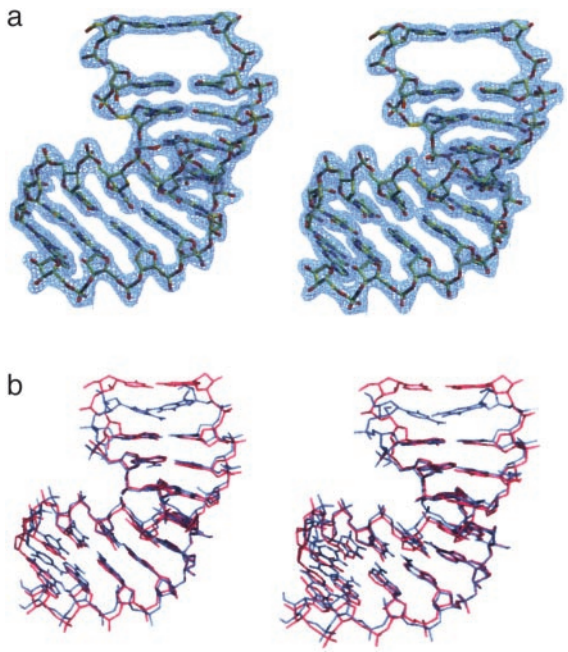

Fig. 3. (*a*) Stereo views of the 2 $F_o - F_c$ electron-density map contoured at the 1σ level. The RNA·DNA hybrid duplex with rC1·dG20 at the top of the helix. (*b*) The superposition of the I222 (purple) with the $P2_12_12_1$ (blue) structure.

BIOCHEMISTRY

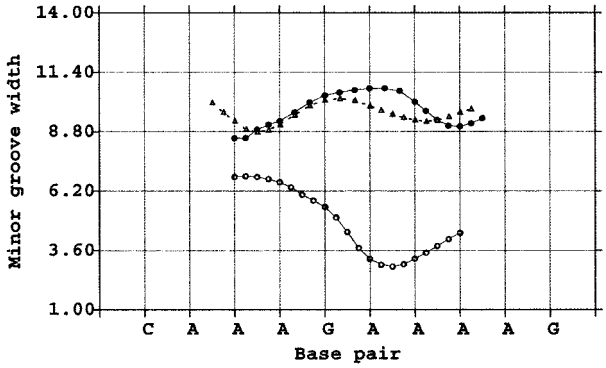

Fig. 4. Comparison of the minor groove widths between RNA·DNA and DNA·DNA in CAAAGAAAAG helices. Open triangles (I222 form) and filled circles ($P2_12_12_1$ form) represent the RNA·DNA hybrids. Open circles represent the DNA·DNA structure. The DNA·DNA shows the B-helix property of a narrow minor groove width, whereas the RNA·DNA hybrids show A-like properties. The $P2_12_12_1$ structure shows a relatively wider minor groove width than the I222 hybrid, especially in the abutting packing region. The minor groove width was calculated by CURVES (33). This program added three intermediate points between base steps to create a smooth curve.

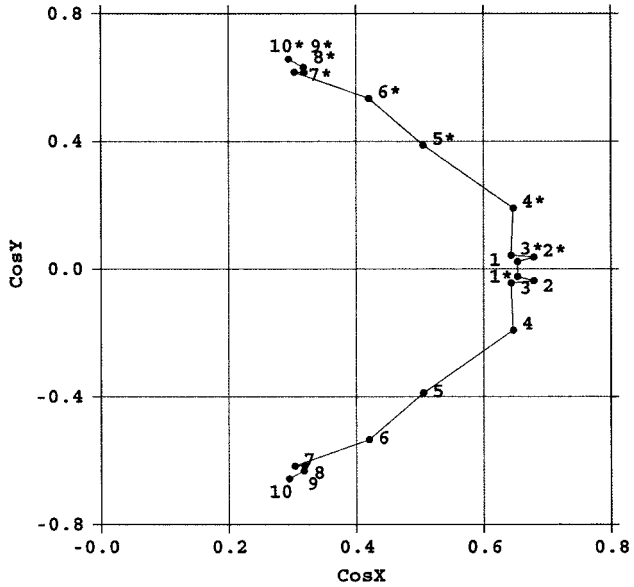

Fig. 5. Normal vector plot for intercalated 20-bp helix dimer. Helices are straight at the end and bent at the AGA steps (for a more extensive explanation of normal vector plots, see figures 4–6 of ref. 38). Notice that intercalation occurs in the middle of the two straight regions.

phosphate backbone was extended, creating a gap. By intercalating a symmetry hybrid molecule packed head to head into this density, the gap was filled (Fig. 2). The electron density of the DNA strand at residues 19 and 20 was better defined than that of steps 1 and 2 in the RNA strand. The average *B* factors of residues 1 and 2 (49.68 Å^2) were slightly higher than those of residues 19 and 20 (35.45 Å^2).

This structure solution was reconfirmed by running MR again with EPMR with a different but related RNA·DNA hybrid structure, d(CTCCTCTTC)·r(GAAGAGGAG) (31), as a search model. This molecule contained a bulge that was removed from the search model. EPMR provided the same clear solution as before, with R = 0.491 (cc = 0.443).

Final refinement was performed with CNS by using bulk solvent correction with the cross-validated maximum-likelihood approach incorporated in simulated annealing refinement (32) and with unrestrained sugar conformations for residues 1 and 2 and 19 and 20. At 10- to 2.38-Å resolution, the 2σ R factor was 0.223 (R_{free} = 0.251), and the R factor for all the data was 0.238 (R_{free} = 0.264). The final structure contains 414 nucleic acid atoms and 66 water molecules.

Results and Discussion

Overall Structure. Fig. 3*a* shows the overall structure of the I222 RNA·DNA hybrid. One of its striking features is a large gap between base pairs 1·20 and 2·19. This gap is filled by residue 1*·20* of a twofold symmetry-related molecule. As a result, base pairs are swapped at the junction of two molecules. The electron densities shown in Fig. 3*a* are reasonable and as continuous in the intercalated regions as in the helix as a whole. Overall, this I222 structure exhibits an A-like conformation. The $P2_12_12_1$ crystals formed from different crystallization conditions (see *Materials and Methods*) also exhibit an A-like helix conformation, but no intercalation occurs. The rms difference between the I222 and $P2_12_12_1$ structures is 1.32 Å for all residues and 0.63 Å for residues 3–18. The main difference (1.38 Å) occurs at the intercalation region of the first CA step (1 and 2 and 19 and 20) (Fig. 3*b*).

The A-like wide minor groove in the I222 hybrid varies in width from 8.8 to 10.3 Å (Fig. 4). The minor groove width in the $P2_12_12_1$ form of the RNA·DNA hybrid is a comparable 8.5–10.7 Å. In contrast, the DNA·DNA helix d(CAAAGAAAAG)·d(CTTTTCTTTG), analogous to our RNA·DNA, crystallized in space group C2 with a typical B-helix conformation exhibiting a narrow minor groove width (ref. 26 and Fig. 4).

Table 2. Average helical parameters

	Helical twist, °	Slide, Å	Xdisp., Å	Rise, Å	Inclination, °
I222 form (this work)					
All steps	29.0	−0.2	−3.6	3.5	2.6
Intercalation CA step	16.1	−0.2	−3.4	6.1	9.3
Without intercalation	30.7	−0.2	−4.0	3.2	2.1
Greatest similarity to standard helix	A	A	A	B	B
$P2_12_12_1$ form (ref. 25)	31.9 (31.3)	−0.1 (−1.3)	−3.3 (−2.1, −3.6)	3.0 (2.7)	5.7 (4.7, 4.4)
DNA·DNA structure of same sequence (ref. 26)	35.5 (35.2)	0.0 (−0.1)	−0.4 (−2.7)	3.3 (3.4)	−2.0 (−2.0)
Region IV* of RT-RNA·DNA (ref. 34)	32.2	0.1	−2.0	3.2	−3.9
A-DNA (ref. 35)	31.1	−1.6	−4.1	2.9	12.0
A-RNA (ref. 36)	33.8	−1.2	−3.6	2.7	17.0
B-DNA (ref. 35)	36.0	0.4	−0.1	3.4	2.4
Nogalamycin/CA step† intercalation (ref. 37)	35.1	0.3	−0.2	7.3	−1.0

Values in parentheses were calculated by using FREEHELIX; all others were calculated by using CURVES. X-displacement (Xdisp.) and inclination are calculated separately for base pairs 1–5 and 6–10 in $P2_12_12_1$ because of their sensitivity to choice of helix axis in a bent molecule.
*Base pairs 14–19 of r(AAGAAA)·d(TTTCTT) of polypurine tract.
†Values from two d(TGTACA)·nogalamycin molecules are averaged.

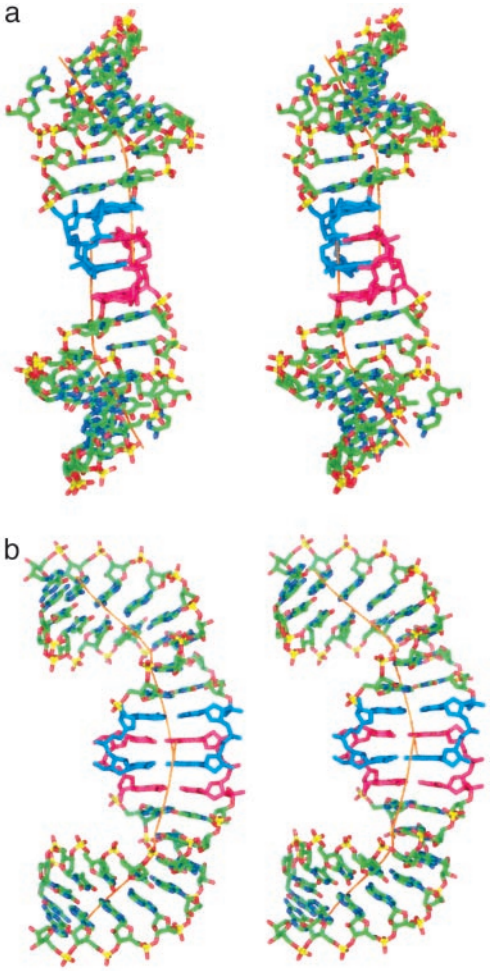

Fig. 6. (*a*) Two intercalated RNA·DNA hybrids stack head to head. Helical axes were drawn by CURVES. (*b*) Ninety-degree rotation view of the hybrid. The central intercalation region is colored in red, rC1·dG20 and blue, rC1*·dG20*.

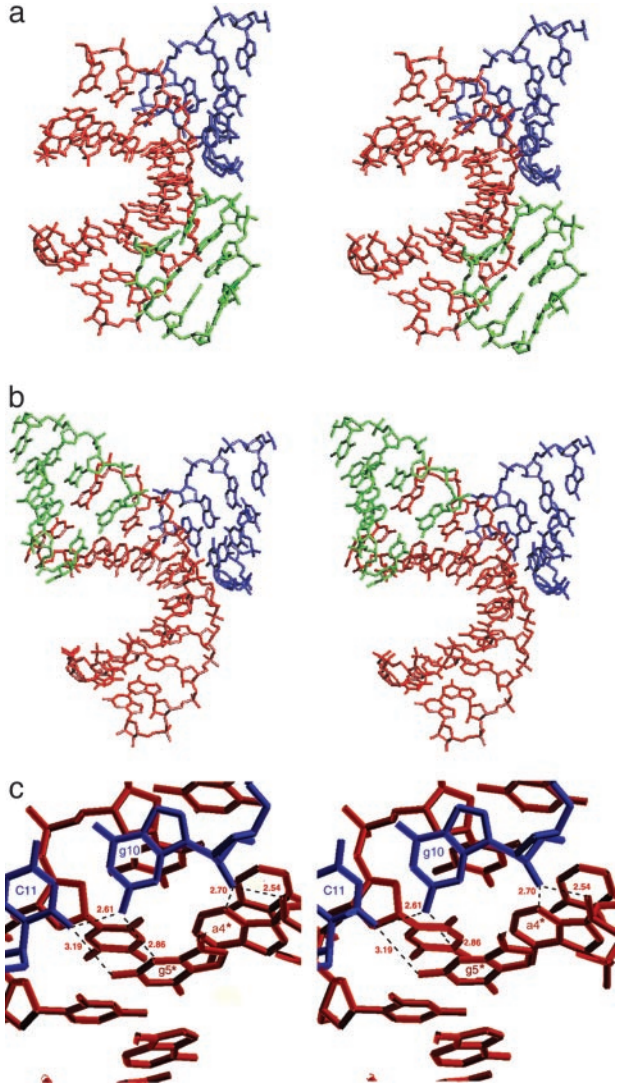

Fig. 7. Intermolecular contacts observed in the $P2_12_12_1$ (*a*) and I222 (*b*) structures. (*a*) In the $P2_12_12_1$ structure, two neighboring molecules (shown as blue and green) pack into the minor groove of the molecule (red), which is typical A-like packing. (*b*) The I222 structure has a new intercalation (green) as well as an A-like abutting packing (blue). (*c*) Close up view of the abutting interactions in the I222 structure with the hydrogen bonds indicated.

Helical Parameters. Overall helical twist, slide, and X-displacement of both the I222 and $P2_12_12_1$ forms are A-like (Table 2). If the intercalation step is omitted, mean helical twist is almost the same in the I222 and $P2_12_12_1$ forms.

The doubling of the rise at the CA intercalation step (6.1 Å) of the I222 hybrid increases the average rise value for all steps to 3.5 Å along the helix (Table 2). However, the average rise without the intercalation step (3.2 Å) is intermediate between A (2.7–2.9 Å) and B (3.4 Å) helices, more B-like than the $P2_12_12_1$ RNA·DNA hybrid (2.7 Å).

Helix Bending. Fig. 5 shows the normal vector plot for the intercalated 20-bp helix dimer consisting of r(CAAAGAAAAG)·d(CTTTTCTTTG) and r(C*A*A*A*G*A*A*A*A*G*)·d(C*T*T*T*T*C*T*T*T*C*) (asterisks indicate a neighboring molecule). Each helix is generally straight at its initial intercalation site (C1-A2), bent at the A4-G5-A6 steps, and straight again at the end of the A6-A7-A8-A9 region (for a more extensive explanation of normal vector plots see figures 4–6 of ref. 38). The relatively small roll angles, except at the AGA step in both I222 and $P2_12_12_1$ forms, are consistent with this helix bend at the AGA step.

A-Like Packing Interactions. In addition to the interhelix intercalation unique to this structure (Fig. 6), the familiar A-like "base pair into minor groove" packing is seen at the other end of the helix (Fig. 7). Watson–Crick hydrogen bonding occurs between C(11)-O2 and G(10)-N2, and the C(11)-O2 is also involved in an intermolecular hydrogen bond with G(5)*-N2. N2 of G(10) is involved in an intermolecular hydrogen bond with N3 of G(5)* as shown in Fig. 7*c*. The O2′ of G(10) makes two intermolecular hydrogen bonds with O2′ and N3 of A(4)*, with bond lengths 2.54 and 2.70 Å, respectively. This provides quadruple, G5*·(G10·C11)·A4*, interactions that reinforce stability in this region (Fig. 7*c*).

Intercalated Helix Conformation. The doubling of the rise at the intercalation region is similar to that which occurs when drugs or small molecules intercalate between base pairs (39–45). However, the helix unwinding at the intercalation step in our structure is very different. Our intercalator is not a free aromatic ring but another DNA base pair connected to its backbone chains. To prevent steric clash between backbones of the two helices at the intercalation site, helical twist must be unwound to 16.1° (Table 2). This is unnecessary, and not observed, in a typical drug CA base intercalation step (Table 2).

Backbone torsion angles at the intercalated site are similar to

BIOCHEMISTRY

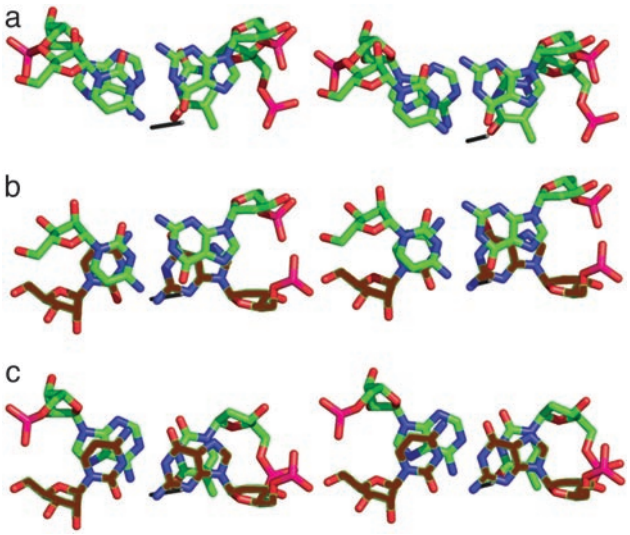

Fig. 8. Local base-pair stacking without (*a*) and with (*b* and *c*) rC1·dG20 intercalation. (*a*) CA step: base pairs rC1·G20 (front) and rA2·dT19 (behind). (*b*) CC* step: base pairs rC1·G20 (front) and 1*·20* (behind; brown). (*c*) C*A step: base pairs rC1*·dG20* (front; brown) and rA2·dT19 (behind). Note that base-pair overlap is enhanced by intercalation.

those found in normal A-DNA or A-RNA. Torsion angles of α(331.7°), ζ(280.5°), and ε(236.7°) are similar to that of A-RNA (313°, 281°, and 213°, respectively; ref. 46). However, the base-sugar torsion angle, χ(205.1°), and main chain, β(193.1°), are quite different from those of A-RNA (13° and 175°, respectively). These values for the conformation angles at the intercalation site agree with the Berman model in which intercalation geometry is A-like except for χ and β (46).

Intercalation does not seem to favor any particular sugar conformation, because several sugar conformations have been found in intercalated structures (42–44). In our case, all sugar puckers are C3′-endo except for residue 2 on the RNA strand. Both C3′-endo and C2′-endo conformations were tried at this locus, and although the difference density was weak, the C2′-endo conformation was favored. This sugar switch at residue 2 is also seen in the high resolution $P2_12_12_1$ structure (ref. 25 and M.L.K., L. Lavelle, G.W.H., H.-L. Ng, and R.E.D., unpublished data).

Deformability of the Initial Pyrimidine-Purine CA Step Rather Than the Final Purine-Purine AG Step. Intercalation occurred only in the very flexible pyrimidine-purine CA step but not in the purine-purine AG step at the other end of the helix. Pyrimidine-purine steps are known to be weak and to exhibit relatively little ring–ring overlap between adjacent base pairs (47). They are especially susceptible to large twist and slide deformations and to bending via a positive roll. In contrast, purine-purine steps are stabilized by extensive ring–ring overlap from one base pair to the next, which suggests that CA steps should be more prone to intercalation than AG steps. Intercalation replaces the weaker stacking by a CA step (base pairs 1·20 and 2·19) (Fig. 8*a*), with a CC* step (base pairs 1·20 and 1*·20*) (Fig. 8*b*), followed by another C*A step (base pairs 1*·20* and 2·19) (Fig. 8*c*).

Role of Helix Intercalation. By base swapping between neighboring duplexes, the unfavorable amine-amine (N4-N6) stacking in the C(1)-A(2) step is replaced with favorable amine-carbonyl (N4-O2*) and carbonyl-amine (O2*-N6) stacking in C(1)-C(1)* and C(1)*-A(2) steps (Fig. 8 *b* and *c*). Also, the O5′ and N4 and N1 of C(1)* are involved in hydrogen bonding with the O1P, N3, and N7 of A(2).

The DNA strand shows a favorable situation as well. The O4-O6, carbonyl-carbonyl stacking in the T(19)-G(20) step is replaced by favorable O4-N2*, carbonyl-amine and amine-carbonyl stacking in the T(19)-G(20)*, and G(20)*-G(20). Thus, favorable O4-N2-O6, carbonyl-amine-carbonyl stacking in T(19)-G(20)*-G(20) occurs by G* intercalation.

Hence, the base-swapped dimer is stabilized both by van der Waals interactions as well as hydrogen bonds at the intercalated region between two molecules.

Comparison of Nucleic Acid Intercalated Structures. As shown in Table 3, our intercalation has similarities as well as striking differences from I-DNA and the zipper-like DNA duplex. All involve intercalation at the CA steps. In the zipper-like DNA duplex, bases are not paired. In I-DNA, although the bases are paired, they are paired by non-Watson–Crick hydrogen bonds; however, ours do show Watson–Crick base-pairing. Our structure resembles 3D domain-swapping structures found in proteins. The identical C·G base pairs were exchanged to form a dimer in the same way that identical domains exchange to form an oligomer, which has been proposed as a mechanism that explains the evolution from monomeric to oligomeric proteins (48).

End-to-End Stacking in RNA·DNA Hybrid. Our base-pair swapping occurs at the end-to-end interaction of the RNA·DNA hybrid. End-to-end packing is uncommon in RNA·DNA structures. In the dominant "A-helix packing" (Table 4 and Fig. 7*a*), neighboring helices pack against the walls of the minor groove. However, end-to-end stacking has been seen in r(UUCGGGCGCC)·d(GGCGCCCGAA) (52). In this structure, residue 1 stacks on top of residue 20* (head to tail), simulating a continuous helix for two molecules, as in a B-helix. In contrast, residue 1 stacks on top of residue 1* in our molecule. In other words, residues 1 and 1* are packed by twofold symmetry (the twofold center is at the center of 1·20 and 1*·20* residues). The large rise (6.1 Å) and unwinding of the helical twist (16.1°) permit its intercalation.

Summary

The RNA·DNA hybrid polypurine tract sequence can crystallize in two space groups: $P2_12_12_1$, with only A-like packing, and I222, with A-like packing at one end and a previously uncharacterized CA intercalation step at the other. I222 diffracts to only 2.38-Å resolution, whereas the $P2_12_12_1$ crystal diffracts to 1.15-Å resolution. A-helix-type packing is stabilized by more hydrogen bonds (Fig. 8*c*) than end-to-end stacking.

Pyrimidine-purine CA tracts of DNA are involved in recombination, transcription gene conversion, and insertion events and in

Table 3. Comparison between nucleic acid intercalated structures

	CA step involved?	Strands	Base pairing	Unusual features	Ref.
Helix intercalation	Yes	Antiparallel	Watson–Crick H bond	Two strands	This work
I-DNA	Yes	Parallel	C·C or A·A	Four strands	13–17 and 20
Zipper-like DNA	Yes	Antiparallel	Non-base pairing	Induced by sheared G·A mismatches	22

Table 4. Single crystal structures for RNA·DNA hybrid duplexes

Sequence	Structure	Space group	Crystal packing	Ref.
r(CAAAGAAAAG)·d(CTTTTCTTTG)	Intercalation RNA-C2′ endo for 2A, C3′ endo for others DNA-C3′ endo	I222	Abutting and head-to-head (1 and 1* and 20 and 20*)	This work
r(CAAAGAAAAG)·d(CTTTTCTTTG)	A-like	$P2_12_12_1$	Abutting	25 and M.L.K., L. Lavelle, G.W.H., H.-L. Ng, and R.E.D., unpublished data
r(GAAGAAGAG)·d(CTCTTCTTC)	A-like	$P6_1$	Abutting	49
r(GAAGAGAGAG)·d(CTCCTCTTC)	A-like flipped-out base	I222	Abutting and head-to-head, looped-out, bulge interaction	31
r(GAAGAGAAGC)·d(GCTTCTCTTC)	A-like	$P2_12_12_1$	Abutting	50
r(GAAGAAGAA)·d(TTCTTBr^{5}CTTC)	A-like	$P6_1$	Abutting	51
r(UUCGGGCGCC)·d(GGCGCCCGAA)	RNA-C3′ endo DNA-C2′ endo for AA, C3′ endo for others	$P4_322$	Abutting and head-to-tail (1 and 20*, and 20 and 1*)	52

the enhancement of recombination (53, 54). In the crystal structure of d(ACCGGCGCCACA)·d(TGTGGCGCCGGT), Watson–Crick base pairing in the major groove of $(CA)_n$ tracts are disrupted by tilting of the bases, and hydrogen bonds were shifted to the 5′ neighbor of their complement (53).

This CA step is responsible for highly variable stacking patterns such as the I-DNA motif (13–17, 20), the zipper-like structure (22), and base-pair shifting to a tilted stack that disrupts Watson–Crick base pairing (53). We now add base-pair swapping involving a CA step. Investigation of its potential role in biology deserves further study.

We thank Thang Kien Chiu for the SHELXDNA program for plotting helical parameters and Yong Xiong for the discussion about the sugar parameters for nucleic acids in CNS refinement. We also thank Ho-Leung Ng and Gail Katsir for helpful discussion and comments in preparing this manuscript and Vivian Yip and Wai-Yin Ng for help with purification of the DNA strand. This work was supported by National Institutes of Health Grant GM-31299 and an Amgen Fellowship (to G.W.H.) from the University of California (Los Angeles) Center for AIDS Research.

1. Ng, H. L., Kopka, M. L. & Dickerson, R. E. (2000) *Proc. Natl. Acad. Sci. USA* **92,** 10767–10771.
2. Vargason, J. M., Henderson, K. & Ho, P. S. (2001) *Proc. Natl. Acad. Sci. USA* **98,** 7265–7270.
3. Lee, J. R., Woodsworth, M. L., Latimer, L. J. & Morgan, A. R. (1984) *Nucleic Acids Res.* **12,** 6603–6614.
4. Parkinson, G. N., Lee, M. P. H. & Neidle, S. (2002) *Nature* **417,** 876–879.
5. Williamson, J. R., Raghuraman, M. K. & Cech, T. R. (1989) *Cell* **59,** 871–880.
6. Sundquist, W. I. & Klug, A. (1989) *Nature* **342,** 825–829.
7. Kim, J., Cheong, C. & Moore, P. B. (1991) *Nature* **351,** 331–332.
8. Deng, J., Xiong, Y. & Sundaralingam, M. (2001) *Proc. Natl. Acad. Sci. USA* **98,** 13665–13670.
9. Lerman, L. S. (1961) *J. Mol. Biol.* **3,** 18–30.
10. Wilson, W. D. & Jones, R. L. (1981) *Adv. Pharmacol. Chemother.* **18,** 177–222.
11. Kielkopf, C. L., Erkkila, K. E., Hudson, B. P., Barton, J. K. & Rees, D. C. (2000) *Nat. Struct. Biol.* **7,** 117–121.
12. Bennett, M. J., Choe, S. & Eisenberg, D. S. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 3127–3131.
13. Leroy, J.-L., Guéron, M., Mergny, J.-L. & Helene, C. (1994) *Nucleic Acids Res.* **22,** 1600–1606.
14. Rohozinski, J., Hancock, J. M. & Keniry, M. A. (1994) *Nucleic Acids Res.* **22,** 4653–4659.
15. Chen, L., Ci, L., Zhang, X. & Rich, A. (1994) *Biochemistry* **33,** 13540–13546.
16. Kang, C., Berger, I., Lockshin, C., Ratliff, R., Moyzis, R. & Rich, A. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 11636–11640.
17. Kang, C., Berger, I., Lockshin, C., Ratliff, R., Moyzis, R. & Rich, A. (1995) *Proc. Natl. Acad. Sci. USA* **92,** 3874–3878.
18. Berger, I., Kang, C., Fredian, A., Ratliff, R., Moyzis, R. & Rich, A. (1995) *Nat. Struct. Biol.* **2,** 416–425.
19. Gehring, K., Leroy, J.-L. & Gueron, M. (1993) *Nature* **363,** 561–564.
20. Guéron, M. & Leroy, J.-L. (2000) *Curr. Opin. Struct. Biol.* **10,** 326–331.
21. Viswamitra, M. A. & Pandit, J. (1983) *J. Biomol. Struct. Dyn.* **1,** 743–753.
22. Shepard, W., Cruse, W. B. T., Fourme, R., delaFortelle, E. & Prange, T. (1998) *Structure (London)* **6,** 849–861.
23. Spackova, N., Berger, I. & Sponer, J. (2000) *J. Am. Chem. Soc.* **122,** 7564–7572.
24. Quigley, G. J. & Rich, A. (1976) *Science* **194,** 796–806.
25. Han, G. W. (2001) *Acta Crystallogr. D* **57,** 213–218.
26. Han, G. W., Kopka, M. L., Cascio, D., Grzeskowiak, K. & Dickerson, R. E. (1997) *J. Mol. Biol.* **269,** 811–826.
27. Otwinowski, Z. & Minor, W. (1997) in *Methods in Enzymology: Macromolecular Crystallography, Part A.*, eds. Carter, C. W., Jr., & Sweet, R. M. (Academic, New York, NY), Vol. 276, pp. 307–326.
28. Kissinger, C. R., Gehlhaar, D. K. & Fogel, D. B. (1999) *Acta Crystallogr. D* **55,** 484–491.
29. Brünger, A. T. (1992) X-PLOR, A System for X-Ray Crystallography and NMR (Yale Univ. Press, New Haven, CT), Version 3.1.
30. Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., *et al.* (1998) *Acta Crystallogr. D.* **54,** 905–921.
31. Sudarsanakumar, C., Xiong, Y. & Sundaralingam, M. (2000) *J. Mol. Biol.* **299,** 103–112.
32. Adams, P. D., Pannu, N. S., Read, R. J. & Brünger, A. T. (1997) *Proc. Natl. Acad. Sci. USA* **94,** 5018–5023.
33. Lavery, R. & Sklenar, H. (1988) *J. Biomol. Struct. Dyn.* **6,** 63–91.
34. Sarafianos, S. G., Das, K., Tantillo, C., Clark, Jr., A. D., Ding, J., Whitcomb, J. M., Boyer, P. L., Hughes, S. H. & Arnold, E. (2001) *EMBO J.* **20,** 1449–1461.
35. Dickerson, R. E. (1992) *Methods Enzymol.* **211,** 67–111.
36. Dock-Bregeon, A. C., Chevrier, B., Podjarny, A., Johnson, J., de Bear, J. S., Gough, G. R., Gilham, P. T. & Moras, D. (1989) *J. Mol. Biol.* **209,** 459–474.
37. Smith, C. K., Brannigan, J. A. & Moore, M. H. (1996) *J. Mol. Biol.* **263,** 237–258.
38. Dickerson, R. E., Goodsell, D. S. & Kopka, M. L. (1996) *J. Mol. Biol.* **256,** 108–125.
39. Saenger, W. (1984) in *Principles of Nucleic Acid Structure*, ed. Cantor, C. R. (Springer, New York), pp. 350–367.
40. Sobell, H. M., Tsai, C. C., Jain, S. C. & Gilbert, S. G. (1977) *J. Mol. Biol.* **114,** 333–365.
41. Wang, A. H.-J., Nathans, J., van der Marel, G., van Boom J. H. & Rich, A. (1978) *Nature* **276,** 471–474.
42. Wang, A. H.-J., Ughetto, G., Quigley, G. L. & Rich, A. (1986) *J. Biomol. Struct. Dyn.* **4,** 319–342.
43. Wang, A. H.-J., Ughetto, G., Quigley, G. J. & Rich, A. (1987) *Biochemistry* **26,** 1152–1163.
44. Kamitori, S. & Takusagawa, F. (1992) *J. Mol. Biol.* **225,** 445–456.
45. Gasper, S. M., Armitage, B., Shui, X., Hu, G. G., Yu, C., Schuster, G. B. & Williams, L. D. (1998) *J. Am. Chem. Soc.* **120,** 12402–12409.
46. Berman, H. M., Neidle, S. & Stodola, R. K. (1978) *Proc. Natl. Acad. Sci. USA* **75,** 828–832.
47. Dickerson, R. E. (1999) in *Oxford Handbook of Nucleic Acid Structure*, ed. Neidle, S. (Oxford Univ. Press, New York), pp. 145–191.
48. Liu, Y., Hart, P. J., Schlunegger, M. P. & Eisenberg, D. S. (1998) *Proc. Natl. Acad. Sci. USA* **95,** 3437–3442.
49. Xiong, Y. & Sundaralingam, M. (2000) *Nucleic Acids Res.* **28,** 2171–2176.
50. Conn, G., Brown, T. & Leonard, G. (1999) *Nucleic Acids Res.* **27,** 555–561.
51. Xiong, Y. & Sundaralingam, M. (1998) *Structure (London)* **6,** 1493–1501.
52. Horton, N. C. & Finzel, B. C. (1996) *J. Mol. Biol.* **264,** 521–533.
53. Timsit, Y., Vilbois, E. & Moras, D. (1991) *Nature* **354,** 167–170.
54. Treco, D. & Arnheim, N. (1986) *Mol. Cell. Biol.* **6,** 3934–3947.