

Molecular mechanics calculations of the structures of polyamide nucleic acid DNA duplexes and triple helical hybrids

(D-loop)

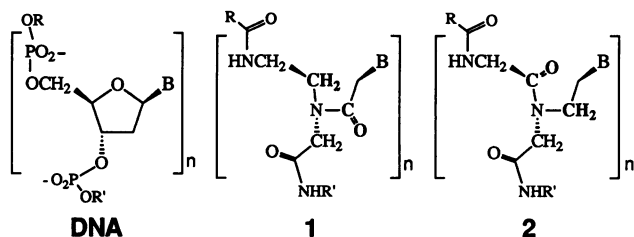
ÖRN ALMARSSON*, THOMAS C. BRUCE*, JANICE KERR†, AND RONALD N. ZUCKERMANN†

*Department of Chemistry, University of California at Santa Barbara, Santa Barbara, CA 93106; and †Chiron Corp., 4560 Horton Street, Emeryville, CA 94608-2916

Contributed by Thomas C. Bruce, April 20, 1993

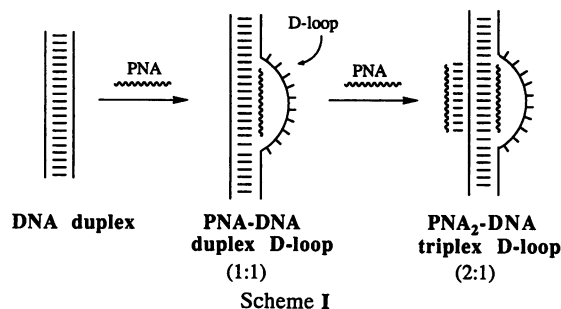
ABSTRACT Polyamide nucleic acids (PNAs) have emerged as useful agents for recognition of single- and double-stranded nucleic acids. Interresidue hydrogen bonds between the amide carbonyl nearest the nucleobase and chain NH moieties provide inherent stability to the helical conformation of PNA 1. Moving the amide carbonyl away from the nucleobase to the backbone, and replacing it with a methylene group, results in 2 lacking the stabilizing hydrogen bond. Oligomers of 2 do not interact with DNA. Modeling suggests that 2 displays a more extended conformation than 1, and nucleobase orientation is disrupted in 2 in the absence of a complementary DNA strand. This is in contrast to 1, which retains a centrosymmetric arrangement of nucleobases. Structures for 1-T₁₀-DNA and (1-T₁₀)₂-DNA species spanned by a pyrimidine strand (D-loop) were constructed. In the triple helical (1-T₁₀)₂-DNA structure, the two PNA strands form the complementary Watson-Crick paired strand and the Hoogsteen base-paired strand in the major groove of the 1-DNA duplex. The PNA strands are proposed to bind antiparallel to one another in (1-T₁₀)₂-DNA structure. The factors suggested to account for the stability of this 2:1 complex are (i) a hydrophobic attraction between two PNA backbones and (ii) a favorable electrostatic effect resulting from replacement of a phosphodiester backbone by a neutral peptide backbone.

The term polyamide nucleic acids (PNAs) has been applied to sequences of *N*-(2-aminoethyl)glycine units in peptide linkage with the glycine nitrogens connected to purine or pyrimidine bases by an acetate linker (Structure 1) (1, 2). Explo-



ration of their capabilities as antigene and antisense agents for DNA (3-5) and RNA (5) has been initiated. (For other examples of PNA surrogates, see ref. 6 and references therein.) Until now, the base used in most studies has been thymine, with occasional cytosines introduced to give mismatches with a purine complementary strand (2). Oligomers of thymine-substituted 1 and complementary adenosine single-stranded DNA react in a 2:1 ratio to provide (1)₂-DNA (3). The PNA is only complementary to the DNA sequence and thus it is required that the intermediate 1-DNA structure precede formation of the stable (1)₂-DNA. Short sequences of 1 react with longer double-stranded (ds) DNA sequences to

displace a portion of one DNA strand, yielding a displacement loop (D-loop; Scheme I).



The phosphates in the loop are highly sensitive to S1 nuclease and the exposed thymines in the loop are sensitive to oxidizing agents (4). The second PNA of the 2:1 complex of Scheme I was proposed to bind in the major groove region of the 1-DNA hybrid by Hoogsteen base-pairing to produce a triple-helix structure (3, 4).

To understand the propensity of 1 to interact with DNA, we have carried out a computer modeling study of the structures of ds(1-DNA) and (1)₂-DNA. Important to our conclusions is the experimental observation that interchanging the positions of —CH₂— and side chain —CO— groups of 1 to provide 2 completely abolishes any interactions with DNA.

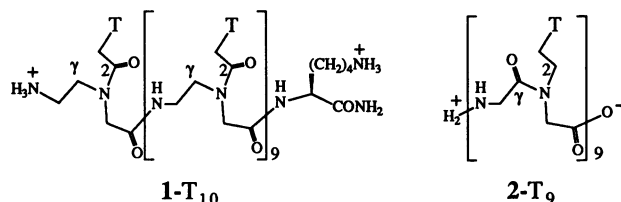
MATERIALS AND METHODS

Molecular Modeling. All computations were performed on a Silicon Graphics (Mountain View, CA) 4D/320GTX workstation. QUANTA (7) v. 3.3 SEQUENCE BUILDER was used to generate the two complementary decamers of single-stranded DNA, as prescribed in the DNAH.RTF topology file for CHARMM v. 22 (8). CHARMM interfaced with QUANTA was used to incorporate the DNA decamers into B-DNA structures. QUANTA's sequence builder (v. 92.0305) generates single strands of B-DNA with averaged torsion angles for ribose rings in the C2'-endo conformation (9). Hybrid molecules with 1 and 2 were built in the following manner. The ribose phosphate chain was peeled away from one strand of constructed dsDNA decamer, leaving the bases all in place. Next, the backbone of 1 or 2 was added. The backbone construction was carried out as follows. For 1, the linker —CH₂—C(=O)— units were overlaid on the positions formerly occupied by the C1' and C2' carbons of the ribose. The PNA glycine C^α and —C(=O)— were superimposed on the

Abbreviations: PNA, polyamide nucleic acid; 1-T₁₀, decameric PNA 1 in which all bases are T; (dAp)₁₀, decanucleotide of adenosine 5'-phosphate; (dTp)₁₀, decanucleotide of thymidine 5'-phosphate; dsDNA, double-stranded DNA; ABNR, adopted basis Newton-Raphson.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

oxygen attached to ribose C3' and the phosphorus, respectively, while the $-\text{CH}_2\text{CH}_2\text{NH}-$ units were superimposed on C4', C5', and O5'. *Trans*-amides were modeled to conform with the known preference for the *trans* $-\text{CONH}-$ linkage (10). The C termini were capped with lysine amide, in accord with the published PNA structures (1, 2), to give structure 1-T₁₀. The strands of 2-T₉ were assembled in a like fashion.



A CHARMM topology file (RTF) was written for the thymine 1 and 2 units by consideration of charges and atom types in DNA (as in DNAH.RTF) and glycine (as in AMINOH.RTF). Charged phosphate diesters were minimized with sodium counterions (charge 1⁺). A dielectric constant of 1 was used throughout. Duplex decamers of complementary 1 or 2 and DNA strands were minimized in CHARMM using 100 steps of steepest descents algorithm (force criterion, 0.001 kcal per 10 steps; 1 cal = 4.184 J), followed by adopted basis Newton-Raphson (ABNR) algorithm for 5000 steps (gradient tolerance, 1×10^{-9} kcal per 10 steps). Hydrogen bonds and nonbonding interactions were cut off at 5.0 and 14.0 Å, respectively, and the angle cutoff for hydrogen bonding was set at 90°. For the ds[1-T₁₀·(dAp)₁₀] and ds[2-T₉·(dAp)₁₀] structures [(dAp)₁₀, decanucleotide of adenosine 5' phosphate], the coordinates of the complementary DNA strand as well as the bases on the polyamide strands were fixed.

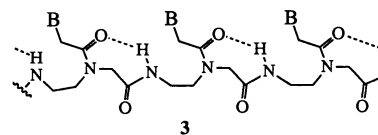
The minimized 1-T₁₀·(dAp)₁₀ decamer structure served as initial coordinates for building the D-loop structure (Scheme I). The positions of bases in the PNA·DNA were overlapped with like bases in a B-DNA ds(GCGT₁₀GCG)/(CGCA₁₀CGC) hexadecamer built in QUANTA's sequence builder. A sequence of five repeating units of the sodium thymidylyl (5' → 3')-thymidylate (5')-hydrate dinucleotide was built from the x-ray coordinates obtained from the Cambridge Structural Database (11) (Refcode THYTHY10). The coordinates were imported from THYTHY10.dat into QUANTA 3.3 and manipulated by CHARMM to give the proposed helix structure with all thymine bases turned away from the helical axis. The resulting (dTp)₁₀ segment (decanucleotide of thymidine 5'-phosphate) was properly patched into the 16-mer dsDNA structure and minimized slightly. The PNA strands 1-T₁₀ were then sequentially introduced. After pairing the exposed adenines with one PNA to give the intermediate 1-T₁₀·DNA duplex D-loop structure (Scheme I), the complete assembly was minimized with ABNR fixing the bases on both strands. The minimized structure was fitted with the second 1-T₁₀ in the same conformation as in the duplex. The Hoogsteen base-paired strand of 1-T₁₀ could only fit in the major groove of the 1-T₁₀·DNA hybrid when placed antiparallel to the first strand of 1-T₁₀. Distance constraints between hydrogen bonding atoms N⁷(A) and H³(1-T) and between NH₂(A) and O⁴(1-T) were applied to properly arrange the hydrogen bonds in the Hoogsteen base pairs. The newly published structure for the subunit of the d(T)_n·d(A)_n·d(T)_n triple helix (12) was used to assist in the building of the (1-T₁₀)₂-DNA structure. ABNR minimization provided the final product of Scheme I.

Synthesis. 2-T₉ was synthesized on a solid support by a three-step cycle. In the first step, bromoacetic acid was coupled to Rink amide polystyrene resin, followed by nucleophilic displacement of the bromine with 1-(2-aminoethyl)thymine (13). An *N*^α[2-(3,5-dimethoxyphenyl)propyl-2-oxycarbonyl]-protected glycine was then coupled to the resin-bound secondary amine with standard benzotriazol-1-yloxytris(pyrrolidino)phos-

phonium hexafluorophosphate chemistry. The *N*^α-protecting group was then removed with 5% trifluoroacetic acid/CH₂Cl₂ to generate the free amine. This process was repeated nine times to generate the full-length product. Alkylation of bis(trimethylsilyl)thymine (14) with bromoacetonitrile, followed by H₂-Pd/C reduction, gave the starting 1-(2-aminoethyl)thymine monomer in 72% overall yield and >90% purity by ¹H NMR. The nonamer was cleaved from the resin with 95% trifluoroacetic acid/H₂O. The final product was obtained in 30% yield and >90% purity after HPLC purification and was characterized by mass spectroscopy. The *T*_m measurements of binding of 2-T₉ to (dAp)₁₀ were performed as described (1) but failed to show any binding.

RESULTS AND DISCUSSION

Inspection of the decamer B-[1-T₁₀·(dAp)₁₀] duplex (Fig. 1) shows that the structure of the 1-T₁₀ strand is predefined by a strong hydrogen bond between the side chain carbonyl oxygen (α to the methylene carrying the base) and the backbone amide NH in the next residue (3). This crucial repeating hydrogen-bonding motif has not previously been recognized as the required stabilizing influence for helical strands of 1. The PNA strand 1-T₁₀ can be removed from the



duplex decamer and subjected to energy minimization without loss of the specific hydrogen bonds between residues. Base orientations in 1 remain centrosymmetric even after extensive minimization in CHARMM (Fig. 2a). Thus, this helical PNA strand possesses an inherent stability.

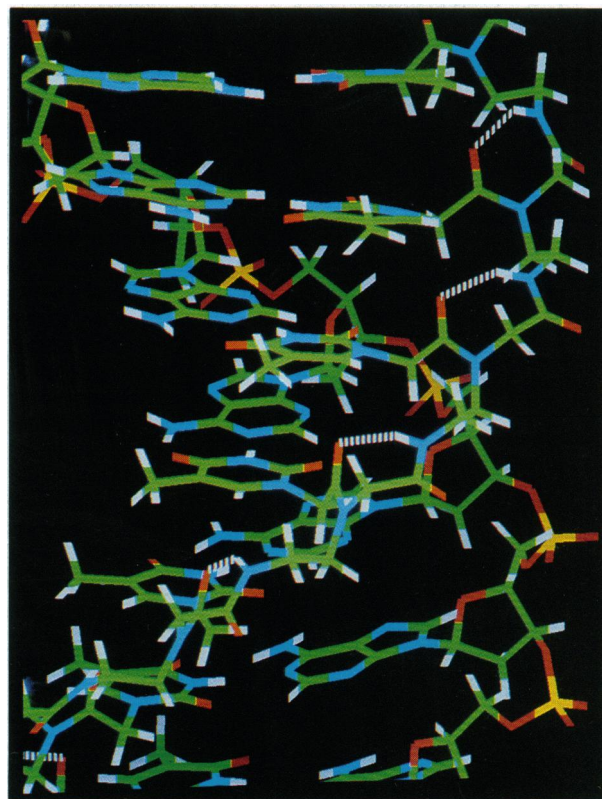


FIG. 1. View of structure of (1-T₁₀)₂·(dAp)₁₀ hybrid duplex showing hydrogen bonding between residues of 1 as shown by 3.

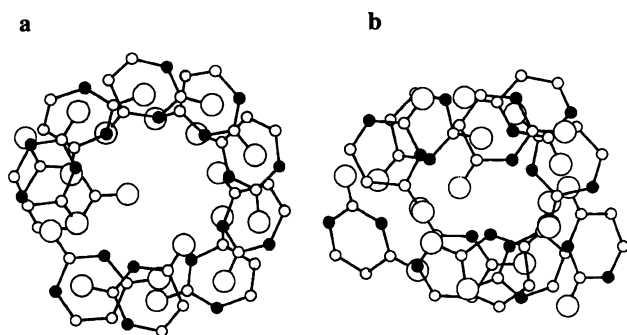
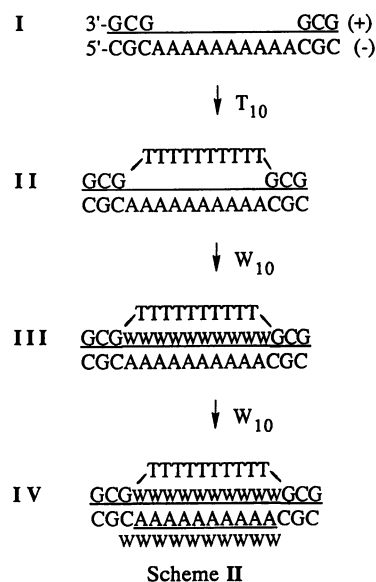


FIG. 2. Views along the helical axes of strands showing only thymine bases. 1- T_{10} strand (a) and 2- T_9 strand (b), with nitrogen atoms shown as solid circles; oxygens are larger than nitrogens and carbons. Both structures were minimized in the absence of complementary DNA strands. Base alignment in 2 is disrupted and duplex formation requires reorganization.

In contrast, the polyamide strand of ds(2- T_9)DNA decamer lacks the intramolecular hydrogen bonding of 1 (shown in 3) required to stabilize the helical backbone conformation. When the helical strand of 2 is minimized in the absence of the complementary DNA strand, it adopts a more extended conformation than can be accommodated in ds(2- T_9)DNA. This is not necessarily the preferred conformation but it shows the helical conformation not to be favored. Also, base orientations are disrupted from the centrosymmetrical arrangement in the helix (Fig. 2b). Therefore, expenditure of energy is required to maintain the helicity and base alignments of 2 required for double helix formation. In addition, comparison of structures 1 and 2 reveals that 1 has only three sp^2 atoms in the backbone of each monomer while 2 has four. The increased rigidity in the backbone of 2 compared with 1 disfavors a helical structure for 2. Although the 2 are polyglycine structures, the alternating N substitution prevents internal hydrogen bonding and hampers helix formation.

For useful comparison, a model for 2 with 10 nucleotides and a lysine amide on the C terminus was constructed as in 1- T_{10} . Table 1 lists the individual contributions to the CHARMM potential energy for the single-stranded 1- T_{10} and 2- T_{10} , after separation from the (dAp) $_{10}$ DNA strand, as well as their hybrid duplexes with (dAp) $_{10}$. The two structures have favorable van der Waals and electrostatic energies, and the bond and improper dihedral terms from the force field are comparable. The angle and dihedral terms indicate a destabilizing effect of angle and dihedral angle deformations in the helical arrangement of 2- T_{10} relative to 1- T_{10} . The stabilizing hydrogen-bonding term is also larger in the case of 1- T_{10} compared with 2- T_{10} , which only displays hydrogen bonding of lysine and N-terminal $-NH_3^+$ groups to the nearest bases.

Description of the method for generation of the D-loop structure of Scheme I is given in *Materials and Methods*. The end-to-end distance (5'-P to 3'-OH) of the helical D-loop (11) was 40 Å, which fits the gap created by removing 10 nucleoside units from one strand of the B-DNA structure. Incorporation of the T_{10} loop structure into I (Scheme II) com-



pletes the modeling of the (+)-strand. In II there is no pairing of A and T bases due to the conformation of the (dTp) $_{10}$ D-loop. The first strand of 1- T_{10} (abbreviated as W_{10} in Scheme II) was introduced by its pairing to the (dAp) $_{10}$ sequence on the (-)-strand. Minimization of the complete assembly was then performed. The resulting model for the intermediate (1- T_{10})·DNA 1:1 complex with (dTp) $_{10}$ loop is shown in Fig. 3. In the exposed (dTp) $_{10}$ loop structure, a few of the thymines are turned toward the purine (-)-strand due to the electrostatic interactions with the phosphates of the latter. Small openings between the duplex and the loop are detected, but the overall structure can be considered quite compact. The terminal $-NH_3^+$ group of the C-terminal lysine on the PNA interacts with two phosphates simultaneously [C2 and G3 on the (+)-strand; see Scheme II], and the protonated ethylamine on the N terminus interacts with the phosphate group on C3 on the (-)-strand.

The second 1- T_{10} strand is accommodated with ease by III to yield IV (Fig. 4 and Scheme II). Few nonbonded interactions are observed between the (dTp) $_{10}$ -loop and the other strands, but thymines close to the junctions of the D-loop and duplex regions hydrogen bond with phosphate groups on the purine strand. Thymine in residue 6, centrally located in the D-loop, hydrogen bonds with an adenine $-NH_2$. While it is difficult to make quantitative predictions based on molecular mechanics calculations as to the relative stabilities of dsDNA

Table 1. Total CHARMM energies and individual energy term contributions for minimized decamer duplexes 1- T_{10} ·DNA and 2- T_{10} ·DNA and separated single-stranded PNA constituents

Structure	Energy							
	Total*	Bond	Angle	Dihedral	Impr	VDW	Elec	H bond
1- T_{10} ·(dAp) $_{10}$	-4037.6	32.6	261.3	223.0	9.6	-146.4	-4328.1	-89.7
2- T_{10} ·(dAp) $_{10}$	-3693.5	29.8	279.8	280.8	12.0	-142.5	-4087.6	-65.8
1- T_{10} †	-1472.9	15.6	88.4	47.5	4.0	-72.5	-1530.5	-25.4
2- T_{10} ‡	-1288.2	13.5	110.3	77.2	6.7	-69.1	-1413.8	-13.1

For definitions of energy terms composing the total energy in CHARMM, see ref. 8. See text for structure of 1- T_{10} ·2-DNA hybrid is described in *Results*. Impr, improper dihedral; VDW, van der Waals; Elec, electrostatic.

*Total CHARMM energy, sum of all contributions in the table.

†Structure minimized in the absence of complementary (dAp) $_{10}$.

‡Structure is analogous to that of 1- T_{10} apart from the required interchange of the γ -methylene and 2-carbonyl groups.

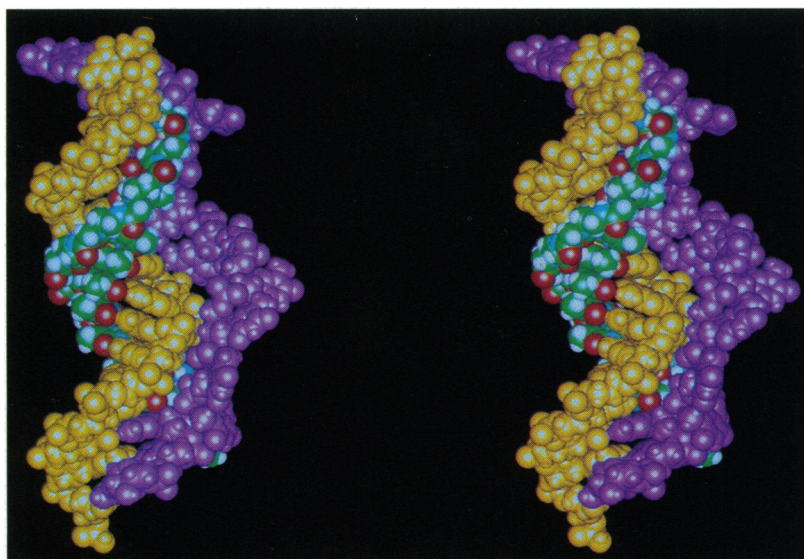


FIG. 3. Stereoview of the 16-mer dsDNA duplex D-loop structure after invasion by 1-T₁₀ (intermediate in Scheme I). Complementary DNA strand (CGCA₁₀CGC; yellow) is in B-DNA conformation displaying C2'-*endo* puckering of ribose rings. Pyrimidine strand (GCGT₁₀GCG; purple) contains the (dTp)₁₀ D-loop structure. The strand of 1-T₁₀, Corey-Pauling-Koltun (CPK) colors with carbons green, is Watson-Crick base-paired with the yellow purine strand.

and 1·DNA on the one hand, and triple-stranded DNA and (1)₂·DNA on the other hand, one can consider individual contributions from the CHARMM force field. The increasingly negative van der Waals and electrostatic contributions for structures III and IV suggest two factors that favor the formation of (1)₂·DNA over other possible structures of 1 and DNA: (i) a van der Waals attraction between the relatively nonpolar peptide strands, and (ii) a stabilizing electrostatic effect arising from removal of one phosphodiester strand. The latter is supported by the observed dependence of the formation of (1)₂·DNA complexes on salt concentration. At high salt concentration, (1)₂·DNA formation is hampered, but once formed the complex is unaffected by increased salt concentration (4). Furthermore, solvation of the phosphates and bases on the displaced loop is expected to be stabilizing according to analysis of x-ray structures of nucleotides (15). The propensity for the 1·DNA 1:1 complex to form the 2:1 adduct with another strand of 1 is not surprising in view of the

fact that poly(dTp)·(dAp) is known to disproportionate to form the triple helix (dTp)_n·(dAp)_n·(dTp)_n and single-stranded (dAp)_n in solution when the salt concentration is raised (16). As constructed, the second PNA strand lies antiparallel with the Watson-Crick paired PNA. The antiparallel arrangement is required to give correct Hoogsteen base pairing between corresponding H³(1-T) and N⁷(A) and between O⁴(1-T) and NH₂(A) when the second strand of 1-T₁₀ is fit into the major groove of the 1:1 PNA·DNA hybrid in the same helical conformation as the first Watson-Crick paired strand of 1-T₁₀. Further modeling work is required to elucidate the importance of strand orientations in the triplexes as well as other factors responsible for the exquisite hybridization/strand invasion properties of PNA.

Gratitude is extended to Charles K. Marlowe for assistance with synthesis. T.C.B. thanks the Office of Naval Research and the National Science Foundation for support.

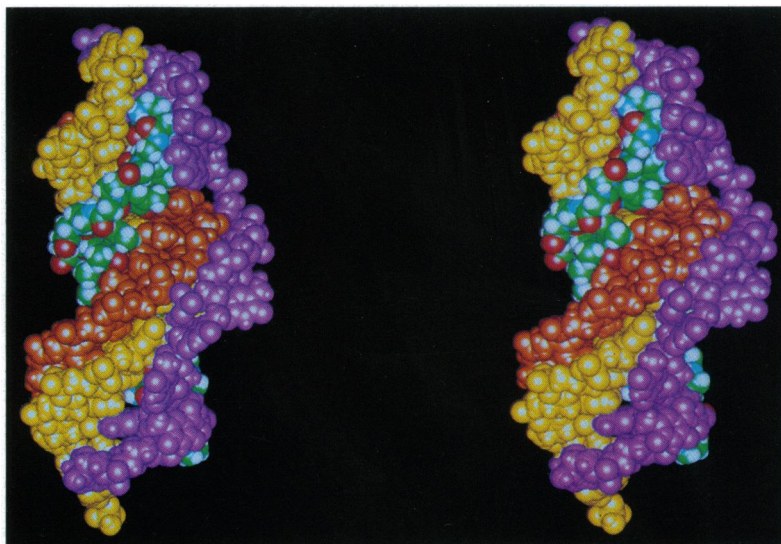


FIG. 4. Stereoview of the 16-mer dsDNA duplex D-loop structure after binding of a second 1-T₁₀ strand (product in Scheme I). See Fig. 3 legend for atom coloring scheme of the duplex. The second strand of 1-T₁₀ (orange) is situated in the major groove of the 1·DNA hybrid duplex, bound by Hoogsteen base-pairing to the purines. The orange PNA strand is antiparallel to the CPK colored PNA strand in the duplex (Fig. 3).

1. Nielsen, P. E., Egholm, M., Berg, R. H. & Buchardt, O. (1991) *Science* **254**, 1497–1500.
2. Egholm, M., Buchardt, O., Nielsen, P. E. & Berg, R. H. (1992) *J. Am. Chem. Soc.* **114**, 1897–1898.
3. Egholm, M., Nielsen, P. E., Buchardt, O. & Berg, R. H. (1992) *J. Am. Chem. Soc.* **114**, 9677–9678.
4. Nielsen, P. E., Egholm, M., Berg, R. H. & Buchardt, O. (1993) *Nucleic Acids Res.* **21**, 197–200.
5. Hanvey, J. C., Peffer, N. J., Bisi, J. E., Thomson, S. A., Cadilla, R., Josey, J. A., Ricca, D. J., Hassman, C. F., Bonham, M. A., Au, K. G., Carter, S. G., Bruckenstein, D. A., Boyd, A. L., Noble, S. A. & Babiss, L. E. (1992) *Science* **258**, 1481–1485.
6. Garner, P. & Yoo, J. U. (1993) *Tetrahedron Lett.* **34**, 1275–1278.
7. Molecular Simulations (1993) CHARMM (Version 22 interface) and QUANTA (Version 3.3) (Molecular Simulations, Waltham, MA).
8. Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S. & Karplus, M. (1983) *J. Comput. Chem.* **4**, 187–217.
9. Saenger, W. (1984) *Principles of Nucleic Acid Structure* (Springer, New York), pp. 92–97.
10. Creighton, T. E. (1983) *Proteins: Structure and Molecular Properties* (Freeman, New York), pp. 159–197.
11. Camerman, N., Fawcett, J. K. & Camerman, A. (1976) *J. Mol. Biol.* **107**, 601–621.
12. Raghunathan, G., Miles, H. T. & Sasisekharan, V. (1993) *Biochemistry* **32**, 455–462.
13. Zuckermann, R. N., Kerr, J. M., Kent, S. B. H. & Moos, W. H. (1992) *J. Am. Chem. Soc.* **114**, 10646–10647.
14. Siegel, S. A. & Lin, T. S. (1985) *Biochem. Pharmacol.* **34**, 1121–1124.
15. Schneider, B., Cohen, D. & Berman, H. M. (1992) *Biopolymers* **32**, 725–750.
16. Arnott, S. & Selsing, E. (1976) *J. Mol. Biol.* **98**, 265–269.