Published online February 27, 2004

1460–1468 *Nucleic Acids Research, 2004, Vol. 32, No. 4*
DOI: 10.1093/nar/gkh312

PKD1 intron 21: triplex DNA formation and effect on replication

Hiren P. Patel, Lu Lu, Richard T. Blaszak[1] and John J. Bissler*

Division of Nephrology and Hypertension, Cincinnati Children's Hospital Medical Center, Cincinnati, OH 45229-3039, USA and [1]Division of Pediatric Nephrology, Arkansas Children's Hospital, Little Rock, AR 72202-3591, USA

Received November 8, 2003; Revised January 16, 2004; Accepted February 5, 2004

ABSTRACT

Although autosomal dominant polycystic kidney disease is transmitted in an autosomal dominant fashion, there is evidence that the pathophysiology of cystogenesis involves a second hit somatic mutation superimposed upon the inherited germline mutation within the renal tubule cells. The polypurine·polypyrimidine (Pu·Py) tract of *PKD1* intron 21 may play a role in promoting somatic mutations. To better characterize this tract and to evaluate its potential to participate in mutagenesis, we investigated the thermodynamics of intramolecular triplex formation by 15 Pu·Py mirror repeat tracts from *PKD1* intron 21 by 2D gel electrophoresis. We demonstrate that intramolecular triplexes form with modest superhelical tensions for all the tracts examined. Primer extension studies demonstrated significant polymerase arrest within the Pu·Py tracts in one direction of replication only. We found correlation between polymerization arrest and both the potential length of the triplex and superhelical tension of intramolecular triplex formation. The presence of a Pu·Py tract also led to a replication blockade and double-strand breakage using an SV40 *in vitro* replication assay with HeLa cell extracts. During DNA replication, the G-rich template of the PKD1 Pu·Py tracts may form a triplex structure with the nascent strand, thereby blocking replication and potentially leading to recombination and mutation.

INTRODUCTION

Autosomal dominant polycystic kidney disease (ADPKD) is one of the most common genetic diseases worldwide (1) and is responsible for up to 10% of cases of end-stage renal disease in the United States (2). However, the expressivity of disease is remarkably variable both within the affected population as a whole and within individual families who share the same germline mutation. In a monozygotic twin study, Levy *et al*. studied 20 pairs of monozygotic twins and concluded that the disease course in the twin pairs was as different as if they were not related (3). Therefore, additional factors beyond the inheritance of a germline mutation must modify the phenotypic expression of the disease.

More than 85% of the cases of ADPKD are associated with a mutation in the *PKD1* gene. Despite sharing the same genetic information, only a minority of the tubules develop into cysts (4). One explanation for this observation could be that cystogenesis requires an initiating event, such as a somatic mutation in the *PKD1* allele from the non-affected parent (5,6). Having lost the normal copy of the *PKD1* gene in the renal tubular epithelial cell, cysts develop and progress. This concept of a somatic mutation superimposed on a heterozygous germline mutation has been referred to as the 'two-hit hypothesis' with the first 'hit' being the inherited germline mutation and the second being the acquired somatic mutation, resulting in loss of heterozygosity (LOH) (7,8). Such a model, initially proposed by Knudson to explain carcinogenesis (9), has been proposed for other autosomal dominant diseases (10–14). The two-hit model in ADPKD is supported by studies demonstrating LOH at *PKD1* within a subset of cysts from diseased individuals (5,15–17).

Based on the two-hit hypothesis, the number of cysts found in most patients with ADPKD would imply a higher spontaneous mutation rate within *PKD1* than would be expected. While previous studies have suggested that renal cells have a greater mutation rate than cells elsewhere (18,19), another explanation would be the presence of sequence motifs within the *PKD1* gene that facilitate somatic mutation. One candidate as a mutation-promoting sequence is the polypurine·polypyrimidine (Pu·Py) tract located in *PKD1* intron 21. Pu·Py mirror repeat sequences are capable of forming intramolecular DNA triplexes and are often found at or near sites of increased recombination frequency (20–22). The *PKD1* intron 21 Pu·Py tract is 2.5 kb in length, contains 65% cytosine and 32% thymidine (97% pyrimidine) on the coding strand, and is one of the largest intragenic Pu·Py tracts in the human genome. There are 23 mirror repeat sequences within the entire Pu·Py tract (23).

We have previously used primer extension reactions following chloroacetaldehyde and potassium permanganate modification to establish that a mirror repeat Pu·Py tract from the *PKD1* gene can form triplex DNA (24). Using this same tract, Tiner *et al*. imaged the intramolecular triplex structure using atomic force microscopy (25). Triplex formation has

*To whom correspondence should be addressed. Tel: +1 513 636–1201; Fax: +1 513 636–7407; Email: john.bissler@cchmc.org

been shown to present a potent block for DNA replication (26–28). However, the features of Pu·Py tracts which favor the inhibition of replication are less clear.

To further evaluate the potential of the *PKD1* intron 21 to participate in DNA-directed mutagenesis, we present the thermodynamic characterization of the Pu·Py sequences within the P1 nuclease-sensitive regions that engage in triplex DNA formation. To begin to understand the characteristics of Pu·Py tracts which may be used to predict interference with DNA replication we evaluated the effect of those sequences on DNA polymerization. We demonstrate evidence that *PKD1* Pu·Py tracts significantly interfere with DNA replication and that there is correlation between replication blockade and both the potential length of triplex and the superhelical tension of intramolecular triplex formation.

MATERIALS AND METHODS

Plasmids and plasmid topoisomers

There are ten individual Pu·Py mirror repeat tracts within the four P1 nuclease-sensitive regions of *PKD1* intron 21 which make up five overlapping pairs. We cloned the ten individual and five overlapping sequences (Fig. 1) into the EcoRI site of pUC19. The constructs were prepared and verified as previously described (24). All plasmids were prepared by alkaline lysis and cesium chloride buoyant density centrifugation (29). The correct size of the insert after cloning was confirmed by restriction endonuclease digestion of the plasmid by EcoRI, followed by T4 polynucleotide kinase labeling with [γ-^{32}P]ATP (Perkin Elmer Life Sciences Inc, Boston, MA, USA) and electrophoretic separation in 5% acrylamide (see Supplementary Fig. S1). pBluescript (a pUC-based cloning vector) was used as a control. Like other pUC-based vectors, it does not exhibit significant conformational changes at low pH. Plasmid topoisomers were generated as previously described (24). All reagents and enzymes were supplied by Gibco (Life Technologies, Rockville, MD, USA) unless otherwise noted.

Two-dimensional gel electrophoresis

Topoisomer mixtures were incubated in electrophoresis buffer (13.5 mM Tris pH 5.0 with 1 mM magnesium acetate) for 60 min at 37°C prior to electrophoresis in the first dimension. All two-dimensional (2D) electrophoreses were performed in 1.25% agarose (20 cm × 20 cm) at 4°C using 5 V/cm for 17 h. The buffer was circulated to avoid the development of a pH gradient. The gel was equilibrated in a second electrophoresis buffer (40 mM Tris, 25 mM sodium acetate, 1 mM EDTA pH 8.3 and 40 mM chloroquine phosphate) for 4 h prior to the second electrophoresis perpendicular to the first dimension. At the completion of electrophoresis the gels were rinsed with distilled water and stained with ethidium bromide (0.5 μg/ml). The amount of superhelical tension required for, and absorbed by, triplex formation can be determined by counting the writhe (number of superhelical turns) at which structural transition occurs using 2D gel analysis. The superhelical tension σ required for structural transition of each plasmid was calculated using the formula:

$$\sigma = \frac{-10.5\omega}{N} \quad \mathbf{1}$$

where ω = the writhe at which structural transition occurred and N = the total number of base pairs in the plasmid. The superhelical tension can be converted to Gibbs free energies of formation and absorption by the formulae:

$$G_{formation}(J) = \frac{1100RT\omega^2}{N} \quad \mathbf{2}$$

$$G_{absorption}(J) = \frac{1100RT(\Delta\omega)^2}{N} \quad \mathbf{3}$$

where R = gas constant (8.31441 J K^{-1} mol^{-1}), T = temperature (Kelvin) and $\Delta\omega$ = number of writhes absorbed by triplex formation.

Primer extension studies with purified polymerase

Template/primer preparation. One microgram of each of the 15 plasmids was digested with 10 units PvuII in a 10 μl reaction for 2 h at 37°C. pUC19 was used as the control for the primer extensions. pUC19 has two PvuII sites, one on either side of the insert. Oligonucleotide primers flanking the EcoRI site of pUC19 within the PvuII sites were synthesized at the University of Cincinnati DNA core facility. One microgram of each primer forward (5′-TGGCGAAAGGGGGATGTG-3′) and reverse (5′-AGCGGATAACAATTTCACACAGGA-3′) were end-labeled with [γ-^{32}P]ATP (Perkin Elmer Life Sciences Inc, Boston, MA, USA) using 10 units T4 polynucleotide kinase in a 25 μl reaction.

Primer extension reaction. A pair of primer extension reactions was run for each plasmid, one for the cytosine-rich template and one for the guanine-rich template. One microliter of labeled primer was mixed with 1 μl of the plasmid digestion product along with 5 units of Taq polymerase, 0.5 mM dNTP, extension buffer (20 mM Tris–HCl pH 8.4, 50 mM KCl) and 3 mM $MgCl_2$. The ratio of primer to template was 100:1. Primer extension conditions were identical for all samples: initial template denaturation at 94°C for 3 min; 30 cycles of 94°C for 1 min, 55°C for 2 min and 72°C for 2 min; final extension at 72°C for 10 min. The primer extension reaction products were mixed with formamide as a denaturant and bromophenol blue as a marker, heated to 94°C for 3 min, snap cooled and resolved on an 8% denaturing polyacrylamide gel. For size determination, a 25 bp molecular weight marker was used. The gel was dried at 80°C for 75 min, exposed to a PhosphorScreen (Molecular Dynamics, Sunnyvale, CA, USA) overnight, and imaged using the Storm 860 (Molecular Dynamics) PhosphorImager. Densitometric analysis using ImageQuant (Molecular Dynamics) software was used to normalize the amount of polymerase blockade by dividing the area of arrested products by the total area of both arrested and full-length products for each of the plasmid/primer combinations. Comparisons done in this manner minimize differences caused by subtle changes in the amount of template. The primer extension reactions were performed

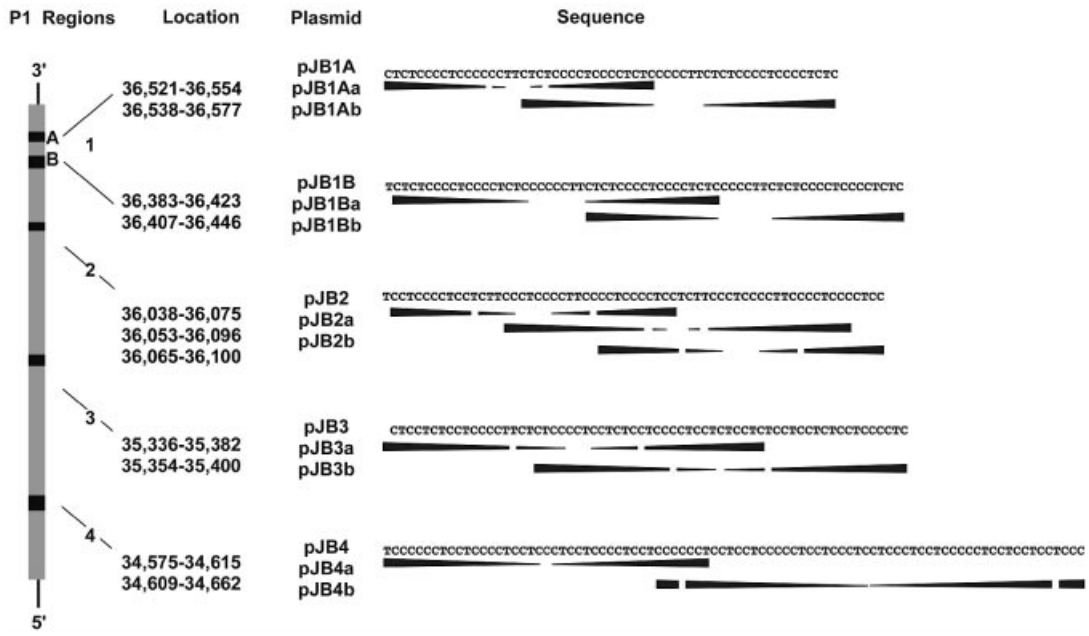

Figure 1. Pu·Py insert sequences. The gray ideogram on the left represents the Pu·Py tract in *PKD1* intron 21. The black bars identify the P1 nuclease-sensitive sites previously described (24). The regions were numbered 1 through 4. There were two regions, A and B, in the first P1-sensitive region that contained mirror repeat sequences capable of triplex formation. The locations in the *PKD1* gene are listed, as is the plasmid name containing the sequence. The sequences are listed and the inverted arrowheads below the sequence represent the mirror repeat sequences. Mismatches within the mirror repeat are identified by a gap within the arrowhead, while the gap between the arrowheads identifies the spacer between the mirror repeats. pJB2a is identical to pJB2c except that pJB2a contains an additional base at the very 5′ and 3′ ends, therefore only pJB2a was cloned.

three times and the average amount of polymerase arrest as a percentage of total product was used for comparisons.

Correlation between polymerase arrest and the following factors was examined: superhelical tension required for intramolecular triplex formation, number of superhelical turns unwound by triplex formation, potential number of base pairs participating in triplex formation, percentage of mirror symmetry, amount of G·C base pairs within the mirror segment of the tract, amount of A·T base pairs within the center of the tract, and length separating the mirror repeat halves. After examining the scatter plots to rule out a nonlinear relationship, the correlation between the above factors and amount of polymerase blockade was calculated using the Spearman rank correlation coefficient.

Primer extensions using a human replication system

We used the clone pJB4 containing two overlapping mirror repeats, including the longest mirror repeat in the entire *PKD1* Pu·Py tract that also has been shown to form triplex structures at physiologic superhelical tensions (24), to generate single-stranded templates. Using the very high fidelity polymerase system described by Barnes (30), the radiolabeled primer was annealed 5′ to the cloning site and this annealed product was incubated with dNTPs and replication-competent extract containing human SSB called replication protein A for 60 min at 37°C. The DNA was phenol/chloroform purified and resolved on a sequencing gel.

Cell culture, extract preparation and *in vitro* replication assay

Mammalian cells (HeLa) were grown in Dulbecco's minimal essential medium at 37°C and 5% CO_2. Cell-free extracts for DNA replication *in vitro* were prepared by the method of Li and Kelly (31) .

DNA replication reactions were carried out (in 30 µl) using 50–60 µg of HeLa cell extract protein (32–34); 20 ng of DNA template; 30 mM HEPES pH 7.5; 7 mM $MgCl_2$; 0.5 mM dithiothreitol; 100 µM each dATP, dGTP and dTTP; 50 µM dCTP; 10 µCi of [α-^{32}P]dCTP; 200 µM each GTP, UTP, and CTP; 4 mM ATP; 40 mM creatine phosphate; 10 µg of creatine kinase; and 1 µg of SV40 large T antigen (Molecular Biology Resources, Madison, WI). Reactions were incubated for 4 h at 37°C, and the reaction products were isolated and purified as described previously (35). Replication products were separated by electrophoresis on a 1% agarose gel containing 5 µg/ml ethidium bromide. Replicated DNA was visualized by autoradiography of the dried gel.

Replication shuttle vectors

This assay uses the viral large T antigen protein to initiate eukaryotic DNA replication. The cloning vector pZ189 contains the SV40 virus replication origin. Large T antigen initiates replication by targeting the origin and melting the duplex DNA so that cell extract proteins can load onto the plasmid and replicate the DNA. We cloned the Pu·Py tract used in the primer extension reactions above into the EcoRI site of pZ189. These plasmids were purified by isopycnic banding and used in replication assays as previously described (33,36,37).

RESULTS

To better understand how the Pu·Py tracts in *PKD1* and more centromeric homologous genes could facilitate mutagenesis,

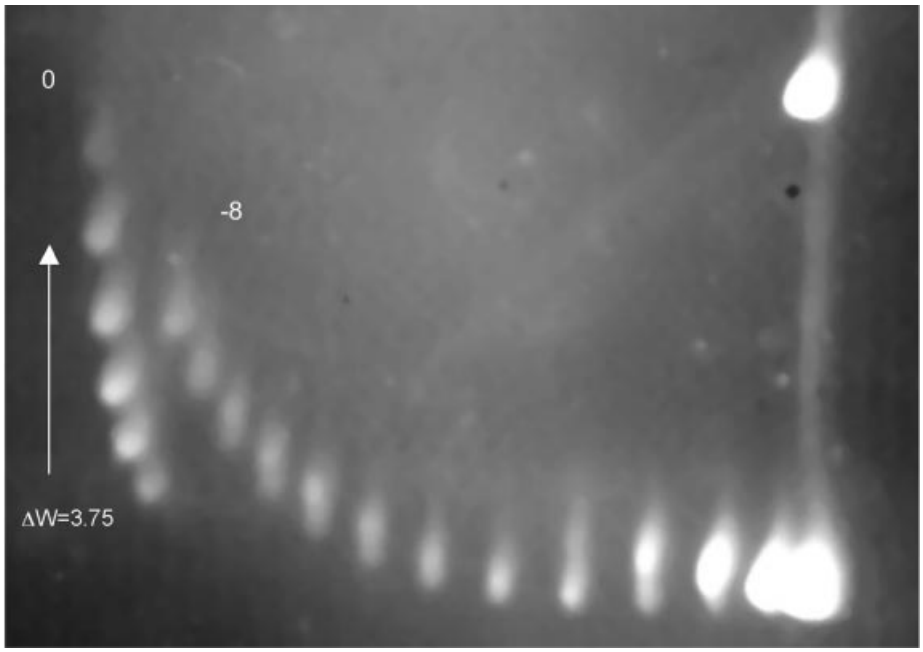

Figure 2. Two dimensional gel of pJB3b. A structural transition consistent with intramolecular triplex formation is demonstrated at writhe –8 (ω = –8). The change in writhe is 3.75 (Δω = 3.75).

Table 1. Spearman rank correlations between factors affecting intramolecular triplex formation and superhelical tension (sigma)

Factor affecting intramolecular triplex formation	r_s	*P*-value
Potential number of bp involved in triplex formation	–0.59	<0.05
Percentage mirror symmetry of tract	–0.07	NS
Amount of G·C bp within mirror segment of tract	0.20	NS
Number of turns unwound by triplex formation	–0.32	NS
Length of centre separating the mirror halves of tract	0.61	$0.05 < P < 0.10$
Amount of A·T bp within the centre of the tract	0.38	NS

The superhelical tension and number of superhelical turns unwound by triplex formation data were obtained from the thermodynamic studies. NS = not significant.

we thermodynamically characterized the structures formed by the mirror repeat sequences. Furthermore, because replication-blocking lesions can lead to recombinational repair, we examined the effects of the Pu·Py tracts on DNA replication.

Thermodynamic characterization of intramolecular triplex formation

Two dimensional gel electrophoresis was used to thermodynamically characterize the triplex structures that were formed by the *PKD1* Pu·Py tract sequences (Fig. 2). Each plasmid exhibited a structural transition. Supplementary Tables S1 and S2 list the thermodynamic data from these experiments. The superhelical tension at which the first structural transition occurred ranged from –0.015 to –0.042 (mean –0.030, median –0.031). Using equation **2**, this corresponds to the energy of formation ranging from 16.5 to 125.8 kJ (mean 64.5 kJ, median 66.5 kJ). Plasmids containing overlapping Pu·Py mirror repeats required less superhelical tension and energy to form intramolecular triplexes than plasmids containing an individual mirror repeat sequence (–0.021 versus –0.034 and 30.7 versus 81.5 kJ, $P < 0.001$ using Mann-Whitney test for both parameters). The number of superhelical turns unwound by the first alternative DNA structure formation ranged from 0.5 to 5.5 (mean 3.1, median 3.5), corresponding to the energy of absorption ranging from 0.26 to 31.19 kJ (mean 12.31 kJ, median 12.62 kJ) using equation **3**. A number of factors may influence the formation of intramolecular triplexes (38). The correlation coefficients between these factors and superhelical tension (σ) are listed in Table 1. Potential length of the triplex was the only factor with a statistically significant correlation.

Primer extension

To begin to understand the effects of the *PKD1* mirror repeat sequences on DNA replication, we used a primer extension assay. In order to compare the DNA structural blockade to an internal control, we cut the plasmid with PvuII (Fig. 3A). If the DNA polymerase were to completely replicate the template using the forward primer, a full-length band of 323 bp plus the insert size (ranging from 34–88 bp) would be predicted

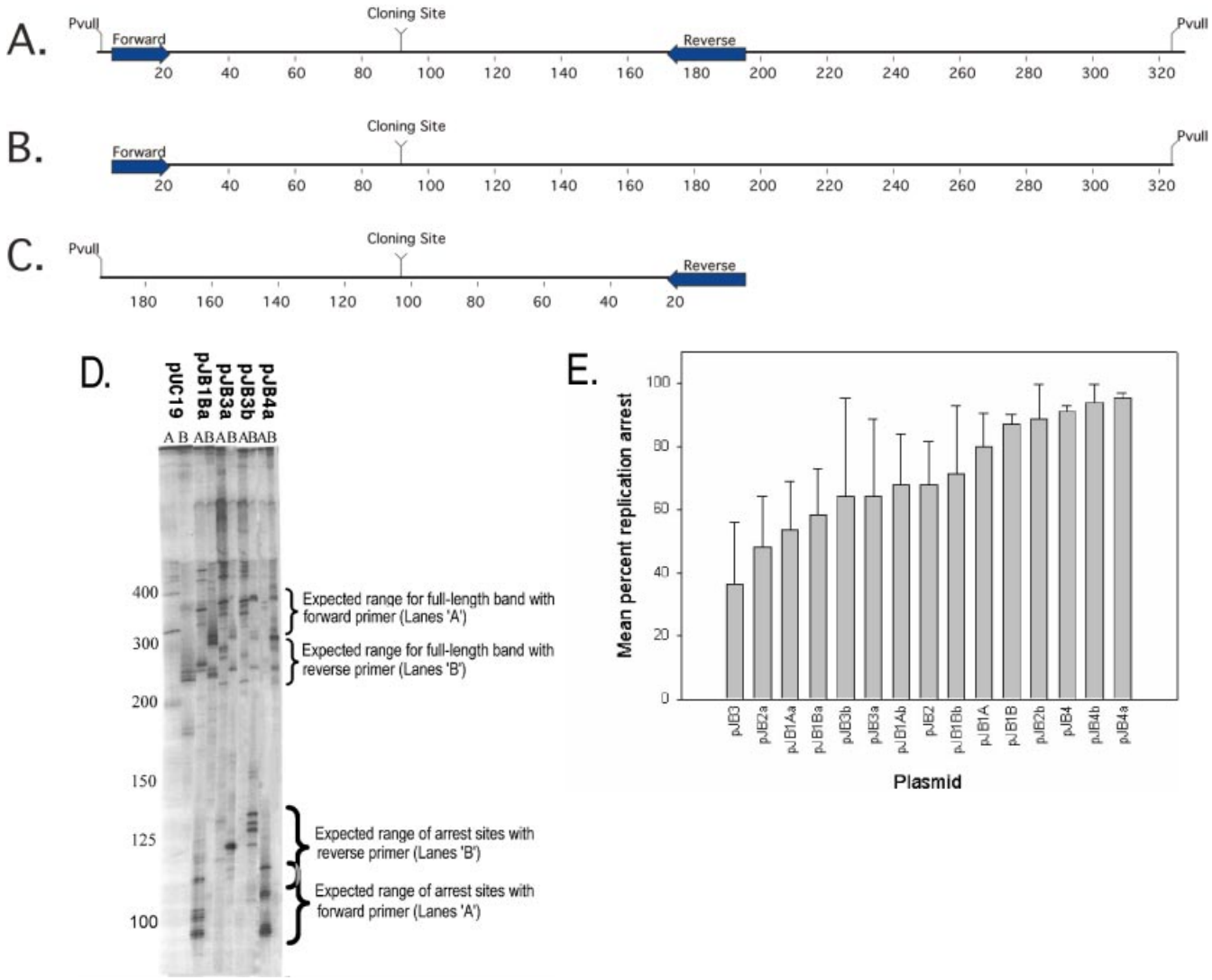

Figure 3. Replication blockade using Taq polymerase. (**A**–**C**) Location of pUC19 primer sites and sequence cloning sites. The Pu·Py tracts were cloned into the EcoRI site. The plasmid was cut with PvuII to generate the templates used in the primer extension reactions. The bars indicate the locations of the primers used. The numbers refer to the base pair position on pUC19. (**D**) Polyacrylamide gel showing primer extension products. The plasmids used to make the single-stranded template are listed at the top. The forward and reverse primers were used in lanes 'A' and 'B' respectively. The expected ranges to find full-length product and product arrested within the Pu·Py insert are indicated on the right side. The numbers on the left indicate the size of the products by the number of base pairs. With templates generated from pJB1Ba and pJB4a, replication arrest occurred in the 'A' lane only. For templates generated from pJB3a and pJB3b, replication arrest occurred in lane 'B' only. Similar results were found for all the other plasmids tested (the lanes for the plasmids not shown have been edited out of this figure). No replication blockade was observed in the templates generated from the control plasmid pUC19. (**E**) Histogram of premature polymerase arrest during replication. *PKD1* intron 21 Pu·Py mirror repeat tracts are ordered on the abscissa by their ability to block replication. The mean amount of premature replication arrest from three experiments is represented by the bar, and standard deviation is depicted by the line.

(Fig. 3B). If the polymerase were to completely replicate the template using the reverse primer, the full-length band would be predicted to be 192 bp plus the insert size (Fig. 3C). Bands corresponding to the full-length product were observed in all lanes in variable amounts. If the DNA polymerase were to arrest at the center of the Pu·Py insert site using the forward primer, a band of ~110–136 bp would be expected, depending on the insert size. With the reverse primer, the expected arrested bands would be ~120–150 bp. As seen in Figure 3D, bands corresponding to polymerase arrest within the Pu·Py tract were observed in either one or the other lane for each insert, but never in both lanes for any specific insert. This is consistent with polymerase being blocked only on the polyguanine template (26,27). Depending on the orientation with which the sequence was cloned into pUC19, the primer extension template could be G-rich using either the forward or reverse primer, explaining why premature polymerase arrest bands were observed in either the 'A' or 'B' lane. The pausing is not seen on the Py-rich template strand because the pH of the extension reaction buffer is nonpermissive. A summary of the densitometric analyses listing the average amount of replication arrest as a percent of total product of the three primer extension reactions is presented in Figure 3E, indicating the amount and range of premature polymerase arrest.

Table 2. Spearman rank correlations between factors affecting intramolecular triplex formation and the percent of polymerization arrest observed

Factor affecting intramolecular triplex formation	r_s	*P*-value
Superhelical tension required for triplex formation	–0.53	<0.05
Potential number of bp involved in triplex formation	0.59	<0.05
Percentage mirror symmetry of tract	0.51	$0.05 < P < 0.10$
Amount of G·C bp within mirror segment of tract	0.47	$0.05 < P < 0.10$
Number of turns unwound by triplex formation	0.18	NS
Length of centre separating the mirror halves of tract	–0.17	NS
Amount of A·T bp within the centre of tract	–0.24	NS

The superhelical tension and number of superhelical turns unwound by triplex formation data were obtained from the thermodynamic studies. NS = not significant.

Analysis of factors affecting intramolecular triplex formation and replication arrest

To determine whether intramolecular triplex formation could be used to predict replication blockade, correlation coefficients between specific factors influencing intramolecular triplex formation (38) and observed replication blockade were calculated and the results are listed in Table 2. After removing the outlying value associated with pJB3 (which also had the greatest relative variability in amount of polymerase arrest), the Spearman rank correlation coefficient (r_s) between these superhelical tension and percent replication blockade was –0.53 ($P < 0.05$) (see Supplementary Fig. S2). The correlation between the potential number of base pairs involved in triplex formation of the individual mirror repeats and the percent polymerase arrest was also statistically significant ($r_s = 0.59$, $P < 0.05$) (see Fig. S3). The percentage of mirror symmetry (i.e. the percentage of nucleotide pairs +n and –n bases around the center that are identical) and the percentage of G·C pairs within the mirror repeat segment involved in triplex formation both approached but did not achieve statistical significance at the 0.05 level. The correlation between replication blockade and number of superhelical turns unwound by triplex formation, center length and amount of A·T base pairs in the center were not significant.

Primer extension using a human replication system

While purified polymerase systems are helpful, they do not adequately reflect the *in vivo* replication process. To better understand the effects of the *PKD1* mirror repeat sequences on mammalian replication, we used replication-competent extracts in primer extension experiments (31). The *PKD1* Pu·Py tract demonstrated significant replication fork blockade (Fig. 4A). The magnitude of the replication blockade was similar to that induced by DNA damage from chemical damage (data not shown). The Pu·Py tract from the *PKD1* gene revealed an intense arrest site on the gel precisely halfway through a mirror repeat in the tract. Like the purified polymerase studies, the replication blockade occurred only on the G-rich template.

In vitro replication in an SV40 system

To detect possible effects of triplex forming sequences on replication, the 88 bp fragment from pJB4 was cloned into the shuttle vector pZ189 to create pZ4.1 (Fig. 4B). These plasmids were then used in an *in vitro* replication assay (Fig. 4C) (33,36,37). As predicted, very little replication took place without the addition of large T antigen in either plasmid. With the addition of this protein, intense bands for both supercoiled and open circular (form II) plasmids were visualized. The addition of the 88 bp fragment from pJB4 resulted in significant changes in the replication reaction products. Linearized molecules (form III) and replication intermediates, felt to be late Cairns structures, were also identified and accounted for ~25% of the resolved replication products. The presence of linearized molecules supports a role for double-strand breakage during the replication process, and the additional replication intermediate band support replication blockade. Given that the only difference in the two plasmids was the presence of the 88 bp Pu·Py tract, it is likely that the replication blockade and double-strand breakage can be attributed to this tract. These findings support the replication blocking ability of this sequence identified in the purified polymerase studies.

DISCUSSION

The biological effects of Pu·Py tracts remain to be fully elucidated. The studies presented here demonstrate that mirror repeat Pu·Py tracts from *PKD1* intron 21 form DNA intramolecular triplex structures under modest energies and inhibit replication. Bacolla *et al*. found that nonB-DNA structures formed by the *PKD1* Pu·Py tract inhibited replication in *Escherichia coli* (39). We demonstrate that these tracts inhibit replication in eukaryotic replication systems as well. Such interference with DNA replication can lead to mutation (40).

We thermodynamically characterized intramolecular triplex formation by the Pu·Py mirror repeats found at the P1 nuclease-sensitive regions of *PKD1* intron 21. Blaszak (24) found that pJB4b formed a triplex at $\sigma = -0.0375$. pJB4b is the longest individual Pu·Py mirror repeat from the region of intron 21 examined. We found that shorter Pu·Py mirror repeats from this region also form triplexes at modest superhelical tension. Sequences containing overlapping Pu·Py tracts required significantly less energy to form triplexes than individual Pu·Py tracts. This observation may indicate that the longer repetitive tracts undergo slipped mispairing more frequently, lowering the energy required to melt the sequence and making transition to a triplex structure more likely.

Although the physiologic relevance of the ability of *PKD1* Pu·Py sequences to adopt an alternative secondary structure

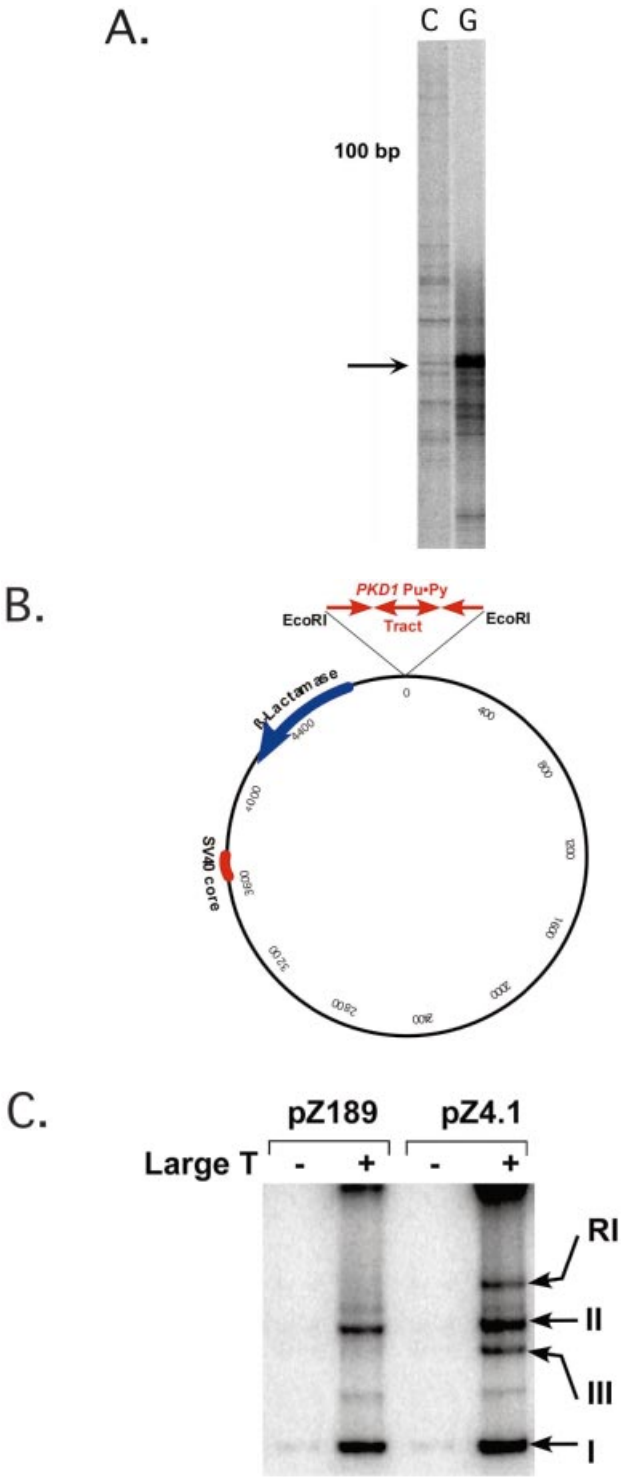

Figure 4. Replication blockade using human replication systems. (**A**) Primer extension reactions using human replication-competent HeLa cell extracts. Replication on the cytosine-rich template (C) is unencumbered, while replication on the guanine-rich template (G) exhibits replication pausing (arrow). Full-length bands not shown. (**B**) Shuttle vectors used for *in vitro* replication. pZ189 contains the β-lactamase gene, and the SV40 origin of replication. pZ4.1 is pZ189 with an 88 bp Pu·Py tract contained in pJB4 cloned into the EcoRI site. This tract contains two mirror repeat sequences. (**C**) Replication blockade during *in vitro* replication. pZ189 and pZ4.1 replicate almost exclusively in the presence of large T antigen. Compared to pZ189, pZ4.1 contains linearized replication products (form III) and persistent replication intermediates (RI).

in vitro under acidic conditions is unclear, it does appear to predict replication arrest. The replication fork blockade demonstrated in Figures 3D and 4A most likely represent replication arrest due to a Pu·Pu·Py triplex formation between the nascent and template strands during replication with the polypurine template strand folding back over the polypyrimidine nascent strand. This interstrand effect on replication would not be dependent on superhelical tension given the single-stranded template. While independent of superhelical tension, however, the foldback-triplex formation would be expected to share common features with the intramolecular triplex formation. For example, the energy required for foldback-triplex formation may correlate with the superhelical energy required for intramolecular triplex formation. Additionally, a longer mirror repeat sequence would be more likely to form either kind of triplex and would probably be more difficult for the polymerase to bypass.

We looked for other features of the Pu·Py tracts that correlate with polymerization arrest in an effort to identify sequence characteristics that can be used to predict which Pu·Py tracts in the human genome are more likely to interfere with replication. Besides superhelical tension of intramolecular triplex formation, we found that only the potential length of the triplex had a statistically significant correlation with polymerization arrest. The relationship between the length of the Pu·Py tracts and polymerase pausing has been suggested previously (28,41). As shown in Figure 3E, the primer extension studies demonstrated a considerable range in the amount of polymerase blockade. Given the correlation coefficients obtained, this variability is only partially explained by the factors examined in this study, reflecting the role of additional factors.

A limitation of this study is that Taq polymerase was used, which is not as processive as other polymerases. This limitation in processivity may explain the multiple bands observed in the primer extension studies that were not consistent with either a full-length product or polymerase arrest within the Pu·Py region. Because repeat sequences are prone to slipped mispairing mutagenesis, digestion of the plasmids with EcoRI was performed and showed a single length insert size, indicating that the starting templates for the primer extensions were homogenous in size (see Fig. S1).

Another limitation is that replication blockade using purified Taq polymerase does not model eukaryotic replication systems. We therefore examined the effect of the *PKD1* Pu·Py tracts on eukaryotic replication systems using both a primer extension assay (Fig. 4A) and and an SV40 system (Fig. 4C). Both assays demonstrated significant replication blockade in the presence of a Pu·Py tract from *PKD1* (Fig. 4A and C). In addition, the SV40 replication assay shows the presence of double-strand breaks which is only observed in the presence of the Pu·Py tract (Fig. 4C). Both findings further support the hypothesis that these sequences play a role in the development of human ADPKD.

A disease model is proposed in which the G-rich template of Pu·Py mirror repeat tracts of *PKD1* or its homologous genes engage in a foldback-triplex structure with the nascent strand during DNA replication resulting in a replication fork blockade (Fig. 5). In an effort to reactivate the blocked replication fork, recombination or double-strand breaks may be required which can compromise replication fidelity, leading

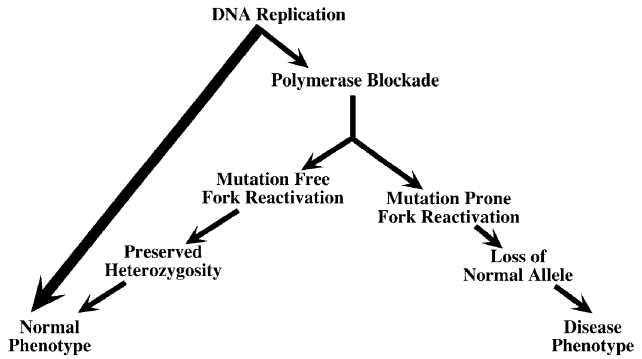

Figure 5. Flow diagram of the disease model. The replication blocking effect of the mirror repeat Pu·Py tract is a probability. A small percentage of the replication fork will be blocked. The majority of blocked forks are reactivated in a mutation-free manner; however, a small percentage will undergo error prone reactivation or repair. Because blocked replication forks lead to double strand breaks, repair may be required. In either scenario, a mutation may result leading to disease.

to a second hit mutation in the normal *PKD1* allele and LOH. Such a model is consistent with the gene conversion mutagenic mechanism proposed by Watnick *et al*. (42). Following replication blockade, gene conversion can occur using the homologous tracts of the *PKD1* gene at chromosome 16p13.1 (42–44).

This model supports the potential role of the Pu·Py tract in *PKD1* or homologous genes as a mechanism leading to somatic mutation and cystogenesis in ADPKD. Variation in the rate at which this second hit occurs may explain the variable expressivity seen with ADPKD. The development of the ability to predict which Pu·Py tracts are more likely to lead to mutagenesis and the development of methods to prevent or slow down the rate of triplex formation during replication may offer an opportunity to alter the disease course in these patients for whom only supportive treatment is currently available.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at NAR Online.

ACKNOWLEDGEMENTS

This work was supported in part by the NIH (DK61458) (J.J.B.), the Polycystic Kidney Disease Foundation (J.J.B.) and the Kidney Foundation of Greater Cincinnati (H.P.P.).

REFERENCES

1. Dalgaard,O.Z. (1957) Bilateral polycystic disease of the kidneys: a follow-up of two hundred and eighty-four patients and their families. *Acta Med. Scand. (Suppl.)*, **328**, 1–251.
2. Gabow,P.A. (1993) Autosomal dominant polycystic kidney disease. *N. Engl. J. Med.*, **329**, 332–342.
3. Levy,M., Duyme,M., Serbelloni,P., Conte,F., Sessa,A. and Grunfeld,J.P. (1995) Is progression of renal involvement similar in twins with ADPKD? A multicentric European study. *Contrib. Nephrol.*, **115**, 65–71.
4. Watson,M.L. and Torres,V.E. (eds) (1996) *Polycystic Kidney Disease*. Oxford University Press, New York.
5. Qian,F., Watnick,T.J., Onuchic,L.F. and Germino,G.G. (1996) The molecular basis of focal cyst formation in human autosomal dominant polycystic kidney disease type I. *Cell*, **87**, 979–987.
6. Watnick,T.J., Torres,V.E., Gandolph,M.A., Qian,F., Onuchic,L.F., Klinger,K.W., Landes,G. and Germino,G.G. (1998) Somatic mutation in individual liver cysts supports a two-hit model of cystogenesis in autosomal dominant polycystic kidney disease. *Mol. Cell*, **2**, 247–251.
7. Pei,Y. (2001) A 'two-hit' model of cystogenesis in autosomal dominant polycystic kidney disease? *Trends Mol. Med.*, **7**, 151–156.
8. Germino,G.G. (1997) Autosomal dominant polycystic kidney disease: a two-hit model. *Hosp. Pract.*, **32**, 81–82, 85–88, 91–92 passim.
9. Knudson,A.G.,Jr (1971) Mutation and cancer: statistical study of retinoblastoma. *Proc. Natl Acad. Sci. USA*, **68**, 820–823.
10. Niida,Y., Stemmer-Rachamimov,A.O., Logrip,M., Tapon,D., Perez,R., Kwiatkowski,D.J., Sims,K., MacCollin,M., Louis,D.N. and Ramesh,V. (2001) Survey of somatic mutations in tuberous sclerosis complex (TSC) hamartomas suggests different genetic mechanisms for pathogenesis of TSC lesions. *Am. J. Hum. Genet.*, **69**, 493–503.
11. Kruse,R., Rutten,A., Hosseiny-Malayeri,H.R., Bisceglia,M., Friedl,W., Propping,P., Ruzicka,T. and Mangold,E. (2001) 'Second hit' in sebaceous tumors from Muir-Torre patients with germline mutations in MSH2: allele loss is not the preferred mode of inactivation. *J. Invest. Dermatol.*, **116**, 463–465.
12. John,A.M., Ruggieri,M., Ferner,R. and Upadhyaya,M. (2000) A search for evidence of somatic mutations in the NF1 gene. *J. Med. Genet.*, **37**, 44–49.
13. Dahia,P.L., Marsh,D.J., Zheng,Z., Zedenius,J., Komminoth,P., Frisk,T., Wallin,G., Parsons,R., Longy,M., Larsson,C. *et al.* (1997) Somatic deletions and mutations in the Cowden disease gene, PTEN, in sporadic thyroid tumors. *Cancer Res.*, **57**, 4710–4713.
14. Henske,E.P., Neumann,H.P., Scheithauer,B.W., Herbst,E.W., Short,M.P. and Kwiatkowski,D.J. (1995) Loss of heterozygosity in the tuberous sclerosis (TSC2) region of chromosome band 16p13 occurs in sporadic as well as TSC-associated renal angiomyolipomas. *Genes Chromosomes Cancer*, **13**, 295–298.
15. Brasier,J.L. and Henske,E.P. (1997) Loss of the polycystic kidney disease (PKD1) region of chromosome 16p13 in renal cyst cells supports a loss-of-function model for cyst pathogenesis. *J. Clin. Invest.*, **99**, 194–199.
16. Koptides,M., Constantinides,R., Kyriakides,G., Hadjigavriel,M., Patsalis,P.C., Pierides,A. and Deltas,C.C. (1998) Loss of heterozygosity in polycystic kidney disease with a missense mutation in the repeated region of PKD1. *Hum. Genet.*, **103**, 709–717.
17. Badenas,C., Torra,R., Perez-Oller,L., Mallolas,J., Talbot-Wright,R., Torregrosa,V. and Darnell,A. (2000) Loss of heterozygosity in renal and hepatic epithelial cystic cells from ADPKD1 patients. *Eur. J. Hum. Genet.*, **8**, 487–492.
18. Colgin,L.M., Hackmann,A.F., Emond,M.J. and Monnat,R.J., Jr (2002) The unexpected landscape of *in vivo* somatic mutation in a human epithelial cell lineage. *Proc. Natl Acad. Sci. USA*, **99**, 1437–1442.
19. Martin,G.M., Ogburn,C.E., Colgin,L.M., Gown,A.M., Edland,S.D. and Monnat,R.J., Jr (1996) Somatic mutations are frequent and increase with age in human kidney epithelial cells. *Hum. Mol. Genet.*, **5**, 215–221.
20. Weinreb,A., Collier,D.A., Birshtein,B.K. and Wells,R.D. (1990) Left-handed Z-DNA and intramolecular triplex formation at the site of an unequal sister chromatid exchange. *J. Biol. Chem.*, **265**, 1352–1359.
21. Kim,S.K., Takahashi,M. and Norden,B. (1995) Binding of RecA to anti-parallel poly(dA).2poly(dT) triple helix DNA. *Biochim. Biophys. Acta*, **1264**, 129–133.
22. Rooney,S.M. and Moore,P.D. (1995) Antiparallel, intramolecular triplex DNA stimulates homologous recombination in human cells. *Proc. Natl Acad. Sci. USA*, **92**, 2141–2144.
23. VanRaay,T.J., Burns,T.C., Connors,T.D., Petry,L.R., Germino,G.G., Klinger,K.W. and Landes,G.M. (1996) A 2.5 kb polypyrimidine tract in the PKD1 gene contains at least 23 H-DNA-forming sequences. *Microbiol. Comp. Genomics*, **1**, 317–327.
24. Blaszak,R.T., Potaman,V., Sinden,R.R. and Bissler,J.J. (1999) DNA structural transitions within the PKD1 gene. *Nucleic Acids Res.*, **27**, 2610–2617.
25. Tiner,W.J.,Sr, Potaman,V.N., Sinden,R.R. and Lyubchenko,Y.L. (2001) The structure of intramolecular triplex DNA: atomic force microscopy study. *J. Mol. Biol.*, **314**, 353–357.
26. Dayn,A., Samadashwily,G.M. and Mirkin,S.M. (1992) Intramolecular DNA triplexes: unusual sequence requirements and influence on DNA polymerization. *Proc. Natl Acad. Sci. USA*, **89**, 11406–11410.

27. Krasilnikov,A.S., Panyutin,I.G., Samadashwily,G.M., Cox,R., Lazurkin,Y.S. and Mirkin,S.M. (1997) Mechanisms of triplex-caused polymerization arrest. *Nucleic Acids Res.*, **25**, 1339–1346.
28. Potaman,V.N. and Bissler,J.J. (1999) Overcoming a barrier for DNA polymerization in triplex-forming sequences. *Nucleic Acids Res.*, **27**, e5.
29. Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) *Molecular Cloning*: *A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
30. Barnes,W.M. (1994) PCR amplification of up to 35-kb DNA with high fidelity and high yield from lambda bacteriophage templates. *Proc. Natl Acad. Sci. USA*, **91**, 2216–2220.
31. Li,J.J. and Kelly,T.J. (1984) Simian virus 40 DNA replication in vitro. *Proc. Natl Acad. Sci. USA*, **81**, 6973–6977.
32. Cagnano,M., Benharroch,D. and Geffen,D.B. (1991) Pulmonary lymphangioleiomyomatosis. Report of a case with associated multiple soft-tissue tumors. *Arch. Path. Lab. Med.*, **115**, 1257–1259.
33. Carty,M.P., Hauser,J., Levine,A.S. and Dixon,K. (1993) Replication and mutagenesis of UV-damaged DNA templates in human and monkey cell extracts. *Mol. Cell. Biol.*, **13**, 533–542.
34. Carty,M.P., Zernik-Kobak,M., McGrath,S. and Dixon,K. (1994) UV light-induced DNA synthesis arrest in HeLa cells is associated with changes in phosphorylation of human single-stranded DNA-binding protein. *EMBO J.*, **13**, 2114–2123.
35. Hauser,J., Levine,A.S. and Dixon,K. (1988) Fidelity of DNA synthesis in a mammalian *in vitro* replication system. *Mol. Cell. Biol.*, **8**, 3267–3271.
36. Carty,M.P., el-Saleh,S., Zernik-Kobak,M. and Dixon,K. (1995) Analysis of mutations induced by replication of UV-damaged plasmid DNA in HeLa cell extracts. *Environ. Mol. Mutagen.*, **26**, 139–146.
37. Kohli,M. and Jorgensen,T.J. (1999) Lack of dependence on p53 for DNA double strand break repair of episomal vectors in human lymphoblasts. *Biochem. Biophys. Res. Commun.*, **264**, 702–708.
38. Sinden,R.R. (1994) *DNA Structure and Function*. 1st edn. Academic Press, New York.
39. Bacolla,A., Jaworski,A., Connors,T.D. and Wells,R.D. (2001) PKD1 unusual DNA confirmations are recognized by nucleotide excision repair. *J. Biol. Chem.*, **276**, 18597–18604.
40. Friedberg,E.C., Walker,G.C. and Siede,W. (eds.) (1995) *DNA Repair and Mutagenesis*. 1st edn. ASM Press, Washington DC.
41. Krasilnikova,M.M., Samadashwily,G.M., Krasilnikov,A.S. and Mirkin,S.M. (1998) Transcription through a simple DNA repeat blocks replication elongation. *EMBO J.*, **17**, 5095–5102.
42. Watnick,T.J., Gandolph,M.A., Weber,H., Neumann,H.P. and Germino,G.G. (1998) Gene conversion is a likely cause of mutation in PKD1. *Hum. Mol. Genet.*, **7**, 1239–1243.
43. Anonymous (1994) The polycystic kidney disease 1 gene encodes a 14 kb transcript and lies within a duplicated region on chromosome 16. The European Polycystic Kidney Disease Consortium. *Cell*, **77**, 881–894.
44. Bogdanova,N., Markoff,A., Gerke,V., McCluskey,M., Horst,J. and Dworniczak,B. (2001) Homologues to the first gene for autosomal dominant polycystic kidney disease are pseudogenes. *Genomics*, **74**, 333–341.