Structural Tools for the Analysis
of Protein-Nucleic Acid Complexes
Advances in Life Sciences




# POLYMORPHISM AND POSSIBLE BIOLOGICAL FUNCTIONS OF PARALLEL-STRANDED DNA

Karsten Rippe[1], Vitaly V. Kuryavyi[1],
Eric Westhof[2], and Thomas M. Jovin[1]

[1]Max-Planck-Institut für Biophysikalische Chemie, Abteilung Molekulare Biologie, Postfach 2841, W-3400 Göttingen, F.R.G., and [2]Institut de Biologie Moléculaire et Cellulaire, Centre National de la Recherche Scientifique, 15 rue René Descartes, F-67084 Strasbourg Cedex, France

SUMMARY: DNA oligonucleotides with appropriate sequences can form a duplex in which the two strands are in a parallel orientation, in contrast to the conventional antiparallel helix of B-DNA. Such parallel-stranded duplexes (denoted generically as ps-DNA) are stable under physiological conditions (10 mM $MgCl_2$, neutral pH) and have been constructed to date from A·T and G·C ($ps_{RY}$-DNA) or G·G and A·A base pairs ($ps_{RR}$-DNA). The glycosyl bonds are in a *trans* orientation. In addition, force field calculations indicate that it is possible to construct ps-DNA with other non-canonical base pairs such as A·C and G·T. We present data demonstrating the formation of $ps_{RY}$-DNA and $ps_{RR}$-DNA by gel electrophoresis, UV absorption, vacuum-UV circular dichroism and monomer-excimer fluorescence of oligonucleotides end-labeled with pyrene. The potential intervention of ps-DNA *in vivo* is discussed in terms of specific mechanisms and models. The latter also include a new quadruplex structure with $A^+$·G·$A^+$·G tetrads.

## INTRODUCTION

It is now well established that certain DNA sequences can form parallel-stranded duplexes with d(A·T) base pairing and with both strands in the same 5'→3' orientation (reviewed in Jovin *et al.*, 1990; Jovin, 1991; Rippe and Jovin, 1992). Parallel-stranded DNA

with d(A·T) base pairs has spectroscopic, thermodynamic and biochemical properties that differ significantly from those of conventional antiparallel B-DNA. It is quite stable under physiological conditions and therefore of potential biological significance. Interspersed d(G·C) base pairs can be incorporated in mixed d(A·T) sequences but they tend to destabilize the ps secondary structure (Rippe et al. 1990; Rentzeperis et al., 1992). However, it has been reported that a 40-nt sequence with a G·C content as high as 40% adopts a parallel-stranded (ps) conformation with a melting point of 53 °C in 0.15 M NaCl (Tchurikov et al., 1992).

The glycosyl bonds of the Watson-Crick A·T and G·C base pairs of B-DNA have a *cis* orientation, defined in relation to a central axis parallel to the hydrogen bonds in the two-dimensional projection of the base pairs. In contrast, the orientation of the base pairs in the ps duplexes described below is *trans*, leading to the designation *trans* base pairs for these constituent elements of secondary structure in ps-DNA. It should be emphasized that all the helical structures under consideration incorporate chemically unmodified bases and are stable at neutral pH.

We distinguish between two types of parallel-stranded duplexes according to differences in helical structure. The first, denoted $ps_{RY}$-DNA, has purine·pyrimidine *trans* A·T, G·C, A·C or G·T base pairing with all bases in an *anti* glycosidic conformation. [If the secondary structure is established by A·T base pairing we use the notation $ps_{AT}$-DNA.] The second ps helix type considered here is based on purine·purine base pairing and is designated $ps_{RR}$-DNA. We recently presented experimental support and a model for $ps_{RR}$-DNA based on the alternating d(G-A) sequence (Rippe et al., 1992). The proposed structure incorporates the homopurine base pairs A·A and G·G with alternations between the *anti* and *syn* glycosidic conformations. In this report we present data and models related to ps-DNA structures adopted by the different sequences and discuss their possible biological functions.

## MATERIALS AND METHODS

### OLIGONUCLEOTIDE SYNTHESIS AND CHARACTERIZATION

The oligonucleotides were synthesized, purified, labeled with [$^{32}$P], and analyzed as described previously (Rippe et al., 1989). Oligonucleotides of the same length showed the same mobility under denaturing conditions; no faster or slower migrating species were detected. Native gel electrophoresis was carried out at 7 °C according to Rippe et al., 1989. The molar extinction coefficients at 70 °C in 10 mM Na-cacodylate, pH 7.0 were taken to be 8.6 mM base$^{-1}$ cm$^{-1}$ at 264 nm for the oligonucleotides D1, D2, D3, D4, F5, F6 and F7 (Ramsing and Jovin, 1988; Rippe et al., 1989; Fig. 1). For $GA_{15}$, $GA_{25}$ and $GAA_{15}$ (Fig. 1), the value of 12.2 mM base$^{-1}$ cm$^{-1}$ at 255 nm was derived from Lee et al., 1980. The molar extinction coefficients at 25 °C of 13.7 mM base$^{-1}$ cm$^{-1}$ at 253 nm for GMP and of 15.4 mM base$^{-1}$ cm$^{-1}$ at 259 nm for AMP were adopted from Sprecher and Johnson, 1977. The values for $T_m$ and $\Delta H_{vH}$ for the helix-coil transition were determined from the temperature dependent UV absorption as described in Ramsing et al., 1989. The vacuum-UV CD spectra were acquired with a 1 mm cell in a Jasco J-720 spectropolarimeter in a solution of 90% $D_2O$, 0.8 mM Na-cacodylate and 2 mM Na-phosphate pH 7.0, and 0.8 mM NaCl.

### FLUORESCENCE MEASUREMENTS OF PYRENE LABELED OLIGONUCLEOTIDES

The oligonucleotides with a 5'-primary aminohexyl linker were labeled with succinimidyl 1-pyrenebutyrate (Molecular Probes, Eugene, OR) and purified by Sephadex G-25 and reversed phase HPLC chromatography as described in Rippe et al., 1992. Fluorescence measurements were made with an SLM 8000 photon counting spectrofluorimeter at 7 °C, unless otherwise noted, and at a DNA concentration corresponding to $A_{260}$ = 0.1-0.2. The emission spectra were corrected by background subtraction and internal standards. Excitation was at 340 nm. For measurements at neutral pH, all solutions contained 10 mM Na-cacodylate, pH 7.0, in addition to other specified salt components.

MODEL FORCE FIELD CALCULATIONS

The SYBYL (Tripos) molecular modelling package was used on an Evans and Sutherland ESV30 graphics workstation to built the $ps_{RY}$-DNA structures and the junction from $ps_{AT}$-DNA to $ps_{RR}$-DNA. The starting structure for the $ps_{RY}$-DNA helix with A·C and G·T base pairs was constructed from the coordinates of standard aps B-DNA with the same designated purine strand. The bonds between all pyrimidine residues were broken, the pyrimidine nucleotides were rotated by 180° about the axis connecting the C6 and C4 atoms of the purine and pyrimidine, respectively, and the T residues were mutated to C or vice versa so as to obtain *trans* A·C and G·T base pairs. The broken bonds were restored, establishing an initial helical structure. The starting model for the $ps_{RR}$-DNA helix was generated and refined on an Evans and Sutherland PS300 graphics system as described in Rippe et al., 1992. All structures were then energy minimized by applying the AMBER-PMF force field (Klement et al., 1990) or AMBER 3.0 (Weiner et al., 1986). The figures were generated with the program SCHAKAL v86b (Dr. E. Keller, Univ. of Freiburg, FRG).

## RESULTS

DUPLEX FORMATION AND STABILITY

We synthesized the oligonucleotide sequences shown in Fig. 1. By mixing the strands with parallel complementary, i.e., D1+D2, D3+D4 and F5+F6, the parallel-stranded duplexes ps-D1·D2, ps-D3·D4 and ps-F5·F6 with A·T base pairing were obtained (Fig. 1). The antiparallel-stranded (aps) reference duplexes were formed from the combinations D1+D3 (aps-D1·D3), D2+D4 (aps-D2·D4) and F5+F7 (aps-F5·F7) (Fig. 1; Rippe et al., 1989; Jovin et al., 1990; Rippe and Jovin, 1992). The gel electrophoretic mobilities of these duplexes under native conditions ($MgCl_2$, 7 °C) in a 14% acrylamide gel were compared to those of the oligonucleotides $GA_{15}$ and $GA_{25}$.


Fig. 1. Sequences and duplexes used in this study. The designation ($NH_2$) indicates that these sequences were also synthesized with a 5'-terminal aminoalkyl group to enable labeling with the fluorescent dye pyrene (see below). To the right are examples of ps and aps duplexes adopted by these sequences.

The ratios of the electrophoretic mobilities to those of single-stranded oligonucleotide of the same length (F5 or D1) were: $GA_{15}$ (0.60), $GA_{25}$ (0.61), ps-F5·F6 (0.66) and ps-D1·D2 (0.65). The values are similar (in some cases ~5% higher) than those for the corresponding antiparallel-stranded duplexes aps-F5·F7 (0.66) and aps-D1·D3 (0.61). We conclude that duplex structures were formed in the DNA concentration range studied (2-100 μM strands). There was no evidence for higher order species such as triplexes or tetraplexes, which would have been appreciably retarded with respect to the duplexes.

The $GA_{15}$, $GA_{25}$ and $GAA_{25}$ duplexes had melting points, $T_m$, determined by UV spectroscopy of 25 °C, 40 °C and 34 °C, respectively, in 10 mM $MgCl_2$, 10 mM Na-cacodylate, pH 7.0, and at a strand concentration of 2 μM. Heating (10 → 70 °C) led to a hyperchromicity of ~24%. The van't Hoff enthalpy difference for the helix-coil transition, $\Delta H_{vH}$, was 210 kJ $mol^{-1}$ for $GA_{15}$, and the entropic change $\Delta S$ was 0.60 kJ $mol^{-1}$ $K^{-1}$. These values correspond to average thermodynamic parameters per formed GA/GA and AG/AG stack of $\Delta H_S \approx$

-18 kJ $mol^{-1}$ and $\Delta S_s \approx$ -0.054 kJ $mol^{-1}$ $K^{-1}$, assuming the nucleation terms $\Delta H_\beta$ = 44 kJ $mol^{-1}$ and $\Delta S_\beta$ = 0.15 kJ $mol^{-1}$ $K^{-1}$ for ps linear duplexes (Rippe and Jovin, 1992). It should be noted that the stacking interactions for the GA/GA as compared to the AG/AG step are very different in the model discussed below. The thermodynamic values for the $ps_{RR}$-DNA helix can be compared to those determined for $ps_{AT}$-DNA: $\Delta H_s$ = -24 kJ $mol^{-1}$ and $\Delta S_s$ = -0.072 kJ $mol^{-1}$ $K^{-1}$ for AA/TT, and $\Delta H_s$ = -18 kJ $mol^{-1}$ and $\Delta S_s$ = -0.058 kJ $mol^{-1}$ $K^{-1}$ for AT/TA stacking interactions. These values were derived from an analysis of the available data for various ps constructs (Rippe and Jovin, 1992). With respect to the melting enthalpy, the ps helix formed by alternating d(G-A) appears to be less stable than that adopted by homo-oligomeric d(A)·d(T) sequences but is similar to $ps_{AT}$-DNA formed with strictly alternating d(A-T).

CIRCULAR DICHROISM SPECTROSCOPY

The duplexes aps-D2·D4 (Fig. 2A) or aps-D1·D3 (Fig. 2B) form right-handed B-type helices with Watson-Crick A·T base pairing. This conformation is characterized by a positive peak from 185-200 nm in the CD spectrum with a maximum at about 190 nm (Fig. 2A, 2B) (Sutherland et al., 1986). Native ps-D3·D4 in 100 mM KF (Fig. 2A) and ps-D1·D2 in 2 mM $MgCl_2$ (Fig. 2B) had similar spectra in the vacuum-UV region, from which we conclude that the $ps_{AT}$-DNA helix is right-handed with all the bases in an *anti* conformation. This result is supported by other spectroscopic and biochemical studies (Germann et al., 1989; Otto et al., 1991; Klysik et al., 1991). From theoretical considerations, a contribution reflecting the different strand orientation is not expected in this region of the CD spectrum (Germann et al., 1989). However, the spectra of the aps and ps duplexes differ significantly between 255-285 nm and provide criteria by which to distinguish between the two DNA conformations.

The CD signals of the single-stranded oligonucleotides arising from ps-D3·D4 and aps-D2·D4 upon denaturation at 50 °C were indistinguishable (Fig. 2C). This result was expected inasmuch as a simple reversal of a given sequence should lead to very minor

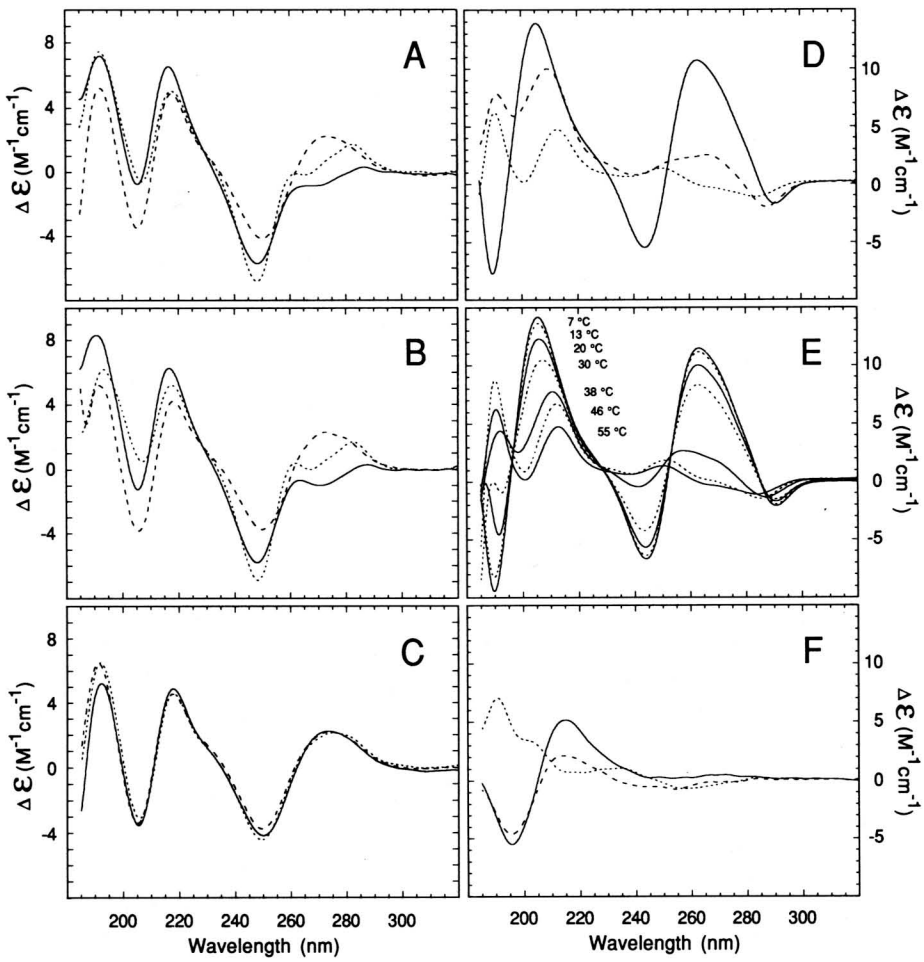


Fig. 2. Circular dichroism (CD) spectra of $GA_{15}$. (A) Spectra with 100 mM KF of ps-D3·D4 at 15 °C (——), at 50 °C (- - -), and of aps-D2·D4 (····) at 15 °C. (B) Spectra with 2 mM $MgCl_2$ of ps-D1·D2 at 15 °C (——), at 50 °C (- - -), and of aps-D1·D3 (····) at 15 °C. (C) Spectra in the denatured state: ps-D3·D4 at 20 °C without salt (- - -) and at 50 °C with 100 mM KF (——), and aps-D2·D4 at 50 °C with 100 mM KF (····). (D) $GA_{15}$ at 7 °C without $MgCl_2$ (- - -), and with 5 mM $MgCl_2$ at 7 °C (——), and at 55 °C (····). (E) $GA_{15}$ in 5 mM $MgCl_2$ at increasing temperatures. (F) 3'-GMP (——), 5'-GMP (- - -), and 5'-AMP (····) at 25 °C.

changes in the spectra of single-stranded oligonucleotides. In addition, it is apparent from the spectrum of ps-D3·D4 at 20 °C without added salt (Fig. 2C) that either monovalent or divalent cations are required to induce and stabilize the secondary structure. A detailed account of the CD of the D-duplexes will be published elsewhere (R. Mohan, K. Rippe and T. M. Jovin, in preparation).

The CD spectrum of $GA_{15}$ demonstrated the formation of secondary structure upon addition of 5 mM $MgCl_2$ at 7 °C (Fig. 2D). There was a strong positive peak at 263 nm and a negative peak at 190 nm. The 263 nm peak has been reported previously for the synthetic polymer poly[d(G-A)] at neutral pH (Lee et al., 1979) and for $d(G-A)_{30}$ at pH 4.0 (Antano et al. 1988). Denaturation of $GA_{15}$ by heating to 55 °C led to the disappearance of the positive CD peak at 263 nm and to the inversion of the peak at 190 nm, i.e. from negative to positive values (Fig. 2D). The temperature dependence of the circular dichroism is displayed in Fig. 2E. The set of spectra recorded at various temperatures demonstrates that the melting of the duplex structure adopted by $GA_{15}$ can be monitored by CD spectroscopy. The apparent melting point of ~30 °C agreed well with the $T_m$ of 28 °C estimated from UV absorption data under the same experimental conditions (5 mM $MgCl_2$, 20 μM strands).

The negative peak at 190 nm in the CD spectrum of $GA_{15}$ in the duplex state is unusual and not observed with the conventional B-DNA conformation. A negative peak at this wavelength has been considered to be a marker of a left-handed helical structure, inasmuch as it has appeared previously only with left-handed Z-DNA (Riazance et al., 1987). Although the model for $ps_{RR}$-DNA presented below is right-handed, it shares several features found in Z-DNA, namely the *syn* glycosyl configuration of dG and a dinucleotide repeat unit. Thus, we would suggest that the sign of the CD peak at or near 190 nm may reflect these helical parameters and not handedness per se. In support of this view are CD measurements of the constituent nucleotide monomers of $GA_{15}$ (Fig. 2F) that reveal negative peaks at 195 nm for both 5'-GMP and 3'-GMP (see also Sprecher and Johnson, 1977). The intensity of the 3' mononucleotide was

~20% higher. NMR data at slightly alkaline pH (Son et al., 1972) demonstrate that about 50% of 5'-GMP and 70% of 3'-GMP are in the *syn* conformation. In a single-stranded oligonucleotide with conventional stacking, dG residues adopt the energetically favored *anti* conformation reflected in the positive peak at 190 nm in the spectra of $GA_{15}$ measured in the absence of $MgCl_2$ or in the denatured state at 55 °C.

PYRENE MONOMER-EXCIMER FLUORESCENCE, A MEASURE OF STRAND ORIENTATION

We determined the strand orientation in the duplex formed by the GA-oligonucleotides by exploiting the fluorescent properties of the polycyclic compound, pyrene (Rippe et al., 1992). Excited state pyrene, generated by exposure to 300-350 nm light, can form a dimer (excimer) with another molecule in the ground state if the latter is in close proximity either as a preformed complex or a rapidly diffusible species (Birks, 1970). The excimer has a distinctive spectral signature – a broad unstructured fluorescence in the region 430-600 nm – in contrast to the structured monomer emission at 380-400 nm. A pyrene molecule was attached to the 5'-end of the oligonucleotides via an aminoalkyl linker. The expectation was that the excimer species should be potentiated in ps-DNA, due to the direct apposition of the 5'-ends of the two strands, whereas with antiparallel B-DNA, in which the 5'-ends at opposite ends of the helix, only the monomer fluorescence would be observed. This expectation was fulfilled with the set of molecules studied here (Fig. 3). Upon addition of $MgCl_2$ to stabilize the duplex species, the characteristic excimer emission developed with pyrene labeled $GA_{25}$, $GA_{15}$ and $GAA_{25}$ (Fig. 3C) as well as with the parallel-stranded, but not the antiparallel-stranded, reference duplexes (Fig. 3A). The parallel duplex structure adopted by $GA_{15}$ was stable in the pH range from 4.0-9.0 indicating that the structure is compatible with, but does not require, protonation of the dA residues (Fig. 3D). Thus, this simple technique served to confirm the proposed relative strand orientation in $ps_{AT}$-DNA and $ps_{RR}$-DNA compared to coventional aps(B)-DNA. The method has general

applicability to other situations, e.g. involving multistranded structures, in which one wishes to distinguish between alternative strand orientations.

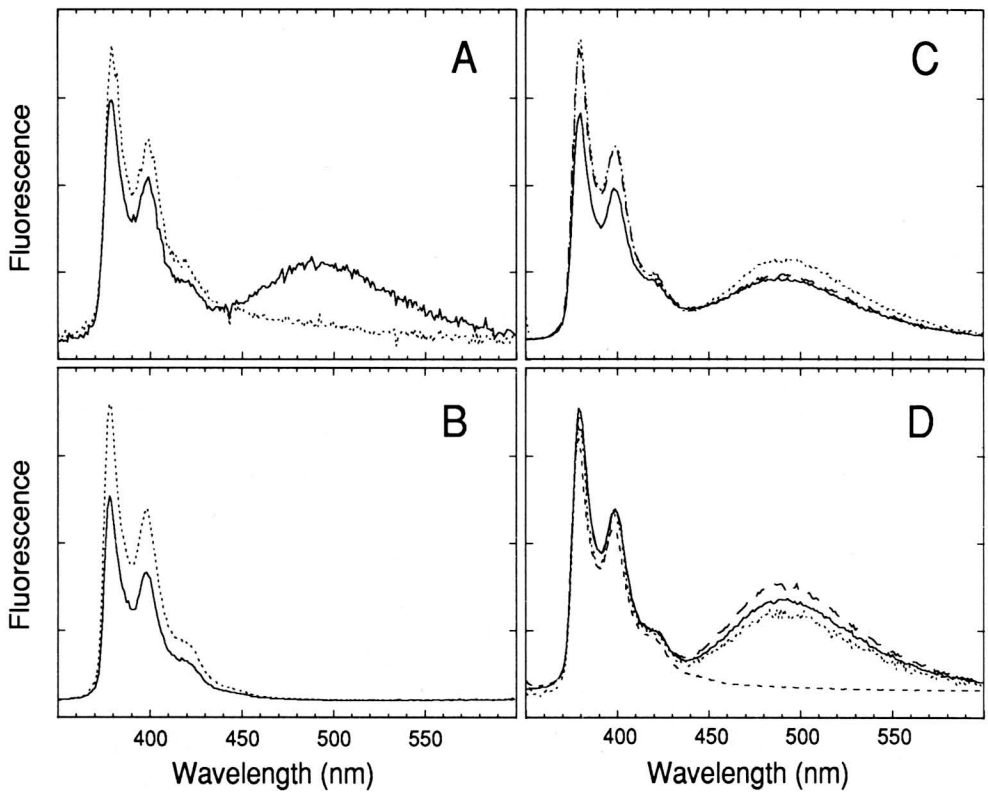


Fig. 3. Fluorescence emission spectra of pyrene-labeled oligonucleotides. (A) Duplexes ps-D3·D4 (——) and aps-D2·D4 (····) in 2 mM $MgCl_2$.(B) Single-stranded oligonucleotide D1 without $MgCl_2$ (····) and in 10 mM $MgCl_2$ (——). Note that addition of $MgCl_2$ did not lead to a fluorescence signal indicative of excimer formation, but rather to a quenching of the monomer fluorescence.(C) $GA_{25}$ (——), $GAA_{25}$ (---), and $GA_{15}$ (····), in 10 mM $MgCl_2$.(D) Effect of pH on the excimer formation of $GA_{15}$ in 10 mM $MgCl_2$ in 10 mM Na-acetate, pH 4.0 (····), in 10 mM Na-cacodylate, pH 7.0 (——), in 10 mM Na-borate, pH 9.0 (– – –) and in 10 mM NaOH, pH 12.0 (---). Data from Rippe et al., 1992.

## BASE PAIRING AND MODEL STRUCTURES

The DNA double helix can be regarded as a hierarchical structure, in which stereochemical, quantum chemical, and electrostatic atomic interactions define regions of stability in a multidimensional phase diagram. It has become apparent that a useful classification of potential ps- and aps-DNA helical structures requires the systematic specification of base pair geometry, glycosyl bond orientation and backbone geometry (Kuryavyi, manuscript in preparation). In particular, one can argue that the stereochemistry of the base pair at the nucleoside level defines the available repertoire for the disposition of the backbone. That is, the relative strand orientation (ps, aps) is determined by: (i) the hydrogen bonding scheme that establishes the glycosyl bond orientation (*cis* or *trans*); (ii) the glycosyl bond torsions (*anti* or *syn*); and (iii) the relative sugar orientation which is correlated to the backbone torsion angles and the sugar pucker. In a particular sequence, nearest neighbor and higher order axial interactions, notably stacking, serve to further delimit the range of feasible helical conformations and junctional transitions. These considerations are described extensively elsewhere (V. Kuryavyi, manuscript in preparation; T. M. Jovin, K. Rippe, V. Kuryavyi, A. Garcia, manuscript in preparation). In the models of parallel-stranded helices presented in this report, the hydrogen-bonding schemes are such that the two glycosyl bonds of each base pair are in a *trans* conformation.

## $PS_{RY}$-DNA HELIX WITH A·T AND G·C BASEPAIRING

The model in Fig. 4 shows a parallel-stranded helix composed of *trans* A·T base pairs with N6H...O2 and N1...HN3 hydrogen bonding and a *trans* G·C base pair in the central position (Figs. 9A, 9B). The G·C base pair has a single N1H...N3 hydrogen bond and exhibits considerable propeller twist and buckle, which serve to minimize clash between the exocyclic amino groups. Force field calculations indicate that this conformation is favored for a single *trans* G·C base pair in a helical segment consisting otherwise of *trans* A·T base pairs, probably because the two base pairs are nearly

isomorphic (Rippe et al., 1990; Figs. 9A, 9B). Thus, stacking

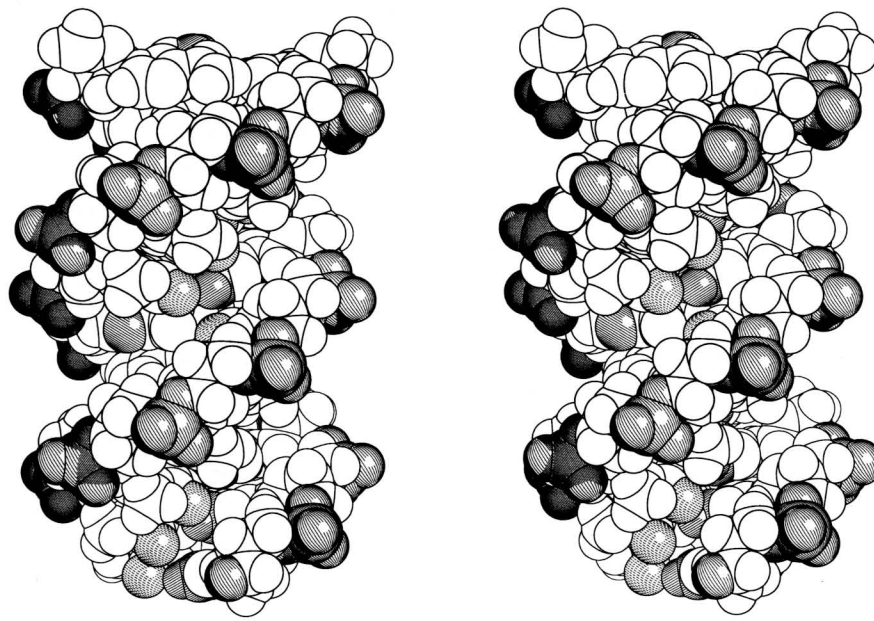

Fig. 4. Space-filling stereo model of $ps_{RY}$-DNA with d(A·T) base pairing and a single d(G·C) base pair in the middle. The sequence ps-d($T_5GA_5$)·d($A_5CT_5$) is shown. Identified atoms: C2H of A (right diagonal), N7 of G (left diagonal), O2 of C (right stippled), and O4 of T (circularly stippled). Phosphates in the backbone are marked by curved lines. The 5'-ends are at the top. Data from Rippe et al., 1990.

interactions may be dominant in establishing the structure.

Within a tract of several *trans* G·C base pairs, hydrogen bonding between N1H...O2 and N2H...N3 (Fig. 9C) appears to be more favorable according to current model calculations because it provides stabilization from the additional hydrogen bond and avoids steric clashes between the N2 amino group of G and the N4 amino group of C. We note that the RNA equivalent of such a base pair is found in yeast $tRNA^{phe}$ between G15 (dihydrouridine loop) and C48 (variable loop) (Kim et al., 1974). The corresponding interaction in yeast $tRNA^{asp}$ is mediated by a *trans* $A_{15}·U_{48}$ base pair (Westhof et al., 1985).

$PS_{RR}$-DNA HELIX

The $ps_{RR}$-DNA helix structure has a two-fold symmetry axis along the helical axis. Thus, the glycosyl bonds have a *trans* orientation

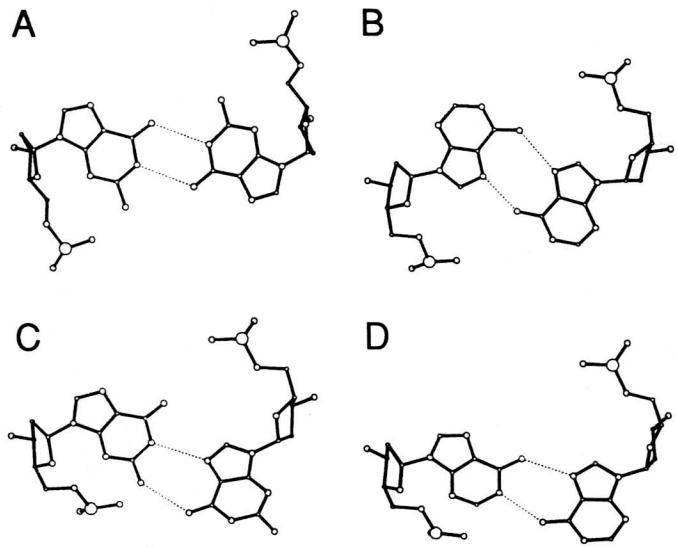


Fig. 5. Base pairing schemes for the proposed parallel-stranded GA duplex structure including non-fully alternating (A,G) sequences (A) Symmetrical $G_{syn}\cdot G_{syn}$ base pair. (B) Symmetrical $A_{anti}\cdot A_{anti}$ base pair. (C) $G_{syn}\cdot G_{anti}$ base pair mimicking $A_{anti}\cdot A_{anti}$ in the AG/G step. (D) $A_{syn}\cdot A_{anti}$ base pair mimicking $G_{syn}\cdot G_{syn}$ in the GA/A step.

and the purine residues form the self-pairs G·G with N1H...O6 hydrogen bonds (Figs. 5A, 9D) and A·A with N6H...N7 hydrogen bonds (Figs. 5B, 9G). These constraints impose the *syn* conformation for the guanine base about the glycosyl bond and the *anti* conformation for the adenine base. A space-filling model is shown in Fig. 6. The sugar pucker of the model is C3'-*endo*. From this feature and the results obtained for purine polymers (Lee et al., 1980), we surmise that an RNA oligomer may adopt a very similar structure, a possibility under current investigation.

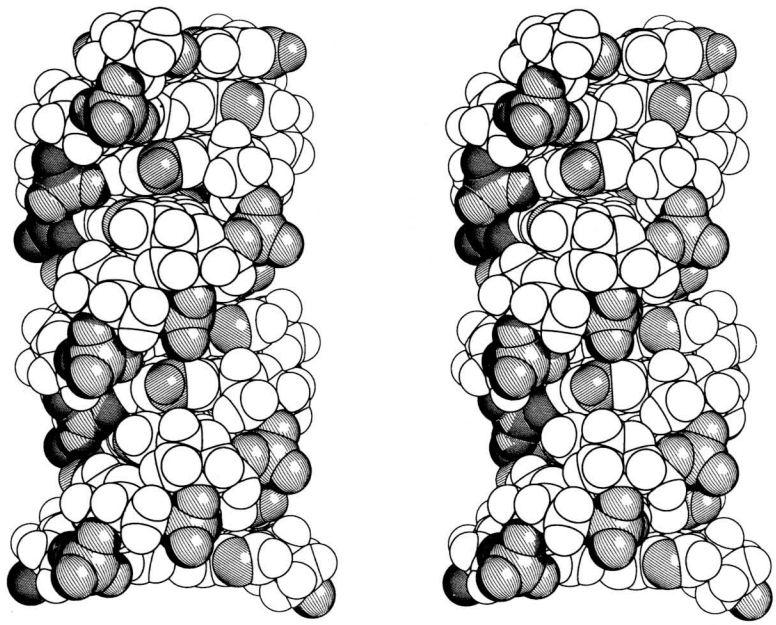

Fig. 6. Space-filling stereo model of $ps_{RR}$-DNA. The sequence ps-d(A-G)$_5$·d(A-G)$_5$ is shown. Atoms are marked as in Fig. 4; the 5'-ends are at the top. Data from Rippe et al., 1992.

Although it breaks the two-fold symmetry, a mixture of *anti* and *syn* conformations for the homopurine base pairs is also possible. These two non-symmetrical base pairings have hydrogen bonds between N1H...N7 and N2H...O6 for the $G_{syn}$·$G_{anti}$ base pair (Figs. 5C, 9E) and between N6H...N1. and N7...HN6 for the $A_{anti}$·$A_{syn}$ base pair (Figs. 5D, 9H). Such base pairings are probably adopted by oligonucleotides lacking a strict alternation of G and A but still capable of forming a similar ps duplex. The fact that the model is able to accommodate a lack of strict (A,G) alternation makes it particularly attractive in terms of biological occurrence and implications.

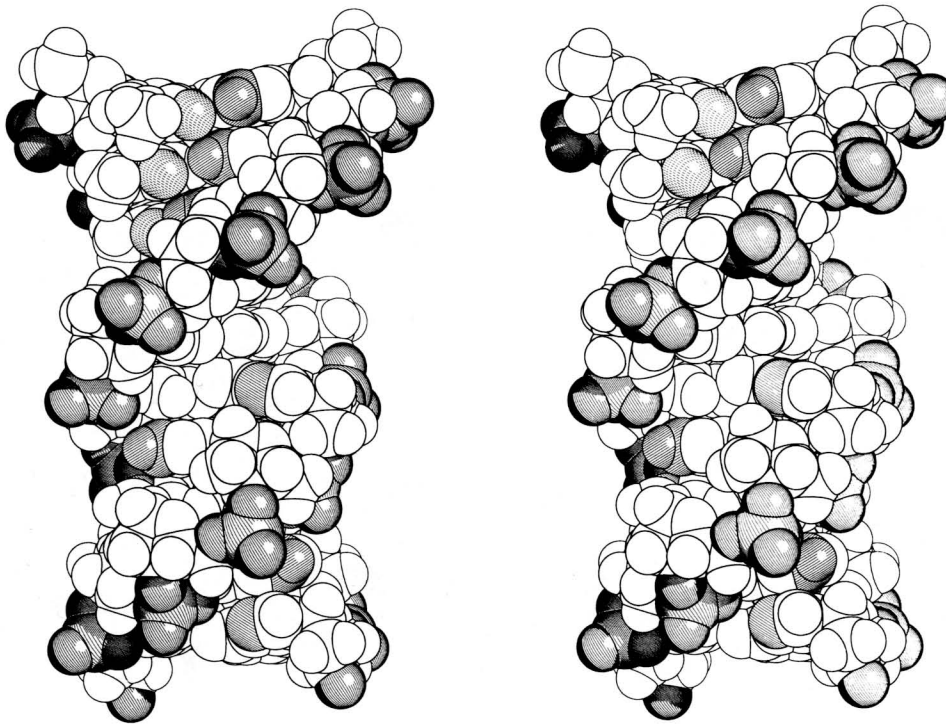

Fig. 7. Space-filling stereo model of combined $ps_{AT}$-DNA and $ps_{RR}$-DNA helices, formed by the juxtaposed sequence ps-d($T_5$(GA)$_2$G)·d($A_5$(GA)$_2$G). Atoms are marked as in Fig. 4; the 5'-ends are at the top.

THE JUNCTION OF THE $PS_{RY}$-DNA HELIX WITH THE $PS_{RR}$-DNA HELIX

The occurrence of parallel-stranded interactions in natural sequences would be facilitated if alternations can occur between $ps_{RY}$-DNA and $ps_{RR}$-DNA helical segments. To pursue this point, we have studied the stereochemical feasibility of such a structure by model building and energy minimization. Our preliminary results indicate that the transition from a $ps_{AT}$-DNA helix (Fig. 4) to a $ps_{RR}$-DNA helix (Fig. 6) is possible without significant distortion of the backbone. The A·T/G·G junction shown in Fig. 7 appears somewhat more favorable than the A·T/A·A step because the distance

between the two 5'-phosphate groups in the G·G nucleotide pair (~14 Å) is closer to that of the A·T pair (~18 Å) than to that of the A·A pair (~11 Å).

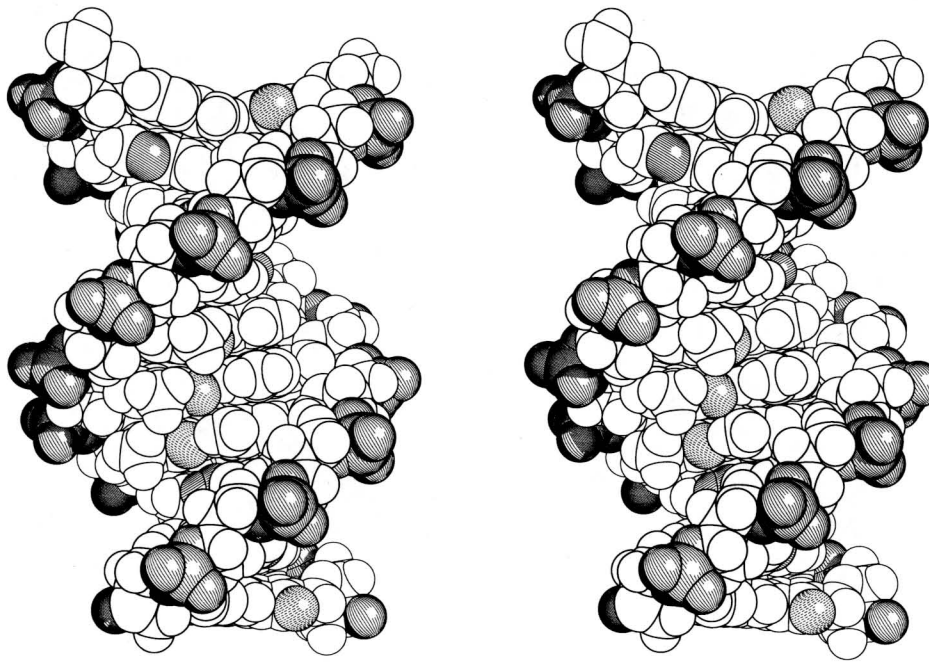

Fig. 8. Space-filling stereo model of a parallel-stranded helix with *trans* A·C and G·T base pairing consisting of the sequence ps-d($A_1GA_2G_3AGA$)·d($C_1TC_2T_3CTC$). Atoms are marked as in Fig. 4; the 5'-ends are at the top.

## $PS_{RY}$-DNA HELIX WITH A·C AND G·T BASEPAIRING

Different pur·pyr base pairs constituted from the four natural bases, stabilized by at least 2 hydrogen bonds, and isomorphic with respect to the positions of the N-glycosidic atoms, can be classified into two sets. The first comprises the canonical Watson-Crick A·T and G·C base pairs of aps-DNA helices. The second set corresponds to ps-DNA constructs with *trans* A·C (Fig. 9I) and G·T (Fig. 9F) base pairs. In *trans* A·C, the hydrogen bonds are

between N1...HN4 and N6H...N3, whereas in *trans* G·T, hydrogen bonding is between N1H...O4 and O6...HN3. Thus, $ps_{RY}$-DNA can have different complementarity principles than aps B-DNA. A space-filling model of the $ps_{RY}$-DNA helix with *trans* A·C and G·T base pairing in the sequence ps-d($A_1GA_2G_3AGA$)·d($C_1TC_2T_3CTC$) duplex is displayed in Fig. 8. The A·C and G·T base pairs are nearly isomorphic and have symmetrical glycosidic bonds with respect to the helix axis. This implies that a very similar structure with any given sequence of purine and pyrimidines can be constructed.

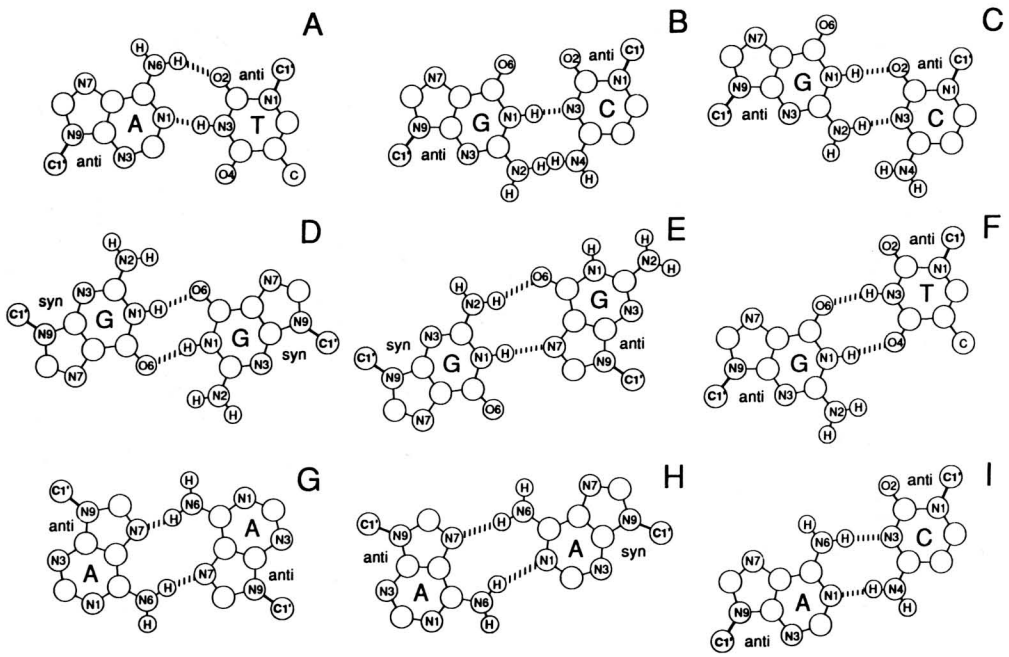


Fig. 9. Base pairing scheme for the proposed parallel-stranded duplex structures. The *trans* base pairs are derived from the positions of the minimized duplex structures.

A TETRAPLEX WITH FOUR PARALLEL STRANDS AND $A^+$·G·$A^+$·G QUARTETS

The model of $ps_{RR}$-DNA (Fig. 6) can be extended to build a fully parallel tetraplex structure based on $A^+$·G·$A^+$·G tetrads with all bases in *anti* conformation and the adenines protonated at the N1

position. Figure 10A shows a stereo view of such a structure in which all four strands are in a parallel orientation. The hydrogen

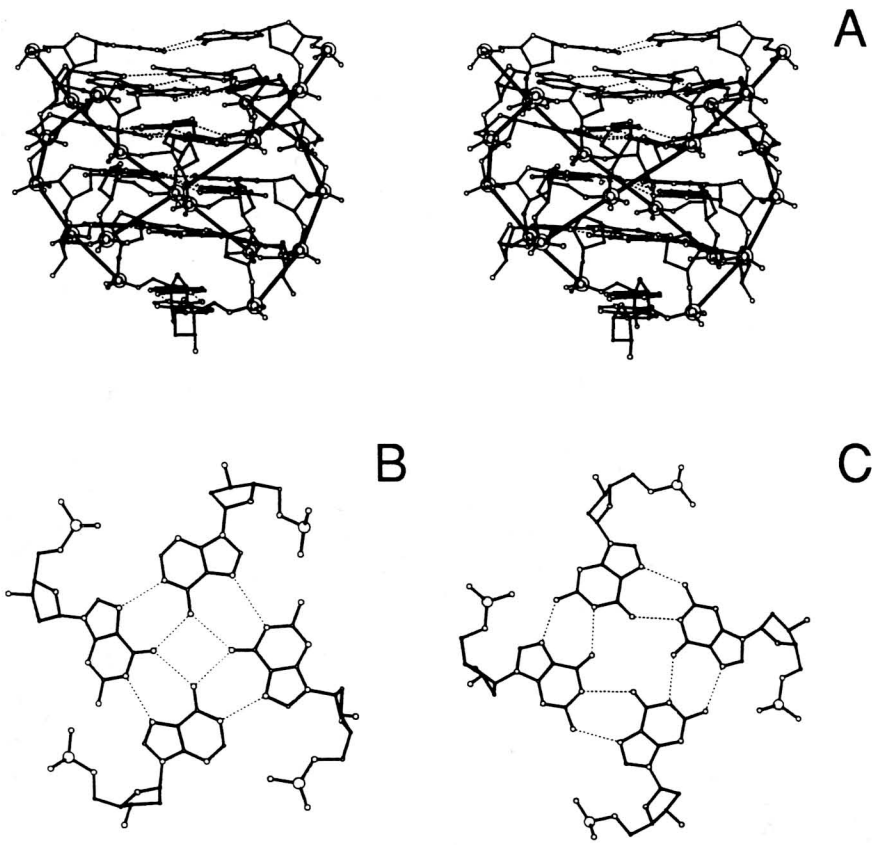


Fig. 10. (A) Stereo views of a quadruplex structure formed from four d(AGAGA) strands with all strands parallel to each other. The dangling ends have symmetrical A·A pairs with N6H...N1 hydrogen bonds. (B) Possible base pairing for a $A^+$·G·$A^+$·G quartet with the adenine bases protonated at N1. (C) Base pairing for a G·G·G·G quartet (Sen and Gilbert, 1988) shown for comparison. Notice the much larger central opening than in the $A^+$·G·$A^+$·G quartet.

bonding is between N6H...O6 and $N1H^+$...N7 (Fig. 10B).

The $A^+$·G·$A^+$·G tetrad is not isostructural to the A·A·A·A

quartets proposed earlier for tetraplexes formed by (G,A) sequences (Lee, et al., 1980; Sundquist & Klug, 1989), since all base pairs entail two hydrogen bonds instead of one as in previous models and steric clash between the central amino groups is avoided. The proposed $A^+ \cdot G \cdot A^+ \cdot G$ quartet is, however, isomorphic with the G·G·G·G quartets formed by four parallel strands and all guanines in the *anti* conformation (Fig. 10C, Sen and Gilbert, 1988; Sundquist and Klug, 1989). In order to build a quadruplex stabilized by $A^+ \cdot G \cdot A^+ \cdot G$ quartets with the strands alternatively antiparallel and parallel, two bases of each tetrad (most probably the guanines) would have to adopt the *syn* conformation.

## DISCUSSION

As demonstrated above, a number of parallel-stranded helical interactions can exist. Structures with *trans* A·T, G·C, A·A and G·G base pairing have been demonstrated experimentally and others, for example with *trans* A·C and G·T base pairs (Figs. 8, 9F, 9I; Kuryavyi, manuscript in preparation), may be possible (see also Il'ychova et al., 1990). Examples exist in the literature (Sarma et al., 1986; Antano et al., 1988) of the third potential canonical ps helix, $ps_{YY}$-DNA; these structures, however, are not stable at neutral pH. The different complementarity relationships compatible with ps-DNA have distinct biological implications. Sequence motifs which would facilitate the formation of parallel-stranded interactions based on $ps_{RY}$-DNA have been identified previously (Ramsing and Jovin, 1988; Jovin, 1991, and references therein) and can be extended by considering the expanded array of possible base pairs. In the following, we discuss some situations illustrating the potential involvement of parallel-stranded DNA and RNA structures in cellular mechanisms.

### CATALYTIC MECHANISM OF REVERSE GYRASE FROM THERMOPHILIC BACTERIA INVOLVING A PARALLEL-STRANDED INTERMEDIATE

A unique topoisomerase activity, reverse gyrase (RG), has been

found in the extremely thermophilic archaebacteria *Sulfolobus acidocaldarius* (growing at 75 °C) and *Desulfurococcus amylolyticus* (growing at 92 °C) (reviewed in Kikuchi, 1990).

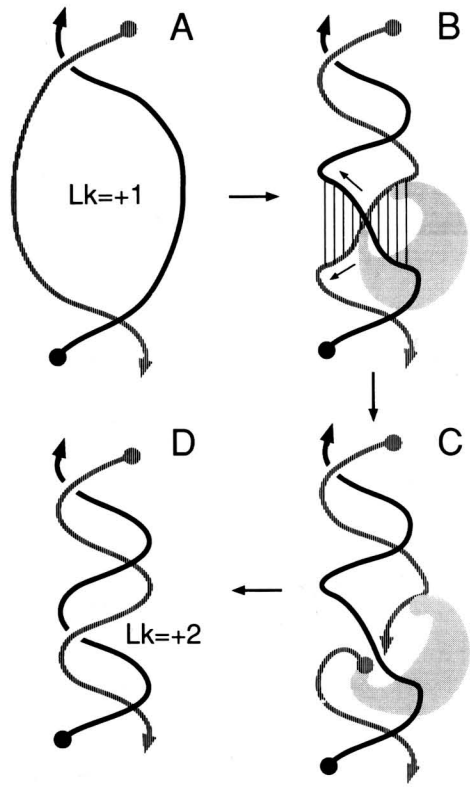


Fig. 11. Model of a mechanism for introduction of positive supercoils by reverse gyrase via a parallel-stranded intermediate. (A) Initial unwound structure with Lk=+1. (B) Binding of enzyme with concommitant formation of parallel-stranded helical segments (putative base pairs shown by vertical lines). The resulting structure is equivalent to that in (A) with respect to the linking number. (C) Strand passage. (D) Religation, resulting in an unitary increase in linking number.

Upon hydrolysis of ATP, the enzyme introduces positive supercoils into covalently closed circular DNA. However, unlike the eubacterial gyrases, RG is a type I topoisomerase, i.e., it introduces a transient single-strand break, thereby leading to an in-

crease in the linking number by steps of 1.

The mechanism for the introduction of positive supercoils by RG has some similarity to that suggested for eubacterial toposisomerases I in that RG binds covalently to a 5'-phosphate and to a cytosine at position -4 to the cleavage site in the 5'-direction via a non-covalent bond (Jaxel et al., 1989; Kovalsky et al., 1990). However, the eubacterial type I topoisomerases can only relax negatively supercoiled DNA and do not require ATP. Slesarev and Kozyavkin (1990) have suggested that RG is able to recognize the strand orientation. Accordingly, we propose the model shown in Fig. 11 for the enzymatic mechanism of RG. This scheme requires the initial denaturation of the antiparallel duplex (Fig. 11A), which potentiates the rearrangement of strands shown in Fig. 11B in which the central part forms an intermediate with a parallel strand orientation stabilized by the enzyme. The strand passage reaction takes place in this segment (Fig. 11C), resulting in an increase of the linking number by 1 after religation (Fig. 11D).

## STABILIZATION OF GENOMIC STRUCTURES IN SINGLE-STRANDED DNA VIRUSES BY PARALLEL-STRANDED INTERACTIONS BETWEEN REMOTE SEQUENCES

The bacteriophage ϕX174 is a small icosahedral virus with a single-stranded, closed circular genome of 5386 nucleotide bases. The capsid has an external diameter of roughly 260 Å, with a protein shell that is ~30 Å thick, resulting in an internal diameter of ~200 Å. (McKenna et al., 1992). These geometrical constraints suggest that distinct physical and biochemical elements are required for compaction of the DNA. From Raman spectroscopy it is concluded that the unpackaged DNA of ϕX174 is highly self-structured because it exhibits about ~70% stacking, compared to the linear double-stranded replicative form, and about a third of the residues have a B-DNA backbone geometry (Benevides et al., 1991). This structure partly disappears upon packaging probably due to interactions of the DNA with the protein shell. From the Raman spectra, the packaged genome has only ~40% stacking and lacks an ordered backbone geometry.

A potential mechanism for compacting single-stranded DNA can be proposed in which remote sequence segments with appropriate complementarity interact so as to form parallel-stranded helical segments that compact the intervening DNA into loops (Ramsing and Jovin, 1988; Jovin, 1991; Fig. 12). A. Bolshoy and E. Trifonov (personal communication) have conducted an analysis of the ϕX174 and related G4 phage genomes based on a search for pairs of AT-rich ps-forming sequences separated by up to 200 bases. A Fourier analysis of the loop size distributions gave evidence for potential periodic ps interactions with a spacing of ~100-120 bases. The statistical significance and generality of these results remain to be established. In addition, the analysis has to be extended to consider the additional $ps_{RY}$ and $ps_{RR}$ complementarities discussed above and site-site separations of > 200 bases. The latter feature is important in view of the long-range interactions of the unpacked genome observed in the electron microscope (Edling and Ihler, 1980).

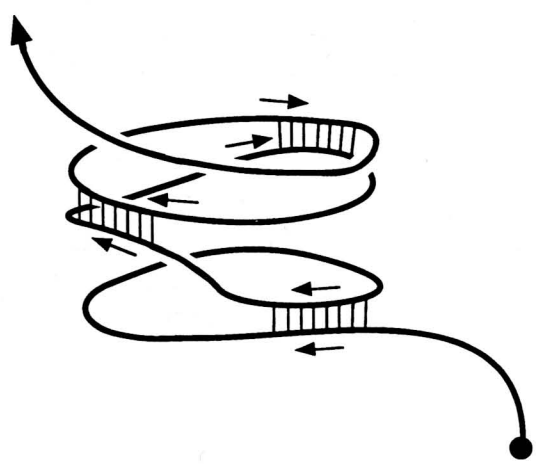

Fig. 12. Potential stabilization of genomic structures in single-stranded DNA bacteriophages by parallel-stranded helical segments. This schematic representation does not depict the topological consequences of such interactions. See text for details.

TRANSCRIPTIONAL CONTROL BY FORMATION OF A TRIPLEX STABILIZED BY A PARALLEL-STRANDED DNA·RNA HYBRID

Antibody class switching occurs by deletional intrachromosomal recombination in highly repetitive DNA sequences. A $d(AGGAG)_{28}$ repeat is found in the murine immunoglobulin IgA switch region (Collier et al., 1988; Reaban and Griffin, 1990). Transcription of this region leads to a loss of about 13 superhelical turns and the generation of a structure sensitive to RNase H (Reaban and Griffin, 1990). It has been proposed (Reaban and Griffin, 1990)

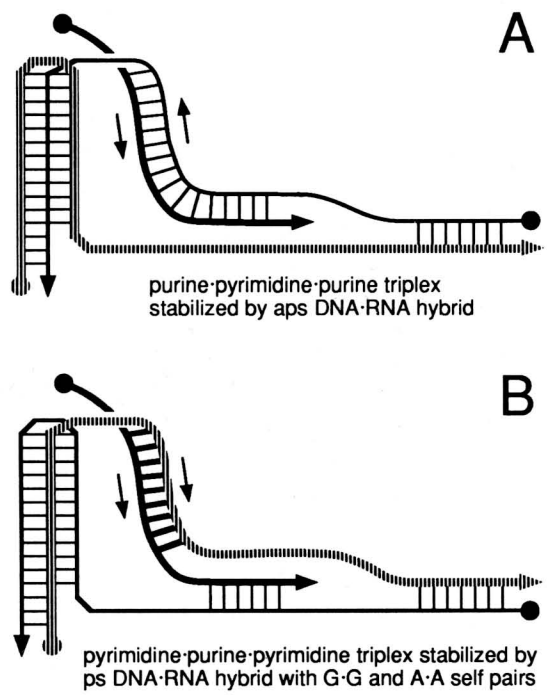


Fig. 13. Two alternative models for the stabilization of triplexes formed in the murine immunoglobulin IgA class switching region during transcription. The purine RNA transcript is shown as a thick solid line. The purine DNA strands are depicted by a line pattern. (A) An aps RNA·DNA hybrid stabilizes a pur·pyr·pur triplex. (B) A ps RNA·DNA duplex with G·G and A·A self-pairs stabilizes a pyr·pur·pyr triplex.

that a pur·pyr·pur triplex is formed in the $(AGGAG)_{28}$ tract that is stabilized by the $(AGGAG)_n$ RNA transcript [an error in the original paper has been corrected (Stavnezer, 1991)]. This structure is depicted in Fig. 13A. However, we note that for mixed GA sequences, the transcription independent formation of this type of a triplex would presumably require $Zn^{2+}$ or $Cd^{2+}$ (Bernués et al., 1990), neither of which was present in the assays (Reaban and Griffin, 1990). An alternative explanation for the experimental results is based on the formation of a parallel-stranded DNA·RNA hybrid with G·G and A·A self-pairs (Fig. 5,6), as shown in Fig. 13B. Such a structure would provide additional thermodynamic stabilization for a more conventional pyr·pur·pyr triplex.

## DIMERIZATION OF THE RETROVIRAL GENOME BY $PS_{RR}$-RNA INTERACTIONS

The genome of all retroviruses consists of two identical RNA strands bound together in a dimer linkage structure (DLS) near the 5'-end (Kung et al. 1976; Bender et al. 1978; Marquet et al., 1991). These dimers are observed by electron microscopy, which reveals a parallel orientation of the two constituent RNA strands (Kung et al. 1976; Bender et al. 1978) The dimerization of the RNA genome is regarded as an essential step in the retroviral life cycle and is catalyzed *in vivo* by the NCp15 protein (Darlix et al., 1990; Prats et al., 1990). Relatively short (~ 100-nt) fragments derived from the DLS region dimerize *in vitro* at neutral pH in the presence of divalent cations or polyamines; proteins are not required (Prats et al., 1990; Roy et al., 1990; Marquet et al., 1991; Tounekti et al., 1992). An example is the segment from 311-415 (cap= +1) in the RNA of the AIDS virus HIV-1, which has a purine content of 75% and 6 GA clusters with lengths of 5-13 residues. This fragment exhibits self-structure in the presence of $MgCl_2$ (Marquet et al., 1991).

The intervention of the polypurine sequences in the dimerization of the retroviral sequences is well established: (a) the dimer linkage structures of all the retroviruses have sequence clusters with a PuGGAPuA consensus (Marquet et al., 1991); (b) the

corresponding antisense RNAs (with polypyrimidine sequences) do not dimerize, also indicating that complementary purine and pyrimidines are not exchangeable as would be the case if Watson-Crick base pairs were involved. These facts have led to the suggestion that dimerization is based upon the formation of a purine quartet (Marquet et al., 1991; Kang et al., 1992). However, an alternative mechanism can be proposed. Inasmuch as self-pairing between A·A and G·G as in $ps_{RR}$-DNA offers the means by which identical sequences can hybridize in a parallel orientation, we postulate that dimerization of retroviral RNAs is mediated by the double-helical RNA analogue of $ps_{RR}$-DNA.

ACKNOWLEDGMENTS

We thank Gudrun Heim for excellent technical assistance, and Reinhard Klement for discussions and help with the DNA modeling.

REFERENCES

Antano, V.P., Gray, D.M., and Ratcliff, R.L. (1988) Nucleic Acids Res. 16, 719-738.

Bender, W., Chien, Y.-H., Chattopadhyay, S. , Vogt, P.K., Gardner, M.B., and Davidson, N. (1978) J. Virol. 25, 888-896.

Benevides, J.M., Stow, P.L., Ilag, L.L., Incardona, N.L., and Thomas, Jr., G.J. (1991) Biochemistry 30, 4855-4863.

Bernués, J., Beltrán, R., Casasnovas, J.M., and Azorin, F. (1990) Nucleic Acids Res. 18, 4067-4073.

Birks, J.B. (1970) Photophysics of aromatic molecules. John Wiley & Sons Ltd, New York.

Collier, D.A., Griffin, J.A., and Wells, R.D. (1988) J. Biol. Chem. 263, 7397-7405.

Darlix, J.-L., Gabus, C., Nugeyre, M.-T., Clavel, F., and Barré-Sinoussi, F. (1990) J. Mol. Biol. 216, 689-699.

Edling, T.D., and Ihler, G.M. (1980) J. Mol. Biol. 142, 131-144.

Germann, M.W., Kalish, B.W., Pon, R.T., and van de Sande, J.H. (1989) Biochemistry 28, 6220-6228.

Il'ychova, I.A., Lysov, Y.P., Chernyi, A.A., Shchyolkina, A.K., Gottikh, B.P., and Florentiev, V.L. (1990) J. Biomol. Struct. Dyn. 7, 879-897.

Jaxel, C., Nadal, M., Mirambeau, G., Forterre, P., Takahashi, M., and M., D. (1989) EMBO J. 8, 3135-3139.

Jovin, T.M., Rippe, K., Ramsing, N.B., Klement, R., Elhorst, W., and Voijtiskova, M. (1990) in Sarma, R.H., and Sarma, M.H.

(eds.), Structure and Methods, Vol. 3: DNA and RNA. Adenine Press, Schenectady, NY, pp. 155-174.
Jovin, T.M. (1991) in Eckstein, F., and Lilley, D.M.J. (eds.), Nucleic Acids and Molecular Biology, Vol. 5 , Springer-Verlag, Berlin/Heidelberg, pp. 25-38.
Kang, C.-H., Zhang, X., Ratliff, R., Moyzis, R., and Rich, A. (1992) Nature 356, 126-131.
Kikuchi, A. (1990) in Cozzarelli, N.R., and Wang, J.C. (eds.) DNA topology and its biological effects, Cold Spring Harbor University Press, Cold Spring Harbor, pp. 285-298.
Kim, S.H., Sussman, J.L., Suddath, F.L., Quigley, G.J., McPherson, A., Wang, A.H.-J., Seeman, N.C., and Rich, A. (1974) Proc. Natl. Acad. Sci. USA 71, 4970-4974.
Klement, R., Soumpasis, D.M., von Kitzing, E., and Jovin, T.M. (1990) Biopolymers 29, 1089-1103 .
Klysik, J., Rippe, K., and Jovin, T.M. (1991) Nucleic Acids Res. 19, 7145-7154.
Kovalsky, O.I., Kozyavkin, S.A., and Slesarev, A.I. (1990) Nucleic Acids Res. 18, 2801-2805.
Kung, H.-J., Hu, S., Bender, W., Bailey, J.M., Davidson, N., Nicolson, M.O., and McAllister, R.M. (1976) Cell 7, 609-620.
Lee, J.S., Johnson, D.A., and Morgan, A.R. (1979) Nucleic Acids Res. 6, 3073-3091.
Lee, J.S., Evans, D.H., and Morgan, A.R. (1980) Nucleic Acids Res. 8, 4305-4320.
Marquet, R., Baudin, F., Gabus, C., Darlix, J.-L., Mougel, M., Ehresmann, C., and Ehresmann, B. (1991) Nucleic Acids Res. 19, 2349-2357.
McKenna, R., Xia, D., Willingmann, P., Ilag, L.L., Krishnaswamy, S., Rossmann, M.G., Olson, N.H., Baker, T.S., and Incardona, N.L. (1992) Nature 355, 137-143.
Prats, A.-C., Roy, C., Wang, P., Erard, M., Housset, V., Gabus, C., Paoletti, C., and Darlix, J.-L. (1990) J. Virol. 64, 774-783.
Ramsing, N.B., and Jovin, T.M. (1988) Nucleic Acids Res. 16, 6659-6676.
Ramsing, N.B., Rippe, K., and Jovin, T.M. (1989) Biochemistry 28, 9528-9535.
Reaban, M.E., and Griffin, J. (1990) Nature 348, 342-344.
Rentzeperis, D., Rippe, K., Jovin, T.M., and Marky, L.A. (1992) J. Am. Chem. Soc., in press
Riazance, J.H., Johnson, Jr., W.C., McIntosh, L.P., and Jovin, T.M. (1987) Nucleic Acids Res. 18, 7627-7635.
Rippe, K., and Jovin, T.M. (1992) Meth. Enzymol. 211, 199-220, in press.
Rippe, K., Ramsing, N.B., and Jovin, T.M. (1989) Biochemistry 28, 9536-9541.
Rippe, K., Ramsing, N.B., Klement, R., and Jovin, T.M. (1990) J. Biomol. Struct. Dynam. 7, 1199-1209.
Rippe, K., Fritsch, V., Westhof, E., and Jovin, T.M. (1992), submitted.
Roy, C., Tounekti, N., Mougel, M., Darlix, J.-L., Paoletti, C., Ehresmann, C., and Ehresmann, B. (1990) Nucleic Acids Res. 18,

7287-7292.
Sarma, M.H., Gupta, G., and Sarma, R.H. (1986) FEBS Lett. 205, 223-229.
Sen, D., and Gilbert, W. (1988) Nature 334, 364-366.
Shimizu, M., Hanvey, J.C., and Wells, R.D. (1989) J. Biol. Chem. 264, 5944-5949.
Slesarev, A.I. (1990) J. Biomol. Struct. Dynam. 7, 935-942.
Son, T.-D., Guschlbauer, W., and Guéron, M. (1972) J. Am. Chem. Soc. 94, 7903-7911.
Sprecher, C.A., and Johnson, Jr., W.C. (1977) Biopolymers 16, 2243-2264.
Stavnezer, J. (1991) Nature 351, 447-448.
Sundquist, W.I., and Klug, A. (1989) Nature 342, 825-829.
Sutherland, J.C., Lin, B., Mugavero, J., Trunk, J., Tomasz, M., Santella, R., Marky, L., and Breslauer, K. (1986) Photochem. Photobiol. 44, 295-301.
Tchurikov, N.A., Shchyolkina, A.K., Borissova, O.F., and Chernov, B.K. (1992) FEBS Lett. 297, 233-236.
Tounekti, N., Mougel, M., Roy, C., Marquet, R., Darlix, J.-L., Paoletti, C., Ehresmann, B., and Ehresmann, C. (1992) J. Mol. Biol. 223, 205-220.
Weiner, S.J., Kollman, P.A., Nguyen, D.T., and Case, D. (1986) J. Comp. Chem. 7, 230-252.
Westhof, E., Dumas, P., and Moras, D. (1985) J. Mol. Biol., 184, 119-145.