2610–2617 *Nucleic Acids Research, 1999, Vol. 27, No. 13*

DNA structural transitions within the PKD1 gene

Richard T. Blaszak, Vladimir Potaman[1] and Richard R. Sinden[1] and John J. Bissler*

The Children's Hospital Research Foundation, 3333 Burnet Avenue, Cincinnati, OH 45229-3039, USA and [1]Center for Genome Research, Laboratory of DNA Structure and Mutagenesis, Institute of Biosciences and Technology, Texas A&M University, Houston, TX, USA

Received April 6, 1999; Accepted May 5, 1999

ABSTRACT

Autosomal dominant polycystic kidney disease (ADPKD) affects over 500 000 Americans. Eighty-five percent of these patients have mutations in the PKD1 gene. The focal nature of cyst formation has recently been attributed to innate instability in the PKD1 gene. Intron 21 of this gene contains the largest polypurine·polypyrimidine tract (2.5 kb) identified to date in the human genome. Polypurine·polypyrimidine mirror repeats form intramolecular triplexes, which may predispose the gene to mutagenesis. A recombinant plasmid containing the entire PKD1 intron 21 was analyzed by two-dimensional gel electrophoresis and it exhibited sharp structural transitions under conditions of negative supercoiling and acidic pH. The superhelical density at which the transition occurred was linearly related to pH, consistent with formation of protonated DNA structures. P1 nuclease mapping studies of a plasmid containing the entire intron 21 identified four single-stranded regions where structural transitions occurred at low superhelical densities. Two-dimensional gel electrophoresis and chemical modification studies of the plasmid containing a 46 bp mirror repeat from one of the four regions demonstrated the formation of an H-y3 triplex structure. In summary, these experiments demonstrate that a 2500 bp polypurine·polypyrimidine tract within the PKD1 gene is capable of forming multiple non-B-DNA structures.

INTRODUCTION

Autosomal dominant polycystic kidney disease (ADPKD) affects over 500 000 people in America, 50% of whom will develop renal failure by the age of 60 (1). The expenses associated with end stage management in this group of patients exceed $1 000 000 000 a year (2) and ADPKD is responsible for 8–10% of all chronic renal replacement therapy. Two of the three genes associated with the ADPKD phenotype, PKD1 and PKD2, have recently been identified (3–6). Eighty-five to 95% of patients with ADPKD have mutations in the PKD1 gene (7). The focal nature of cyst formation, affecting only 1–2% of nephrons (8,9), may be the result of a two-step process. In addition to an inherited germline mutation at one of the ADPKD alleles, cystogenesis appears to require inactivation or loss of heterozygosity at the normal allele (10–12). Undoubtedly, this would require a very high rate of somatic mutations to explain the large number of cysts (10 000–20 000) present in individual kidneys. This mutation frequency suggests that the PKD1 gene contains sequence characteristics that bestow innate genetic instability.

Analysis of the PKD1 gene from patients with ADPKD revealed that intron 21 of the PKD1 gene is prone to deletion and insertion mutations, with several size variants (9). Intron 21 of the PKD1 gene contains a 2500 bp polypyrimidine tract that, to date, is the largest identified in the human genome (9). Polypurine·polypyrimidine (Pu·Py) mirror repeats are known to form intramolecular DNA triplexes *in vitro* (13) and *in vivo* (14–17). Triplex DNA is hypothesized to be involved in genetic recombination (16,18–21), gene regulation (22–26), mutagenesis (27) and termination of DNA replication (28–31).

We hypothesize that this long polypyrimidine tract can form intramolecular triplex structures that then predispose the normal allele to mutation and loss of function. An analysis of this region identified many mirror repeats that could potentially engage in intramolecular triplex formation (32). Therefore, using chemical modification, two-dimensional gel electrophoresis and P1 nuclease sensitivity, we analyzed DNA structural transitions adopted by intron 21 in a recombinant plasmid.

MATERIALS AND METHODS

Plasmids

Plasmid pCW31 is a 3.1 kb *Sau*3A fragment of the PKD1 gene, containing intron 21 and a small amount of exonic sequence, cloned in pBluescript (a kind gift of Peter C. Harris, Institute of Molecular Medicine, John Radcliff Hospital, Oxford, UK). Intron 21 is 2.5 kb in length and contains 65% cytosine and 32% thymidine arranged in an imperfect $(CCTCCCCT)_n$ repeat. pCW2966 contains a 46 nt perfect mirror repeat (CCTCCTCCTCCCCCTCCTCCCTCCTCCCTCCTCCCCC-TCCTCCTCC) from the PKD1 gene (bp 34613–34658) cloned into the *Eco*RI site in pUC19. This sequence was selected for analysis because it represents the largest perfect mirror repeat within intron 21 and maps to a P1-sensitive site in pCW31 (see below). pCW2966 was produced by digestion of pUC19 with *Eco*RI, followed by phenol/chloroform extraction, ethanol precipitation and dephosphorylation with calf intestinal alkaline phosphatase (CIP). Oligonucleotides were synthesized at the University of Cincinnati DNA core facility and phosphorylated

*To whom correspondence should be addressed. Tel: +1 513 636 4531; Fax: +1 513 636 7407; Email: john.bissler@chmcc.org

with T4 polynucleotide kinase. Complementary strands were mixed in equimolar amounts and annealed by heating to 95°C and then slowly cooling to room temperature. The resulting double-stranded oligonucleotides were gel purified on a 12.5% polyacrylamide gel, eluted and concentrated by ethanol precipitation. The phosphorylated annealed oligonucleotides were ligated with *Eco*RI/CIP-treated pUC19 using T4 DNA ligase (16°C, 18 h, insert to vector molar ratio 1:10). The plasmid was propagated in a DH5α host. Constructs were verified by nucleotide sequencing of the insert using the dideoxy termination method. pBR325, a 6 kb plasmid, was used as a size control as it was not known to exhibit any conformational changes at low pH. All plasmids, including pBluescript, were prepared by alkaline lysis and cesium chloride buoyant density centrifugation (33). All reagents and enzymes were supplied by Gibco BRL unless otherwise noted.

Generation of plasmid topoisomers

Plasmid topoisomers were generated by incubating 20 μg of supercoiled plasmid DNA with a topoisomerase extract (prepared from CEM cells as described; 34,35) in 400 μl of a buffer containing 10 mM Tris (pH 7.5), 1 mM EDTA, 50 mM NaCl and ethidium bromide in concentrations from 0 to 4.8 μM, for 18 h at 37°C. Individual topoisomer preparations were pooled, ethidium and proteins were removed by three phenol/chloroform extractions and the DNA was ethanol precipitated and dissolved in 10 mM Tris (pH 8.0), 1 mM EDTA, to give a final concentration of 2.5 mg/ml. This created topoisomers with superhelical densities from $\sigma = 0$ to -0.05.

Two-dimensional gel electrophoresis

Mixtures of topoisomers were incubated in electrophoresis buffer (13.5 mM Tris, pH adjusted from 8.0 to 5.0 with the addition of glacial acetic acid) with or without 1 mM magnesium acetate for 60 min at 37°C prior to electrophoresis in the first dimension. Unless otherwise specified, all two-dimensional electrophoreses were performed in 1.25% agarose (20 × 20 cm) at 4°C at 5 V/cm for 17 h. The buffer was circulated vigorously to avoid the development of a pH gradient between the electrodes. The gel was then equilibrated in a second electrophoresis buffer (40 mM Tris, 25 mM sodium acetate, 1 mM EDTA, pH 8.3) and 20 μg/ml (40 mM) chloroquine phosphate for 4 h prior to electrophoresis in the second dimension, which was performed perpendicular to the first dimension. At the completion of electrophoresis the gels were extensively rinsed with distilled water prior to staining with ethidium bromide (0.5 μg/ml).

P1 nuclease reactions

Supercoiled plasmid DNA (1 μg) was digested with 0.22 U of P1 nuclease (Pharmacia Biotech) in 100 μl reactions containing 10 mM sodium acetate (pH 4.6) or 10 mM Tris (pH 6.0 or 7.6), 50 mM NaCl, 10 mM $MgCl_2$, at room temperature for 1 min. DNA was recovered by phenol/chloroform extraction and ethanol precipitation. For mapping experiments, pCW31 DNA was digested with the restriction endonuclease *Eco*RI, *Sph*I, *Xmn*I, *Sca*I or *Bgl*I and then analyzed on a 0.85% agarose gel. pCW2966 was digested with *Ssp*I, *Nde*I, *Xmn*I and *Bgl*I for mapping.

Chemical modification

For sequencing and primer extension, a 15:1 (v/v) mixture of *KlenTaq1* (25 U/μl; Ab Peptides, St Louis, MO) and *Pfu* (2.5 U/μl; Stratagene, La Jolla, CA) polymerases was used (36). In a standard experiment, each sample containing 1 μg of supercoiled pCW2966, at an average superhelical density of $\sigma = -0.05$ was incubated in 40 μl of 50 mM sodium acetate, 1 mM EDTA (pH 5.0). After a 1 h incubation at 37°C, samples were treated with chemicals at room temperature by adding either 1.6 μl of 50% chloroacetaldehyde (CAA) or 2 μl of 20 mM $KMnO_4$. $KMnO_4$ reactions were halted after 5 min by adding 5 μl of 2.5 M β-mercaptoethanol. CAA reactions were terminated after 2.5 h by extracting CAA with diethyl ether. Following ethanol precipitation, modified DNA was digested with *Pvu*II and the reaction mixtures were then extracted with phenol/chloroform. Digestion with *Pvu*II prior to primer extension increases the signal-to-noise ratio in subsequent analysis. After an additional ethanol precipitation, samples were resuspended in 20 μl of TE (10 mM Tris, pH 7.5, 1 mM EDTA) and 5 μl was used for primer extension analysis with standard M13 24mer sequencing primers. Primer extension reactions for M13 forward primer contained 0.025 mM radiolabeled primer, 250 mM dNTPs, 0.5 U of *KlenTaq1* and 0.0035 U of *Pfu* polymerases, 2 M betaine and 9% dimethyl sulfoxide in *KlenTaq1* buffer. Primer extension reactions for M13 reverse primer employed 4-fold higher concentrations of polymerases and a 10-fold excess (relative to the labeled primer) of a 'helper oligonucleotide' (5'-TCCCTCCTCCCTCCTCCCCC) that prevented hairpin formation in the purine-rich template. Primer extension products were resolved on a 7% denaturing polyacrylamide gel in TBE buffer (90 mM Tris–borate, pH 8.3, 2.5 mM EDTA) with sequence markers obtained by dideoxy sequencing of the plasmid using a protocol similar to that described above. Gels were dried and exposed to X-ray film or a PhosphorImager plate for analysis of the radioactivity pattern using ImageQuant software (Molecular Dynamics).

RESULTS

A mirror repeat region from intron 21 forms an H-y3 intramolecular triplex DNA

The long Pu·Py tract found in intron 21 of the PKD1 gene contains imperfect repeats of $(CCTCCCC)_n$ with mirror repeat symmetry (32). Such regions have the potential to form intramolecular triplex DNA structures at low pH (25,37) and have been implicated in mutagenesis (38). To determine whether an intramolecular triple helical DNA structure could form from a mirror repeat in the intron 21 Pu·Py tract, we analyzed plasmid pCW2966 containing a 46 bp perfect mirror repeat from intron 21. Two-dimensional gel electrophoresis of pCW2966 topoisomers demonstrated a transition at pH 5.0 that was absent at pH 7.0 (Fig. 1A and B). The transition occurred at topoisomer –9, corresponding to a superhelical density of $\sigma = -0.0375$. The DNA secondary structure transition removed ~4.8 superhelical turns, consistent with the formation of a triplex within the 46 bp mirror repeat (Fig. 1B). No transitions were observed at pH 5.0 in the parent plasmid, pUC19 (Fig. 1C). The pH-dependent structural transition in plasmid pCW2966 is

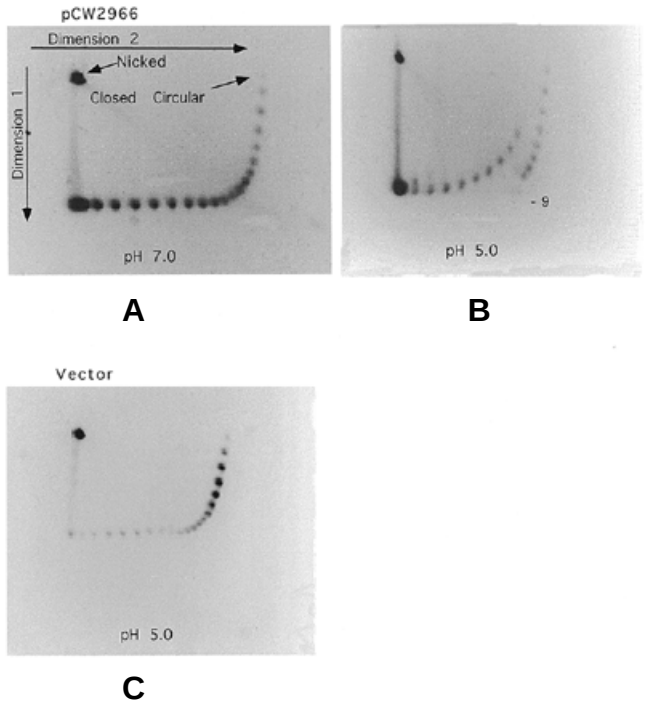

Figure 1. Two-dimensional agarose gel analysis of a DNA secondary structure transition in pCW2966. (**A**) Two-dimensional gel electrophoresis of topoisomers of plasmid pCW2966 containing the 46 bp perfect mirror repeat at neutral pH. The dark 'spot' in the upper left hand corner corresponding to nicked plasmid DNA and the faint spot in the upper right hand corner corresponding to relaxed (closed circular) plasmid are indicated. Each discrete spot, from top to bottom and from right to left, represents a topoisomer with increasing negative supercoils. The first dimension was run from top to bottom while the second dimension was run from left to right after the addition of chloroquine. The gel was run in both dimensions at 5 V/cm for 17 h in 13.5 mM Tris (pH 7.0), 1 mM $MgCl_2$ buffer. (**B**) Two-dimensional agarose gel analysis of pCW2966 at pH 5.0 in the first dimension. The conditions for this gel were identical to those shown in (A) with the exception that the sample was incubated and then run at pH 5.0 in the first dimension. A structural transition occurs at topoisomer –9 due to the relaxation of ~4.8 supercoils. (**C**) Two-dimensional agarose gel analysis of pUC19 (without the Pu·Py mirror repeat) at pH 5.0. These plasmid topoisomers were incubated at pH 5.0 and run at pH 5.0 in the first dimension. No structural transition was apparent.

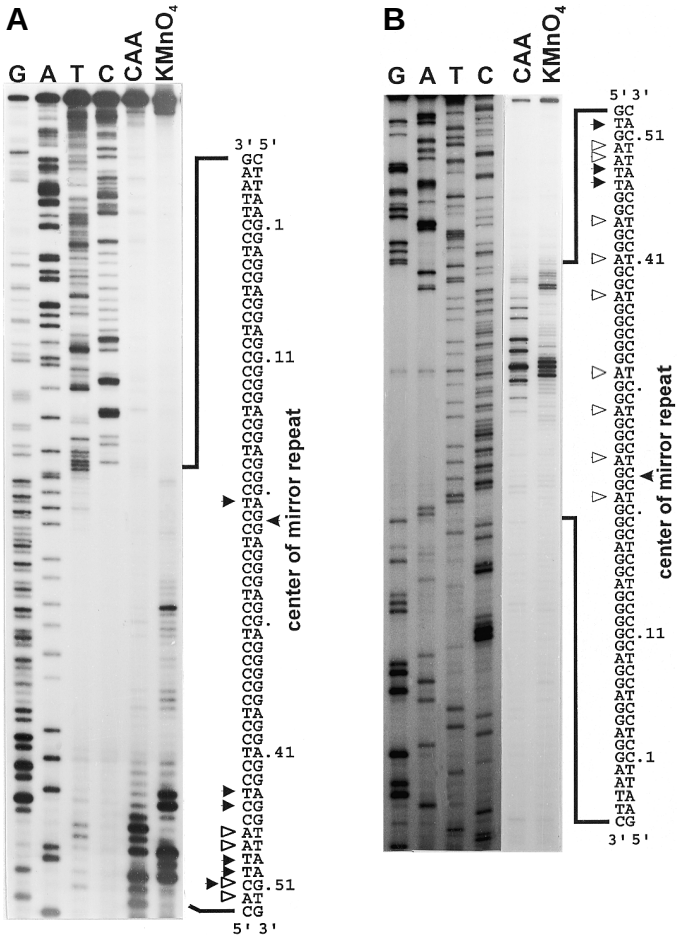

Figure 2. Chemical modification of plasmid pCW2966. (**A**) The modifications on the pyrimidine strand are adjacent to the sequencing ladder. The sequence and chemical modification lanes are marked at the top of the gel. Because dideoxy termination was used to generate the sequencing ladder, it is purine-rich and the modified bases are on the complementary strand. The duplex DNA is illustrated to the right. The modified bases are depicted by arrowheads. Filled arrowheads denote modification by $KMnO_4$, while open arrowheads identify bases modified by CAA. The sequence is numbered beginning at the *Eco*RI site. (**B**) The purine strand chemical modifications are adjacent to the sequencing ladder as labeled at the top of the gel. The duplex sequence is depicted to the right of the gel and the filled arrowheads identify $KMnO_4$-modified bases while open arrowheads identify CAA-modified bases.

consistent with the formation of a protonated triple helical DNA structure.

Chemical probing at pH 5.0 shows the presence of H-DNA. The chemical modification patterns for pCW2966 of native superhelical density at pH 5.0 are shown in Figure 2. CAA, which recognizes unpaired adenines and cytosines and to a lesser extent guanines, and $KMnO_4$, which recognizes unpaired thymines, were used to reveal single-stranded regions. To analyze nucleobases accessible to the chemicals, we used primer extension with thermophilic DNA polymerases that easily work on the pyrimidine template strand of the mirror repeat sequence (Fig. 2A). However, polymerases exhibited multiple pause sites in the 3'-half and middle of the mirror repeat purine template. A combination of changes to the usual linear amplification PCR protocol were employed to overcome this problem (V.Potaman and J.Bissler, unpublished results). These were: (i) using a combination of *KlenTaq1* and *Pfu* polymerases, which are known to perform well on G+C-rich templates; (ii) adding betaine and dimethyl sulfoxide that reduce structural effects of the template on DNA polymerization; (iii) hybridizing a pyrimidine oligonucleotide within the purine tract to prevent its folding into a hairpin (Fig. 2B). Probing of the pyrimidine strand revealed an unpaired thymine at position 22 that likely belongs to a single-stranded loop at the tip of the triple helix and unpaired thymines at positions 49 and 50 at the triplex–duplex boundary (Fig. 2A, lane $KMnO_4$). In addition, adenines at positions 47, 48 and 52 and a cytosine at position 51, also at the triplex–duplex boundary, were reactive to CAA (Fig. 2A, lane CAA). Adenines in the 5'-half of the purine strand (sequence 23–50) were reactive to CAA, indicating that these bases were unpaired (Fig. 2B, lane CAA).

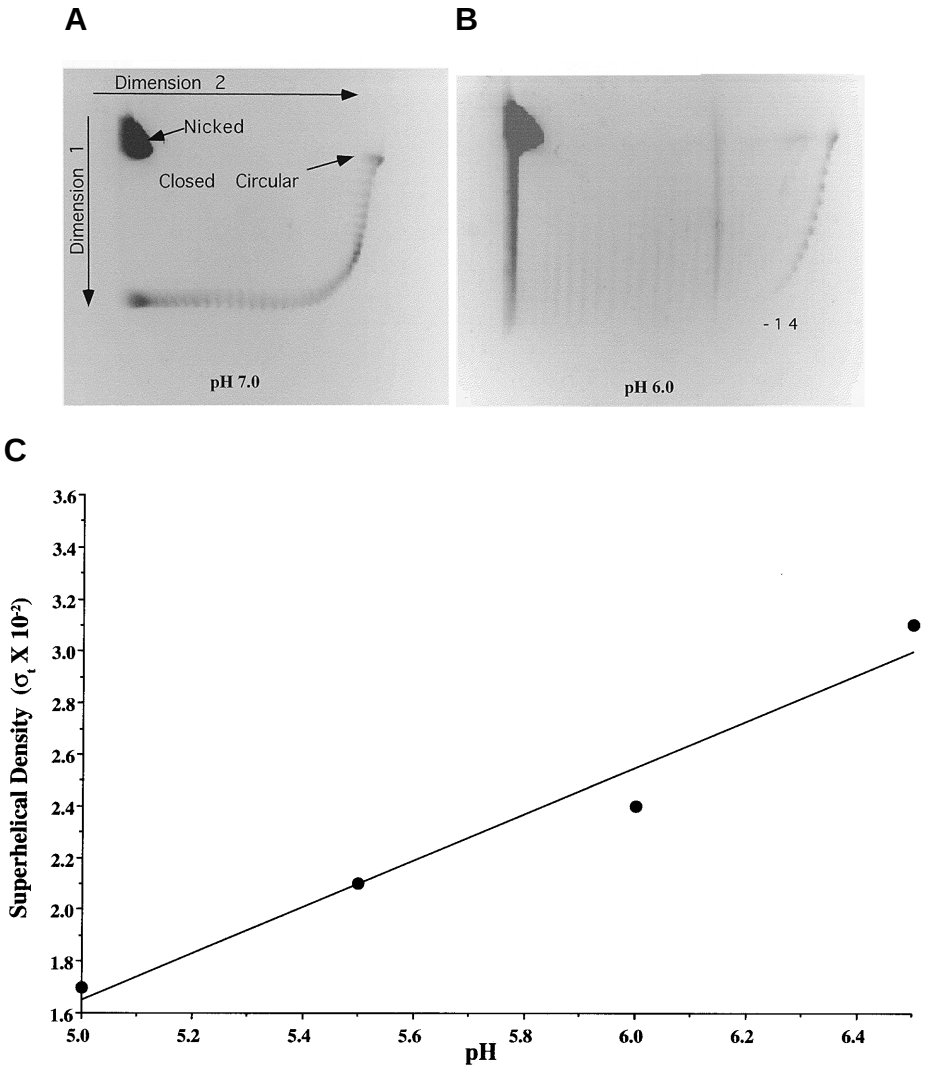

Figure 3. Two-dimensional agarose gel analysis of pCW31. (**A**) The two-dimensional agarose gel was run at pH 7.0 as described in Methods and Materials and in the legend to Figure 1. The positions of the nicked and covalently closed circular DNA topoisomers are indicated. (**B**) The two-dimensional agarose gel shown for pH 6.0 contains the same set of topoisomers as shown in (A) (for electrophoresis at pH 7.0). The series of distinct bands halts at topoisomer number –14, where the pattern becomes complex due to multiple DNA secondary transitions. (**C**) Influence of pH on negative superhelical density at which a structural transition occurs. Individual data points were obtained from analysis of two-dimensional gels as shown in (A) and (B).

$KMnO_4$ also detected unpaired thymines at positions 47, 48 and 52 that abut the 5'-end of the purine sequence (Fig. 2B, lane $KMnO_4$). There are several non-specific polymerase stops in the $KMnO_4$ lane. The triplex structure loop may be tightly packed as several unpaired bases (cytosines at positions 23 and 24 and thymine at position 25) are not accessible to chemicals.

Structural transition(s) within the 2500 bp Pu·Py tract

To test the hypothesis that multiple alternative DNA secondary structures could form in the Pu·Py region of the PKD1 gene, pH-dependent structural transitions in plasmid pCW31 were analyzed by two-dimensional gel electrophoresis. Figure 3 shows two-dimensional gels of pCW31 at pH 7.0 and 6.0 in the presence of 1 mM magnesium. The pH 7.0 gel showed approximately 35 topoisomers of pCW31 in a typical array expected for a plasmid with no DNA secondary structural transition (Fig. 3A). At pH 6.0, structural transitions are evident at topoisomer –14 (σ = –0.024), after which the topoisomers no longer resolve into discrete bands (Fig. 3B). This pattern is expected from the formation of multiple structural transitions involving variable numbers of base pairs resulting in the loss of variable numbers of supercoils, as observed previously with the formation of multiple triplex structures (39,40). Additionally, at pH 5.0 there was a large amount of DNA at the position of relaxed DNA in the second dimension (data not shown). This is likely DNA that was relaxed in the first dimension and continued to be relaxed in the second dimension. It is also possible that the DNA (with large amounts of single-stranded DNA) became nicked prior to electrophoresis in the second

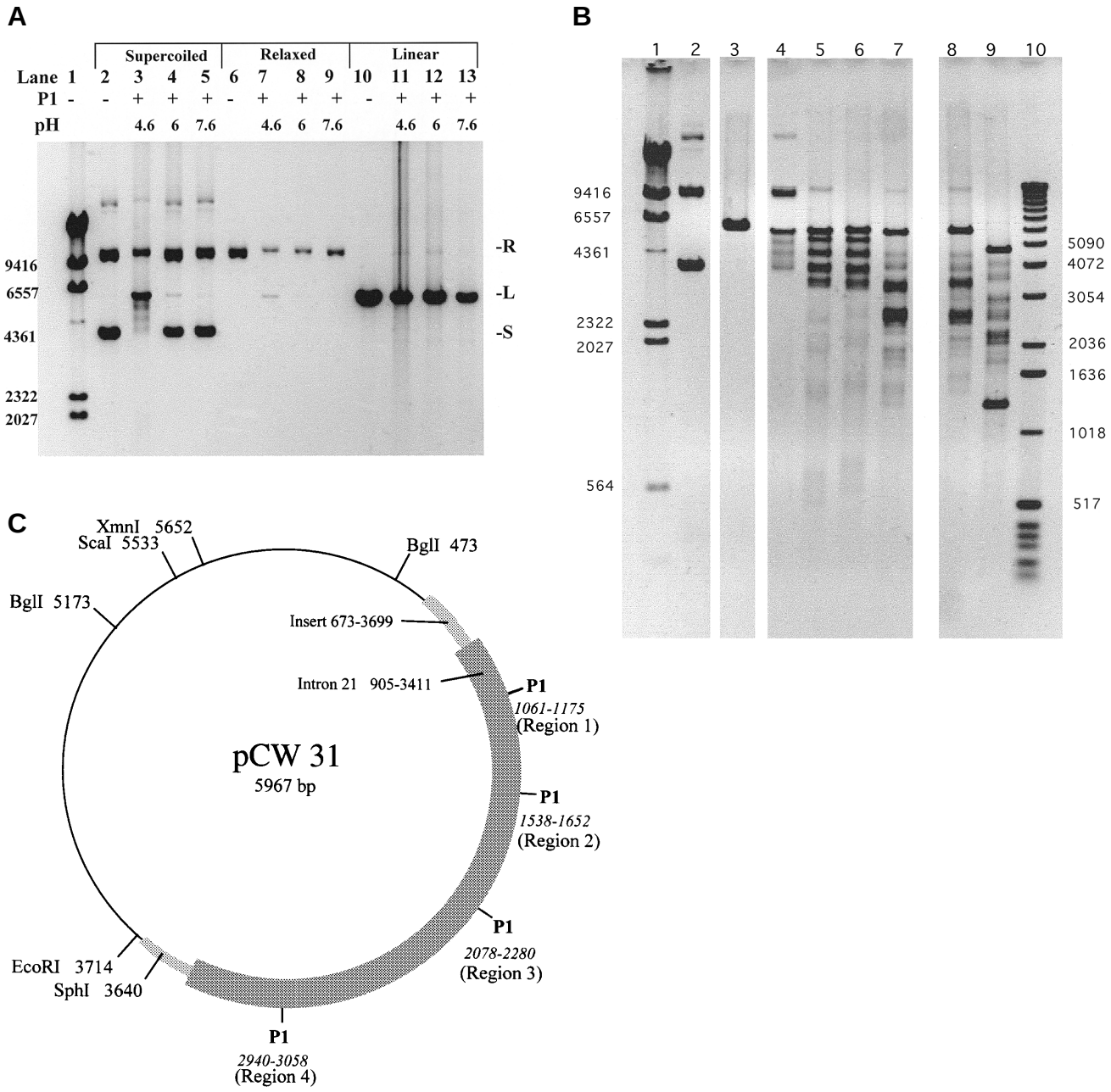

Figure 4. P1 nuclease mapping analysis of DNA secondary structure transitions in pCW31. (**A**) Superhelical density and pH dependence of P1 cutting of pCW31. Lane 1, *Hin*dIII digest of λ phage DNA. Lanes 2–5, supercoiled DNA. Lane 2 shows two predominant bands. The lower band is supercoiled monomer, while the middle band is relaxed monomer and supercoiled dimer plasmids. The upper fainter band is supercoiled trimer. Lanes 6–9, DNA relaxed by treatment with topoisomerase. Lanes 10–13, DNA linearized by treatment with *Eco*RI. Samples in lanes 3–5, 7–9 and 11–13 were treated with P1 nuclease (indicated by +). Samples were treated with P1 nuclease at pH 4.6 (lanes 3, 7 and 11), 6.0 (lanes 4, 8 and 12) or 7.6 (lanes 5, 9 and 13). (**B**) Mapping of P1 nuclease hypersensitive sites in pCW31. Lanes 1 (*Hin*dIII digest of λ phage DNA) and 10 (1 kb ladder) are molecular weight markers. Lane 2, supercoiled pCW31 DNA. Lane 3, pCW31 cut with *Eco*RI. Lane 4, supercoiled pCW31 digested with P1 nuclease. Lanes 5–9, supercoiled pCW31 digested with P1 nuclease and then: *Sph*I (lane 5), fragment sizes 5279, 4530, 3881, 3388, 2465, [2068], 1393, [588]; *Eco*RI (lane 6), fragment sizes 5150, 4486, 3881, 3388, 2561, 2083, 1481, [635]; *Xmn*I (lane 7), fragment sizes 4530, 4007, 3255, [2393], [1853], 1469, 1261; *Sca*I (lane 8), fragment sizes 4400, 3881, 3388, 2961, 2541, [1985], 1562, 1393; *Bgl*I (lane 9), fragment sizes 4072, 3583, 2984, 2580, [2115], 1943, 1807, 1242, 1082. Fragment sizes in brackets represent the means of a series of bands clustered in 150 bp. (**C**) Locations of P1 nuclease hypersensitive sites. The locations of the Pu·Py insert from intron 21 and flanking PKD1 DNA sequences are indicated by dark and lightly shaded areas, respectively. P1 nuclease-sensitive regions 1–4 are indicated.

dimension. These patterns are consistent with a large, variable change in DNA topology.

To further examine the pH dependence of this transition, other two-dimensional gels were analyzed and the superhelical

Figure 5. The H-y3 structure inferred from the chemical modification of the 46 bp perfect mirror repeat from the intron 21 Pu·Py sequence. The 3' pyrimidine end folds back into the major groove. The single-stranded loop and purine stand are susceptible to chemical modification. Closed arrowheads denote $KMnO_4$ modification, while open arrowheads denote CAA modification. The symbols O and + identify T·A and C^+·G Hoogsteen hydrogen bonds respectively while the symbol | indicates Watson–Crick hydrogen bonds.

density at which the transitions were first observed (σ_t) was plotted as a function of pH in Figure 3C. There was a near linear relationship between σ_t and pH.

Mapping P1 nuclease-sensitive sites

Single strand-specific nucleases, P1 and S1, have been used previously for the identification of non-B-DNA stuctures (reviewed in 41). This characteristic can be exploited to locate superhelical density-dependant DNA structures containing single-stranded regions, such as triplex DNA, Z-DNA, cruciforms or other unwound structures. Alternative DNA secondary structure formation in plasmids pCW31, pCW2966 and pUC19 was investigated using P1 nuclease at pH 4.6, 6.0 and 7.6 (42). Plasmids were incubated for 60 min at the designated pH to promote structural transitions and then digested with P1 nuclease at room temperature in 50 mM NaCl and 10 mM $MgCl_2$ to cleave regions of single-stranded DNA. The reaction was halted by two phenol/chloroform extractions and the DNA was precipitated and resuspended in 10 mM Tris, 1 mM EDTA, pH 8.

P1 cutting at sites of non-B-DNA structure typically results in the linearization of supercoiled plasmids. As shown in Figure 4A (lane 3), P1 nuclease digestion of supercoiled pCW31 preincubated at pH 4.6 (conditions that form intramolecular triplex DNA) resulted in the loss of supercoiled DNA, with the concomitant appearance of linear molecules and three other smaller products. This is indicative of the presence of one and, in some fractions of the molecules, multiple non-B-DNA structures. Only minimal P1 linearization was observed in DNA preincubated at pH 6.0, while no digestion was observed in DNA preincubated at pH 7.6 (Fig. 4A, lanes 4 and 5, respectively). Relaxation of superhelical tension by topoisomerase treatment (lanes 6–9) or linearization by *Eco*RI digestion (lanes 10–13) prevented subsequent P1 nuclease cutting. A small amount of P1 cutting was observed in relaxed DNA preincubated at pH 4.6 (lane 7) due to cutting at an A+T-rich region. These results demonstrate that the formation of non-B-DNA structures was pH and superhelical density dependent. A structural transition was observed at pH 6.0 in the two-dimensional agarose gel analysis but not using P1 nuclease. The 10-fold increase in magnesium concentration and presence of NaCl in the P1 buffer may suppress the structural transition in the P1 reactions, while the long duration run combined with the buffer conditions in the two-dimensional gel experiments allowed the transition.

The location of the P1 nuclease cutting sites can be determined by cutting DNA with a set of restriction enzymes and analyzing the fragment lengths. This analysis is shown in Figure 4B. Restriction endonucleases *Eco*RI, *Sph*I, *Xmn*I, *Sca*I and *Bgl*I were used to map the location of the P1-sensitive sites. In a few instances the fragments resolved into broad bands or a series of bands within a size span of ~150 bp. The analysis is consistent with the presence of four P1-sensitive sites, denoted regions 1–4 (Fig. 4C). Several points can be made from this analysis. First, the three products observed in supercoiled DNA treated with P1 (sizes 5000 ± 100, 4500 ± 70 and 3900 ± 100) are consistent with P1 cutting at multiple sites in supercoiled plasmid. Moreover, the probability of cutting at the four sites can be determined from analysis of the intensities of the products shown in Figure 4B (see Discussion).

In an effort to determine if there was a density-dependent hierachy to the nuclease cleavage sites, nine different topoisomer preparations with average superhelical densities varying from σ = –0.0045 to –0.0228 were constructed. P1 nuclease digestion of these lower density preparations resulted in a single 6 kb linearized band, in contrast to the multiple bands observed in plasmids with a superhelical density of σ = –0.05. Fine mapping experiments using plasmids with an average superhelical density of σ = –0.009 and the restriction endonucleases *Eco*RI, *Sph*I, *Xmn*I, *Sca*I and *Bgl*I revealed that all four P1 nuclease sites were present within the population of fragments represented by the single band following P1 digestion. Similar experiments with less supercoiled plasmids resulted in much fainter bands.

DISCUSSION

We studied DNA structural transitions within the 2.5 kb polypurine·polypyrimidine tract from intron 21 of the PKD1 gene. Our results show that multiple pH-dependent structural transitions occur at modest superhelical densities within this sequence. The pH dependence of the observed change in topology suggests that at least some component of the DNA alternative secondary structure formed is protonated, consistent with the formation of either protonated intramolecular triple helical DNA containing G·C^+ Hoogsteen bonds or hairpins containing C·C^+ bonds (43,44).

Of the four P1 nuclease-sensitive sites in plasmid pCW31, site 4 appears to be the strongest cutting site since the three bands observed in the supercoiled DNA sample correspond to cutting at sites 1 and 4, 2 and 4 and 3 and 4, but not 1 and 2 or 2 and 3 (cutting at sites 1 and 3 could not be determined as its product overlaps with that from cutting at sites 3 and 4). The relative band intensities in the *Xmn*I, *Sca*I and *Bgl*I digests suggest that the strength of cutting may be 4 > 3 > 2 > 1. This pattern correlates with the length of the Pu·Py tracts identified in Table 1. Region 4 contains an 87 bp Pu·Py tract composed of two overlapping mirror repeats of 41 and 55 bp. The 46 bp perfect mirror repeat that forms the H-y3 structure is contained within the 55 bp mirror repeat (as discussed below). Region 3 contains a 64 bp Pu·Py tract composed in two overlapping

Table 1. Location of P1 sites and corresponding mirror repeats

	LENGTH (bp)	pCW31	PKD1
REGION 1			
CTCTCCCCTCCCCTCTCC-*CCCCT*-TCTCTCCCCTCCCCTCTC	41	1201-1241	36383-36423
CTCTCCCCTCCCCTCTCC-*CCCT*-TCTCTCCCCTCCCCTCTC	40	1047-1086	36538-36577
CTCTCCCCTCCCCTCTCC-*CCCT*-TCTCTCCCCTCCCCTCTC	40	1178-1217	36407-36446
CTCTCCCCTCCCCCC-*TTCT*-CTCCCCTCCCCTCTC	34	1070-1103	36521-36554
REGION 2			
CCCTCCCCTTCCCCTCCCCTC-*CT*-CTTCCCTCCCCTTCCCCTCCC	44	1528-1571	36053-36096
CCTCCCCTCCTCTTCC-*CTCC*-CCTTCCCCTCCCCTCC	36	1524-1559	36039-36074
REGION 3			
CTCCTCTCCTCCCCTTCTCTCC-*CCT*-CCTCTCCTCCCCTCCTCTCCTC	47	2242-2288	35336-35382
CTCCCCTCCTCTCCTCCCCTC-*CTCTC*-CTCTCCTCCTCTCCTCCCCTC	47	2224-2270	35354-35400
REGION 4			
TCCCCCCTCCTCCCCTCCTC-*C*-CTCCTCCCCTCCTCCCCCCT	41	3009-3049	34575-34615
CCCCCCTCCTCCTCCCCCTCCTCCCT-*CCT*-CCCTCCTCCCCCTCCTCCTCCTCCC	55	2962-3015	34609-34662
OTHER LARGE MIRRORS (>35NT) NOT MAPPING TO P1 SITES			
TTCCCCTCCCCTTCTCTTCTCC-*T*-CCCCTTCTCTTCCCCTCCCCTT	45	1416-1460	36164-36208
CCCCTTCCTCCCCTCC-*TCCC*-CCTCCCCTCCTCCCCC	36	3110-3145	34479-34514
CCCCTCCCCTCCTCCCCTCCC-*CTAGC*-CCTTCCCCTCCTCCCCTCCCC	47	3321-3368	34257-34303

Because of the simple repetitive nature of the sequence, many mirror repeats are possible. The largest repeats corresponding to the P1 digestion sites are mapped. Their location in pCW31 and in the PKD1 gene are listed. The insert in pCW31 is in the reverse orientation, therefore the sites are in ascending order. The corresponding sites in the PKD1 gene are also listed and they are in descending order. The nucleotides in italics separated from the rest of the sequence by dashes are the spacers and the underlined nucleotides are the unpaired sites.

47 bp imperfect mirror repeats. Region 2 contains a 57 bp imperfect mirror repeat with two overlapping 36 and 44 bp mirror repeats. Four 34–41 bp imperfect mirror repeats are found in region 1, which has the weakest P1 cutting site.

The structure of a non-B-DNA structure formed within a 46 bp perfect mirror repeat from region 4 was characterized using two-dimensional agarose gel analysis and chemical probing. This mirror repeat sequence forms an alternative DNA secondary structure that is dependent on both low pH and energy from DNA supercoiling. The relaxation of 4.8 superhelical turns upon structure formation is consitent with the formation of an intramolecular triplex structure within the 46 bp sequence. Moreover, chemical modification of this non-B-DNA structure yields a modification pattern compatible with the formation of a $C^+ \cdot G \cdot C$ triplex DNA structure in which the 3'-half of the pyrimidine strand is involved in Hoogsteen base pairing in the triplex. The H-y3 type intramolecular triplex structure shown in Figure 5 is consistent with the results.

The long Pu·Py tract in the PKD1 gene may be particularly prone to non-B-DNA structure formation, given its length and the large number of potential structures that can from. Unrestrained supercoiling, important for the formation of intramolecular triplex structures, is common in and around many genes in human cells (45,46). Chromatin remodeling or transcription may also create regions of increased DNA supercoiling (47) that promotes intramolecular triplex formation. It is clear that protonated triplex structures can exist *in vivo* (17) and polycations in cells may contribute to the stabilization of protonated H-DNA (48). A pre-formed triple helical DNA structure may present a significant challenge to traversal and faithful replication of polypurine·polypyrimidine sequences by DNA polymerase (30,31). The triple-stranded barrier for DNA polymerase may also form during the process of DNA polymerization (29,49). In fact, in our experiments, we had to develop a novel technical strategy to reduce the formation of triplex DNA at neutral pH in order to allow the polymerase successful passage through the mirror repeat sequence, even at high PCR temperatures in the presence of DNA denaturants. In this case the triple-stranded structure involved two purine strands. It is likely that either one or another type of triple-stranded structure might be involved in mutagenesis of intron 21 of the PKD1 gene, consistent with the earlier observation of a high frequency of mutagenic events, such as deletions, in the polypurine·polypyrimidine sequences (27) and other mutations,

including recombination events, at intermolecular triplex regions (38,50).

Our findings may be important in understanding the molecular pathology of ADPKD caused by mutations in the PKD1 gene. The current hypothesis regarding onset of renal cysts requires a somatic mutation in the functional PKD1 allele (7). Non-B-DNA structures pose a significant challenge for high fidelity DNA replication and are recombination substrates (21). In the PKD1 mutation literature there are reports attributing the cause of the second mutation to somatic mutagenesis (51) and gene conversion (52). We hypothesize that this intron 21 region may participate in the mutagenic mechanism that disables the normal allele in tubular cells leading to cystogenesis.

REFERENCES

1. Gabow,P.A. (1993) *N. Engl. J. Med.*, **329**, 332–342.
2. United States Renal Data System (1996) *Am. J. Kidney Dis*., **28**, S12–S20.
3. European Polycystic Kidney Diseases Consortium (1995) *Hum. Mol. Genet.*, **4**, 575–582.
4. European Polycystic Kidney Diseases Consortium (1995) *Cell*, **81**, 289–298.
5. European Polycystic Kidney Diseases Consortium (1994) *Cell*, **77**, 881–894.
6. Hughes,J., Ward,C.J., Peral,B., Aspinwall,R., Clark,K., San Millan,J.L., Gamble,V. and Harris,P.C. (1995) *Nature Genet.*, **10**, 151–160.
7. Germino,G.G. (1997) *Hosp. Pract.*, **32**, 81–92.
8. Waldherr,R., Zerres,K., Gall,A. and Enders,H. (1989) *Lancet*, **2**, 274–275.
9. Watson,M.L. and Torres,V.E. (eds) (1996) *Polycystic Kidney Disease*. Oxford University Press, New York, NY.
10. Qian,F., Watnick,T.J., Onuchic,L.F. and Germino,G.G. (1996) *Cell*, **87**, 979–987.
11. Brasier,J.L. and Henske,E.P. (1997) *J. Clin. Invest.*, **99**, 194–199.
12. Koptides,M., Constantinides,R., Kyriakides,G., Hadjigavriel,M., Patsalis,P.C., Pierides,A. and Deltas,C.C. (1998) *Hum. Genet.*, **103**, 709–717.
13. Lyamichev,V.I., Mirkin,S.M. and Frank-Kamenetskii,M.D. (1985) *J. Biomol. Struct. Dyn.*, **3**, 327–338.
14. Lee,J.S., Latimer,L.J., Haug,B.L., Pulleyblank,D.E., Skinner,D.M. and Burkholder,G.D. (1989) *Gene*, **82**, 191–199.
15. Kohwi,Y., Malkhosyan,S.R. and Kohwi-Shigematsu,T. (1992) *J. Mol. Biol.*, **223**, 817–822.
16. Kohwi,Y. and Panchenko,Y. (1993) *Genes Dev.*, **7**, 1766–1778.
17. Ussery,D.W. and Sinden,R.R. (1993) *Biochemistry*, **32**, 6206–6213.
18. Collier,D.A., Griffin,J.A. and Wells,R.D. (1988) *J. Biol. Chem*., **263**, 7397–7405.
19. Weinreb,A., Collier,D.A., Birshtein,B.K. and Wells,R.D. (1990) *J. Biol. Chem.*, **265**, 1352–1359.
20. Kim,R.H., Shapiro,H.S., Li,J.J., Wrana,J.L. and Sodek,J. (1994) *Matrix Biol.*, **14**, 31–40.
21. Rooney,S.M. and Moore,P.D. (1995) *Proc. Natl Acad. Sci. USA*, **92**, 2141–2144.
22. Kato,M., Kudoh,J. and Shimizu,N. (1990) *Biochem. J.*, **268**, 175–180.
23. Kato,M. and Shimizu,N. (1992) *J. Biochem.*, **112**, 492–494.
24. Sarkar,P.S. and Brahmachari,S.K. (1992) *Nucleic Acids Res*., **20**, 5713–5718.
25. Firulli,A.B., Maibenco,D.C. and Kinniburgh,A.J. (1994) *Arch. Biochem. Biophys*., **310**, 236–242.
26. Bacolla,A., Ulrich,M.J., Larson,J.E., Ley,T.J. and Wells,R.D. (1995) *J. Biol. Chem.*, **270**, 24556–24563.
27. Jaworski,A., Blaho,J.A., Larson,J.E., Shimizu,M. and Wells,R.D. (1989) *J. Mol. Biol*., **207**, 513–526.
28. Lapidot,A., Baran,N. and Manor,H. (1989) *Nucleic Acids Res.*, **17**, 883–900.
29. Baran,N., Lapidot,A. and Manor,H. (1991) *Proc. Natl Acad. Sci. USA*, **88**, 507–511.
30. Dayn,A., Samadashwily,G.M. and Mirkin,S.M. (1992) *Proc. Natl Acad. Sci. USA*, **89**, 11406–11410.
31. Samadashwily,G.M., Dayn,A. and Mirkin,S. (1993) *EMBO J*., **12**, 4975–4983.
32. Van Raay,T.J., Burns,T.C., Connors,T.D., Petry,L.R., Germino,G.G., Klinger,K.W. and Landes,G.M. (1996) *Microbial Comp. Genomics*, **1**, 317–327.
33. Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd Edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
34. Hancock,R. (1974) *J. Mol. Biol.*, **86**, 649–663.
35. Germond,J.E., Hirt,B., Oudet,P., Gross-Bellark,M. and Chambon,P. (1975) *Proc. Natl Acad. Sci. USA*, **72**, 1843–1847.
36. Barnes,W.M. (1994) *Proc. Natl Acad. Sci. USA*, **91**, 2216–2220.
37. Soyfer,V.N. and Potaman,V.N. (1996) *Triple-Helical Nucleic Acids*, 1st Edn. Springer-Verlag, New York, NY.
38. Wang,G., Seidman,M.M. and Glazer,P.M. (1996) *Science*, **271**, 802–805.
39. Htun,H. and Dahlberg,J.E. (1988) *Science*, **241**, 1791–1796.
40. Htun,H. and Dahlberg,J.E. (1989) *Science*, **243**, 1571–1576.
41. Wells,R.D. (1987) In Guschlbauer,W. and Saenger,W. (eds), *DNA–Ligand Interactions: From Drugs to Proteins*. Plenum Publishing, New York, NY.
42. Collier,D.A. and Wells,R.D. (1990) *J. Biol. Chem*., **265**, 10652–10658.
43. Edwards,E.L., Patrick,M.H., Ratliff,R.L. and Gray,D.M. (1990) *Biochemistry*, **29**, 828–836.
44. Catasti,P., Chen,X., Deaven,L.L., Moyzis,R.K., Bradbury,E.M. and Gupta,G. (1997) *J. Mol. Biol.*, **272**, 369–382.
45. Kramer,P.R. and Sinden,R.R. (1997) *Biochemistry*, **36**, 3151–3158.
46. Ljungman,M. and Hanawalt,P.C. (1992) *Proc. Natl Acad. Sci. USA*, **89**, 6055–6059.
47. Liu,L.F. and Wang,J.C. (1987) *Proc. Natl Acad. Sci. USA*, **84**, 7024–7027.
48. Potaman,V.N. and Sinden,R.R. (1998) *Biochemistry*, **37**, 12952–12961.
49. Samadashwily,G.M., Dayn,A. and Mirkin,S.M. (1993) *EMBO J.*, **12**, 4975–4983.
50. Faruqi,A.F., Seidman,M.M., Segal,D.J., Carroll,D. and Glazer,P.M. (1996) *Mol. Cell. Biol*., **16**, 6820–6828.
51. Watnick,T.J., Torres,V.E., Gandolph,M.A., Qian,F., Onuchic,L.F., Klinger,K.W., Landes,G. and Germino,G.G. (1998) *Mol. Cell*, **2**, 247–251.
52. Watnick,T.J., Gandolph,M.A., Weber,H., Neumann,H.P. and Germino,G.G. (1998) *Hum. Mol. Genet.*, **7**, 1239–1243.