Journal of Cellular Biochemistry 89:1193–1205 (2003)

Non-Random Features of Loop-Size Chromatin Fragmentation

Ildikó Szilágyi,[1] Tamás Varga,[1] Lóránt Székvölgyi,[1] Éva Hegedüs,[1] Katalin Goda,[1] Viktória Kaczur,[1] Zsolt Bacsó,[1] Yuji Nakayama,[4] János Pósafi,[2] Sándor Pongor,[3] and Gábor Szabó, Jr.[1]*

[1]Department of Biophysics and Cell Biology, University Medical School of Debrecen, 4012 Debrecen, Nagyerdei krt. 98, Hungary
[2]New England Biolabs, 32 Tozer Rd., Beverly, Massachusetts 01915
[3]ICGEB-International Center for Genetic Engineering and Biotechnology, Padriciano 99, 34012 Trieste, Italy
[4]Lawrence Berkeley National Laboratory, 1 Cyclotron Rd., University of California, Berkeley, California 94720

Abstract Upon isolation of DNA from normal eukaryotic cells by standard methods involving extensive proteolytic treatment, a rather homogeneous population of loop-size, double-stranded DNA fragments is regularly obtained. These DNA molecules can be efficiently end-labeled by the DNA polymerase I Klenow fragment, as well as by a 3′- to -5′-exonuclease-free Klenow enzyme, but not by terminal transferase (TdT) unless the ends have been filled up by Klenow, suggesting that dominantly 5′ protruding termini are generated upon fragmentation. The filled-up termini were used for cloning the distal parts of the ~50 kb fragments. BLAST analysis of the sequence of several clones allowed us to determine the sequence of the non-cloned side of the breakpoints. Comparison of 25, 600 bp-long breakpoint sequences demonstrated prevalence of repetitive elements. Consensus motives characteristic of the breakpoint sequences have been identified. Several sequences exhibit peculiar computed conformational characteristics, with sharp transition or center of symmetry located exactly at the breakpoint. Our data collectively suggest that chromatin fragmentation involves nucleolytic cleavages at fragile/hypersensitive sites delimiting loop-size fragments in a non-random manner. Interestingly, the sequence characteristics of the breakpoints are reminiscent of certain breakpoint cluster regions frequently subject to gene rearrangements. J. Cell. Biochem. 89: 1193–1205, 2003.

Key words: apoptosis; chromatin; fragmentation; fragility; terminal transferase; Klenow; 50 kb

Routine procedures of DNA isolation involving extensive proteolytic digestion of cell lysates, phenol extraction, and ethanol precipitation usually yield DNA of ~50 kb average size. Since fragmentation of DNA can be avoided when live cells are embedded into agarose plugs prior to lysis, mechanical factors are generally believed to play the primary role in this loop-size fragmentation observed when lysis is performed in a cell suspension. Based on this view, random breakage of the long DNA molecules would be expected. Several observations, however, run counter this view: the average size of DNA molecules obtained during protein denaturing treatments of lysates prepared from normal cells without embedding is always around ~50 kb (forming in some cellular systems, a rather well-focused band, unexplained by a compression artifact, see Szabó et al., 1990); loop-size chromatin fragmentation can also be observed upon rapid alkaline lysis of live cells in suspension [Szabó and Bacsó, 1996], and when DNA is isolated from fixed cells [Gál et al., 2000]; this phenomenon has also been demonstrated when agarose plugs containing

Abbreviations used: PFGE, pulse field electrophoresis; FIGE, field-inversion electrophoresis; TdT, terminal deoxynucleotidyl transferase; ss, single-strand; ds, double-strand

Grant sponsor: Hungarian Scientific Research Fund (OTKA), National Committee for Technological Development (OMFB) and Ministry of Health, Social and Family Affairs (ETT); Grant numbers: OTKA 032563, OMFB-02692/2000, ETT T01/103.

*Correspondence to: Gábor Szabó, Jr., Department of Biophysics and Cell Biology, University Medical School of Debrecen, 4012 Debrecen, P.O.B. 39.
E-mail: szabog@jaguar.dote.hu

Received 23 October 2002; Accepted 5 May 2003
DOI 10.1002/jcb.10591

deproteinized chromatin of mammalian or yeast cells are melted to 60–85°C [Varga et al., 1999]. Based on the stepwise, large-scale chromatin fragmentation observed upon digestion of isolated nuclei with endogenous nucleases and after treatment with topoisomerase II poisons it was suggested that chromatin loops contain approximately 50 kb of DNA [Kokileva, 1988; Filipski et al., 1990]. The size distribution of the DNA in the experiments above are reminiscent of the early stages of chromatin fragmentation generated by apoptotic nucleases [Oberhammer et al., 1993]. A common denominator of these findings may be the unique fragility of chromatin at predilection sites marking loop-size regions. The mechanism responsible for the generation of loop-size fragments may involve nucleases operating at hypersensitive sites, similar to the models explaining apoptotic DNA degradation phenomena, or an abrupt release of torsional stress becoming manifest at the bases of chromatin loops upon deproteinization. Characterization of these predilection points might be useful for understanding how apoptotic DNA fragmentation is superimposed on analogous phenomena observed in non-apoptotic cells. Certain aspects of chromatin structure may also be reflected by the phenomenon studied, in view of the fact that loop-size fragmentation accompanying apoptosis is thought to reflect the loop-level of higher-order chromatin organization [Razin et al., 1991; Oberhammer et al., 1993; Gromova et al., 1995; Lagarkova et al., 1995; Khodarev et al., 2000]. Furthermore, molecular analysis of the breakpoints involved in ~50 kb fragmentation could perhaps open up leads to the possible relationship between these fragile sites and the hot-spots of pathological gene rearrangements.

Degradation of DNA into megabase, and/or ~50 kb, loop-size, discrete fragments, ultimately to form an oligonucleosomal ladder (for review see Nagata, 2000), are frequently observed gelelectrophoretic symptoms of apoptotic cell death. The free DNA termini generated upon chromatin fragmentation can be usually detected via end-labeling by terminal transferase (TdT) or by the *E. coli* Klenow enzyme. Degradation of DNA upon apoptosis may be part of the physiological program evolved to get rid of potentially harmful cell debris. On the other hand, maintenance of DNA integrity is a key element in the faithful propagation of genetic information, and is, therefore, carefully guarded against environmental and endogenous challenges by an elaborate machinery of DNA repair and checkpoint control. Unrepaired discontinuities may ultimately invoke cell death programs. In a series of studies, chromatin-bound nucleases were suggested to carry out the $\geq$50 kb cleavages, while soluble enzymes were implicated in the oligonucleosomal fragmentation steps [Pandey et al., 1994, 1997; Liu et al., 1998, 1999]. Others have demonstrated the role of a caspase-activated DNase (CAD) in both steps of DNA degradation [Sakahira et al., 1998]. Reconciliation of these and many other disparate findings may be through the consideration of complexity, assuming that several different nucleases may be responsible for the different aspects of DNA fragmentation, and that the fragmentation phenomena observed in the chromatin derived from apoptotic cells may be superimposed on a vulnerability already present in the chromatin of normal cells.

We addressed the question whether the ends of the ~50 kb fragments generated upon isolation of DNA from normal eukaryotic cells by standard methods involving proteolytic treatments possess distinctive features. First, the end-labeling characteristics of the fragments were studied, using urea-TAE agarose denaturing gelelectrophoresis. A dominantly 5′-protruding end configuration was determined, and this feature made the cloning of the ~50 kb end-sequences possible. The sequences of several clones were determined and compared with each-other, and with the databank sequences.

METHODS

Cells

The HL60 (human promyelocytic leukemia) and the Jurkat (human T cell lymphoma) cell lines were maintained in RPMI 1640, the NIH 3T3 (mouse fibroblast) cells were maintained in DMEM. Both media were supplemented with 20 mM glutamine, 10% fetal calf serum (Gibco, Invitrogen Corporation, Carlsbad, California), and antibiotics. The cells were washed twice with PBS (phosphate-buffered saline, pH 7.4) before further processing.

DNA Preparation

At a concentration of 10^7 cells/ml, the cells were resuspended in TE (10 mM TRIS-HCl, 1 mM EDTA pH 8.0) containing 1 mg/ml Proteinase K (PK, Sigma-Aldrich Kft., Budapest,

Hungary, molecular biology grade, self digested), Sarkosyl (*N*-lauroylsarcosine Na-salt, Sigma), and 20 mM EDTA. After incubation at 55°C overnight, the samples were diluted 5× with TE then fresh PK (to 0.5 mg/ml) and SDS (sodium dodecyl sulfate, Sigma; 1%) were added for a second cycle of digestion. The samples were then extracted with phenol/chloroform, precipitated with ethanol, and redissolved in TE.

Enzyme Reactions

Terminal deoxynucleotidyl transferase (TdT), Klenow, and the Klenow exo$^-$ fragment of DNA polymerase I were purchased from New England Biolabs (Beverly, MA) and used at 25, 5, and 5 U/ml concentration, respectively. The restriction enzymes were from Amersham Biosciences UK, Ltd. (Buckinghamshire, England). α-[^{32}P]dATP (110 TBq/mmol) was from Izinta Trading Co., Ltd., Hungary, Budapest. The protocols suggested by the manufacturer were followed. Labeling reactions using the Klenow enzymes were performed in the presence of dGTP, dCTP, dTTP, and α-[^{32}P]dATP. In the case of the TdT reactions, unlabeled dTTP and dGTP were included in the reaction mixture along with α-[^{32}P]dATP. The unincorporated nucleotides were removed by gel filtration on Sephadex G25 columns (Pharmacia Biotech).

Gelelectrophoresis and Autoradiography

Samples containing 1 μg DNA were loaded in the appropriate sample buffer onto 1% agarose gels that were run at 6–8 V/cm in 1× TAE (40 mM Tris-acetate, 1 mM EDTA) using an MJ Research PPI 200 power inverter, in a cold room. Running parameters were set to maximize resolution in 50–400 kb range. Lambda DNA and its concatemers (both from Sigma) were used as molecular weight standards. After gelelectrophoresis, DNA was transferred to a Hybond-N$^+$ membrane (Amersham), the filters were exposed on a PhosphorImager plate for up to 16 h and analyzed by a Molecular Dynamics PhophorImager. In some experiments, standard X-ray film autoradiography was used, as indicated in the figure legends. Urea-agarose gelelectrophoresis was carried out according to Materna et al. [1998]; (see also Varga et al., 1999), resolving single-stranded (ss) and double-stranded (ds) DNA in the same gel. This technique allows recognition of ss discontinuities, as opposed to standard non-denaturing electrophoresis. The procedure involves denaturation of DNA by heat-treatment and electrophoresis in the presence of urea. Urea (0.012 g) was added to each 25 μl sample then non-denatured and heat-denatured sample pairs were prepared, with the heat-denatured samples placed to 95°C for 5 min before applying to the gel. During gel preparation, urea was added to a final concentration of 1 M when the solution cooled below 60°C. The electrophoresis buffer was also prepared from 1× TAE containing 1 M urea. Wide-bore pipette tips were applied for handling the samples to minimize mechanical shearing. Staining of the gels with 0.5 μg/ml ethidium bromide was carried out after electrophoresis.

Cloning and Sequencing of the ~50 kb Fragment Ends

Blunt ends were created by Klenow enzyme and the gel-purified, blunt-ended ~50 kb fragments were ligated to Sma I digested pUC18. The predigested, dephosphorylated "Ready-To-Go" pUC18 vector was from Amersham Biosciences. Cleavage of this sample by Sac I created (3′-protruding) cohesive termini, leaving only the distal fragments of the ~50 kb DNA attached to the plasmid. After a second cycle of ligation, the purified DNA was electroporated using a BTX ECM600 electroporator into SURE (Stratagene, Amsterdam, The Netherlands) bacteria providing blue/white selection of recombinants on X-Gal. For sequencing reactions, the M13 forward primer was used. Sequences were determined partly at the Biological Research Center of the Hungarian Academy of Sciences, Szeged, partly in János Sümegi's lab at University Nebraska Medical Center (Omaha, NE).

Sequence Analysis

The MEME/MAST motif discovery tool [Bailey and Elkan, 1994] was accessed through the following url: http://meme.sdsc.edu/meme/website/meme-intro.html. Repetitive elements were identified by the CENSOR program (http://www.girinst.org/Censor_Server-Data_Entry_Forms.html). Motives with potential transcription factor binding property were identified by the program AliBaba2.1 (www.alibaba2.com). For BLAST analysis, we used the following urls: http://www.ncbi.nlm.nih.gov/BLAST/ and http://dove.embl-heidelberg.de/Blast2/ (WU-BLAST); nucleotide BLAST against UniGene Cluster Sequences was made possible by http://

cgap.nci.nih.gov/Genes/SeqFinder. For pairwise comparison of different sequences we used the BLAST 2 program, or a dot-plot procedure (Dotlet): http://www.isrec.isb-sib.ch/java/dotlet/dotlet_about. html. The plot.it® Server (http://www3.icgeb.trieste.it/~dna/plot_it.html) was used to calculate conformational characteristics of the breakpoint sequences, setting window sizes of 3 for short, and 10 for longer sequences. The predicted 3-D models of the consensus motives were first built based on a consensus model of DNA curvature [Munteanu et al., 1998], and the raw models were refined by simulated annealing [Vlahovicek and Pongor, 2000]. The long models (Fig. 3) were built without simulated annealing. CpG distribution was calculated by: EMBOSS::CpGPlot (http://www.ebi.ac.uk/emboss/cpgplot/). The following MAR-finder program was used: http://www. futuresoft.org/MAR-Wiz//. The breakpoints in the middle of the 600 bp-long sequences analyzed are listed below (clone number, accession number, Genebank position of breakpoint; respectively): (1, AL139403, 43243); (2, AL353738, 41680); (5, AC025478, 68009); (9, AC009702, 143766); (10, AL109740, 113617); (11, AC012479, 159759); (12, AL161773, 41277); (13, AC023487, 123554); (14, AL157397, 8788); (15, AC024171, 111877); (19, AC019262, 100776); (20, AL137023, 30275); (21, X02418, 1433); (25, AC004993, 21562); (27, AL121983, 39061); (29, AC002073 32227); (30, AC021329, 72010); (33, AC004499, 8242); (34, AL132670, 36238); (36, AL121919, 10130); (38, AC015722, 61453); (39, AC010195, 154494); (40, AC012076, 5848); (41, AL008722, 14745); (43, AC012230, 67385).

RESULTS

In order to clone a representative sample of the ~50 kb fragment ends, their end-configuration was first characterized.

Figure 1A–C demonstrate that the ~50 kb fragments can be efficiently labeled with α-[^{32}P]dATP using the Klenow enzyme. Figure 1 also shows that DNA fragments electroeluted from the different size bands on the gel, precipitated and redissolved, could be re-run without major degradation, suggesting that artefactual mechanical breakage is not an important factor determining the ds size of the DNA molecules in our system. The labeling characteristics were also studied on urea-agarose gels, analyzing ds and ss samples simultaneously (Fig. 2). In view of the ~50 kb peak in the ^{32}P-labeled fragment length distribution, a major part of labeling by Klenow occurs at the termini in addition to labeling at internal sites, according to PhosphorImager line-scans of the heat-denatured samples (Fig. 1B, lower panel). Labeling by the Klenow enzyme is made possible by the presence of an accessible hydroxyl on 3′-recessed ends. However, phosphorylated hydroxyls that may be present as a result of various genotoxic events could also be processed by the Klenow enzyme which retains a 3′- to 5′-exonuclease activity. As Figure 1C shows, an exo$^-$ mutant Klenow enzyme, which lacks this activity, also efficiently labeled the ~50 kb fragments. Thus, the ~50 kb fragments have recessed 3′-ends with hydroxyl groups accessible to enzymes.

Unexpectedly, the ~50 kb fragments could not be labeled by TdT using either Mg^{++} or Co^{++} as the cofactor of the reaction, in spite of their strong labeling with both Klenow enzymes, as shown in Figures 1B and 2A. When an aliquot of the same DNA sample was partially digested with the restriction enzyme Pst I (producing 3′-protruding ends) prior to addition of TdT, strong labeling was achieved in the complete size range of fragments present, demonstrating that the lack of TdT labeling in the unrestricted samples was due to the absence of an appropriate DNA substrate (Fig. 2A). Figure 2B compares the efficiency of TdT labeling in the case of Lambda phage DNA PstI and EcoRI digests (generating 3′- and 5′-protruding ends, respectively). Figure 2C shows that incorporation of ^{32}P by TdT into 50 kb ds fragments was well detectable when the TdT reaction followed the extension of the recessed 3′-ends by the Klenow enzyme.

The dominantly Klenow$^+$/TdT$^-$ configuration of the fragment termini was used for cloning the ends (for details of the strategy see Methods). The sequence of several clones was determined and the location of our sequences in the genome identified by BLAST searching of the databank. Using sequence elements containing 300–300 bp from both sides of the breakpoints, we searched for homologies within GeneBank sequences collected into a non-redundant set of gene-oriented clusters (UniGene, see Methods). About 50% of the breakpoint sequences have produced many hits, showing that our clones contain sequence elements that are frequently shared by genes and their sequence environment. Notably, many hits were from hot-spots of chromosomal rearrangements: clone 2,

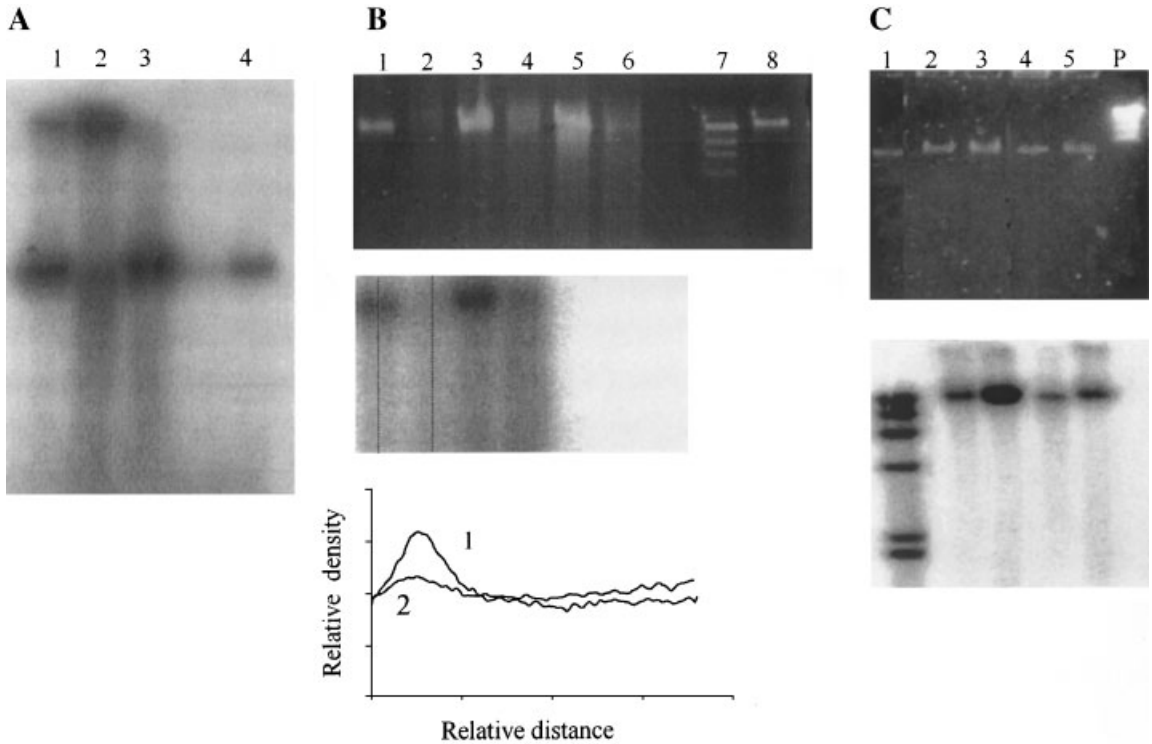

Fig. 1. **A**: α-[^{32}P]dATP-labeling of HL60 DNA by Klenow enzyme. **Lane 1**: HL60 DNA, containing both ~50 kb and larger size fragments; **lane 2**: re-run of electroeluted DNA derived from the high molecular weight species shown in **lane 1**; **lane 3**: re-run of electroeluted DNA derived from the ~50 kb band of the previous samples; **lane 4**: λ DNA. The DNA samples were run on neutral agarose gels by FIGE, ^{32}P incorporation was detected by film autoradiography. **B**: Urea-agarose gelelectrophoretic analysis of the ss length distribution of the ~50 kb (ds size) HL60 DNA samples following Klenow- or TdT-labeling with α-[^{2}P]dATP. **Upper panel**: ethidium bromide stained gel; **middle panel**: ^{32}P incorporation detected by PhosphorImager. **Lower panel**: line-scans of **lanes 1** and **2** in panel B. Samples in **lanes 1** and **3**: Klenow-labeled DNA, non-denatured; **lanes 2** and **4**: Klenow-labeled DNA, denatured. **Lanes 5** and **6**: non-denatured and denatured DNA, respectively, after TdT reaction. **Lane 7**: λ Hind III fragments; **lane 8**: λ DNA. **C**: α-[^{32}P]dATP-labeling of ~50 kb DNA by Klenow or Klenow exo$^-$ enzymes, resolved on a neutral agarose gel by FIGE. **Upper panel**: ethidium bromide stained agarose gel; **lower panel**: ^{32}P incorporation detected by PhosphorImager. **Lane 1**: λ EcoRI fragments ^{32}P-labeled by Klenow; **lane 2**: ~50 kb DNA fragments derived from Jurkat cells labeled by Klenow; **lane 3**: ~50 kb DNA fragments derived from NIH 3T3 cells labeled by Klenow; **lane 4**: ~50 kb DNA fragments derived from Jurkat cells labeled by Klenow exo$^-$; **lane 5**: ~50 kb DNA fragments derived from 3T3 cells labeled by Klenow exo$^-$; P: pulse-marker (λ concatemers).

14q32.11; 12, 8p21.1; 14, 3q29, 17p13.3, 18q12, etc.; 15, 17p13.3, 13q12.2-q13.3, 4q32-q34, 15q15.3, etc.; 27, 19p13.3, 17q21.32, 19p13.12, etc.; 29, 7p15.3, 12p13, 3q29, etc.; 41, 15q24.3, 4q22.1, 5p14.3-q13.3, etc.; 11, 1p32.3, 1q42.3, 18q12, etc.; 21, 1p34.1, 38, 5q31.1, 1p35.2, 4q11, etc. (the recurrent ones are underlined). According to the modeling of the breakpoint sequences (by the plot.it server, see Methods), they tend to contain curved DNA (for representative examples see Fig. 3). Based on the 600 bp long sequences containing the breakpoints, we looked for possible sequence regularities. Analysis of 25 breakpoints demonstrated common repetitive elements that are also shared by known breakpoint regions frequently involved in gene rearrangements. As shown in Table I, 19 clones had repetitive elements within the sequences, usually involving the breakpoints themselves. Figure 3C shows the localization of the ALU element in certain breakpoint sequences: in the case of clone 14, 29, and 38, the ALU element is on one side of the breakpoint, while in clone 15, 27, and 11 ALU is bisected by the breakpoint. The frequency of regions with potential MAR/SAR characteristics, based on any combination of the criteria listed in the MAR-finder program applied (see Methods) to screen the assembled sequences of the clones, did not appear to exceed the density expected in random DNA. However, we note that several clones (5/25) possess an AT-rich sequence element suggested to play a role in these structures [Liebich et al., 2002], right in the vicinity of the breakpoints. The sequences

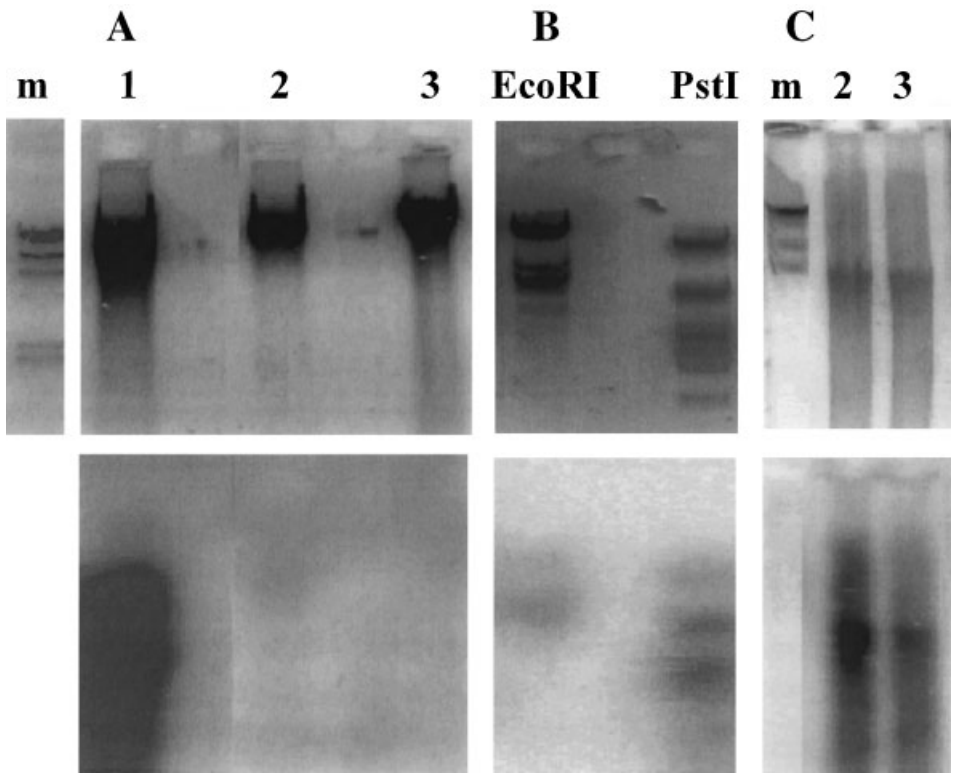

Fig. 2. **A**: α-[^{32}P]dATP labeling by TdT of ~50 kb HL60 DNA and its Pst I digest. **Upper panel**: ethidium bromide stained agarose gel; **lower panel**: ^{32}P incorporation detected by PhosphorImager. **Lane 1**: DNA was partially digested with Pst I and then TdT reaction was performed adding the enzyme and the triphosphates directly into the digest. **Lanes 2** and **3**: undigested DNA samples after TdT-labeling; Co^{++} (lane 2) or Mg^{++} (lane 3) was used as the cofactor of the enzyme. Marker (m): λ Hind III fragments (no ^{32}P label). **B**: Incorporation of α-[^{32}P]dATP into λ DNA digested with PstI or EcoRI restriction enzymes by TdT. **Upper panel**: ethidium bromide stained agarose gel; **lower panel**: ^{32}P incorporation detected by PhosphorImager. **C**: α-[^{32}P]dATP labeling of ~50 kb DNA fragments by TdT, following extension of their ends by Klenow enzyme in the presence of all four unlabeled triphosphates. **Upper panel**: ethidium bromide stained agarose gel; **lower panel**: ^{32}P incorporation detected by PhosphorImager. ~50 kb DNA fragments were from HL60 (**lane 2**), or NIH 3T3 cells (**lane 3**). Marker (m): λ concatemers (no ^{32}P label).

have also been searched for possible base unpairing regions (BURs; [Kohwi-Shigematsu and Kohwi, 1990]); only a minority of the sequences were suspected to have such features based on sequence characteristics and in fact most of these proved to be BUR-negative in gel-shift assays of recombinant SATB1 binding (data not shown). Several sequences of our clones exhibit peculiar conformational characteristics, with the center of symmetry or the transition point in the computed parameters located exactly at, or close to, the breakpoint. Examples of such features are demonstrated in Figure 4, panels A–D. Some of the breakpoints exhibited a striking symmetry exactly at the breakpoint, in terms of CpG dinucleotide frequency (Fig. 4, panels E, F). Using the MEME motif discovery program [Bailey and Elkan, 1994], we were searching for short (6–12 bp) and long (6–50 bp) multilevel consensus motives. Among the short elements shared by several breakpoints, the palindrome CCAGCCTGG, a topoisomerase II binding consensus motif and AAAAAAACAAAA were identified. These motives are potential transcription factor (including C/EBPα, HNF-3, Hb,YY1, and AP-2α) binding sites. The ratio of observed/expected occurrences of the above motives in our sequences, based on the complete sequence of ten human chromosomes, is around 5 and 11, respectively, considering only the perfect matches. Examples of the palindromic multilevel consensus sequence in the positive clones are shown in Table II. According to modeling of this motif (see Methods), CCAGCCTGG is strongly curved toward the major groove (Fig. 3B). There appears to be no general regularity as to the localization of the individual motives, although in several sequences they appear to be clustered around the breakpoints themselves (data not shown). In

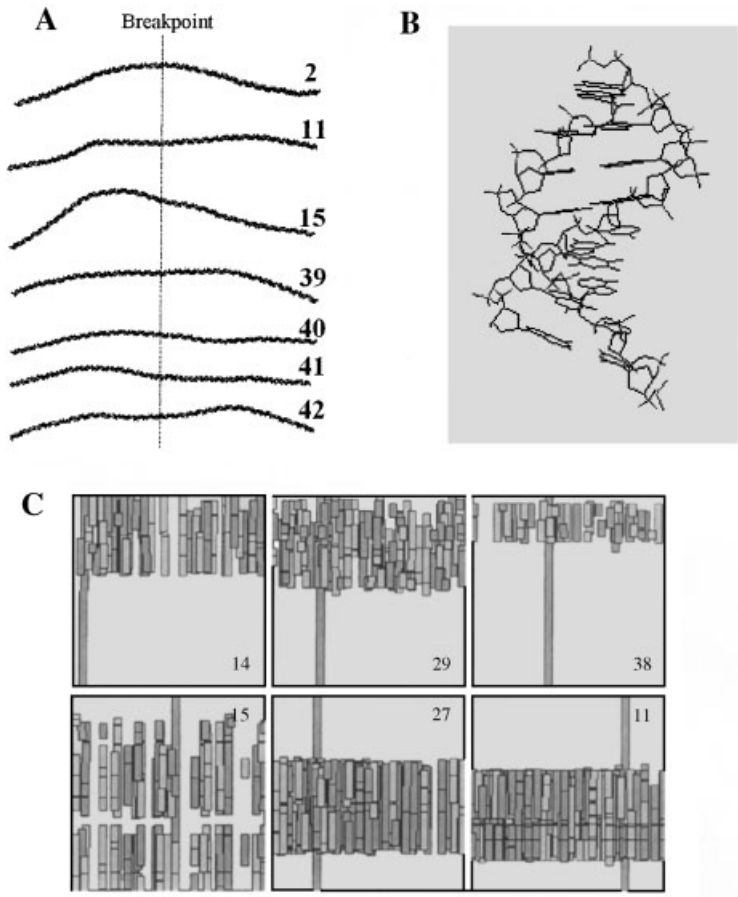

Fig. 3. **A**: The curved configuration predicted in some of the breakpoint sequences (numbers denote clone numbers; see their list in Methods). The 3-D views show segments of 300 nucleotides aligned at the experimentally observed breakpoint (thin perpendicular line). Conformations were predicted with the model.it program ([Vlahovicek and Pongor, 2000], using the Plot.it® server; see Methods). **B**: Atomic detail representation of the curved conformation predicted at the short consensus motif CCAGCCTGG. Conformation was predicted with the model.it program of the same server. **C**: Localization of ALU repetitive elements in some clones, based on BLAST 2 analysis (see Methods). In the case of each clone shown (see numbers 14, 29, 38, 15, 27, and 11), the 600 bp long sequence is compared with the complete Genebank clone it is part of. In this *dotplot* representation, the breakpoint sequence is aligned on the Y axis, and the corresponding Genebank clone sequence is assigned to the X axis; the gray level of bar-hatching indicates the degree of homologies. The position of the bars vertically encompassing the full range (600 bp) indicate the position of the breakpoint sequence within the Genebank clone; in the case of clone 15, e.g., the 600 bp clone is at bp 111,877 within the 208,881 bp-long Genebank clone, AC024171. Due to the differences in the length of the clones (not indicated) the bars partially overlap at the parameter settings applied.

a longer (up to 50 bp) analysis window, further consensus motives were found, represented in our sequences at an average frequency of approximately one per breakpoint sequence (data not shown). For comparison, arbitrarily chosen restriction enzyme cleavage sites within the ~150 kb human breakpoint cluster region (BCR) gene (on chromosome 22q11.21, involved in the BCR/ABL rearrangement of chronic myeloid leukemia; accession number U07000) were selected to define 600 bp long sequences which were then analyzed in a similar fashion. Only ~1/3 of these sequences exhibited long consensus motives, identified in the same window setting of the MEME program. No multilevel consensus sequences were found in the scrambled breakpoint sequences. One of the long motives detected by MEME, a 50 bp-long ALU sequence, and one of the short motives that is also part of ALU (CCAGCCTGG) co-localized with known hot-spots of BCR rearrangements (e.g., around nucleotide 125000 in U07000). In the 20–300 bp window, three (114, 77, and 39 bp long) consensus sequences have been detected by MEME, two of them shared by five clones (clones 29, 11, 27, 15, 14; all scoring below the probability level of 2.43e−50); the third one is present in these clones as well as in further four

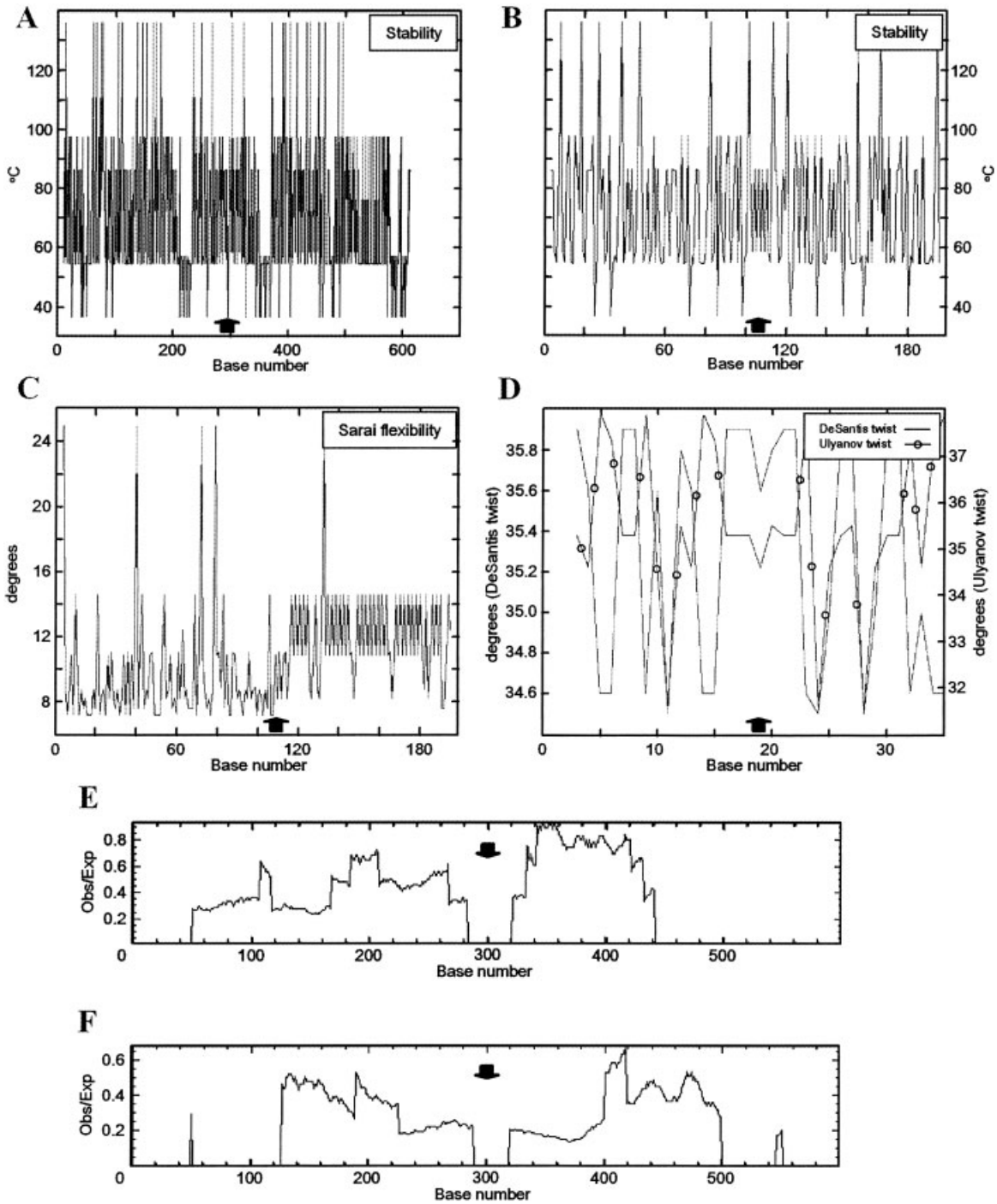

Fig. 4. **A–B**: Double helix stability determined from melting profiles predicted according to Gotoh and Tagashira [1981]. A: breakpoint sequence no. 15 (600 bp); B: no. 33 (200 bp). The thick arrow points at the breakpoint. (Conformational analysis by Plot.it; see Methods.) **C**: Flexibility of DNA obtained from conformational energy calculations [Sarai et al., 1989], expressed as dinucleotide twist, roll and tilt angles, in the case of breakpoint sequence no 10 (200 bp). Thick arrow: breakpoint. (Conformational analysis by Plot.it; see Methods.) **D**: Twist angles in sequence no. 1 (40 bp) determined from NMR data (empty circles; [Ulyanov and James, 1995]), and as predicted based on conformational energy calculations [De Santis et al., 1990]. Thick arrow: breakpoint. (Conformational analysis by Plot.it; see Methods.) **E**, **F**: Observed to expected ratio of C plus G to CpG calculated by CpGPlot [Larsen et al., 1992]; see Methods. (The expected number of CpG patterns is the number of CpG dinucleotides expected based on the frequency of C's and G's in that window. The observed number is the number of times a 'C' is found followed immediately by a 'G'.) E: no. 1 (600 bp); F: no. 2 (600 bp). Thick arrow: breakpoint.

TABLE I. Repetitive Elements in the Breakpoint Sequences

5	(A)
9	MIR@2
10	HERV39
11	ALU-Jo, ALU-Sg
12	L1ME_ORF2
14	ALU-Spqxz
15	(TTTTG), ALU-Jb, ALU-Sxzg (CAAAA), L1M2C_5
20	L1M3A_5
21	ALR1, ALR2
25	HERVIP10FH, HAL1
27	ALU-Spqxz, ALU-Sx
29	ALU-Sx
30	L2A
33	MER20
34	LTR67, MIR
36	L1M4B
38	HERV70_I (TTTTG), ALU-Ya5
41	L1
43	L2A

Numbers indicate the code name of the sequences found positive by the Censor program (see Methods) from among 25 sequences analyzed.

clones (all below level 1.09e−09). When the nucleotides of the same training set of 25 sequences were shuffled, no multilevel consensus sequences were detected.

DISCUSSION

Dominantly 5′-Protruding Configuration of the 50 kb Fragment Termini

In view of the autoradiographic pictures made of urea-agarose denaturing gels (Fig. 1), labeling by Klenow enzyme occurred at the termini, rather than at internal sites. Incorporation of [^{32}P]dATP mostly at random internal nicks or gaps would rather be expected to give rise to ≪50 kb average ss size of the labeled fragments. The intensive labeling by Klenow is suggestive of a dominantly 5′-protruding nature of the ∼50 kb fragments, with free and accessible hydroxyls on recessed 3′-ends. This interpretation could be in line with the fact that the optimal substrates for TdT reaction are 3′ protruding termini [Roychoudhury and Wu, 1980; Takemoto et al., 1998]. Thus, a prevalent 5′-protruding nature of the termini could explain Klenow-positivity and TdT-negativity at the same time. However, a uniformly AT-rich terminal configuration would probably also allow for TdT labeling of the 5′- overhangs due to the "breathing" of the ds DNA. While labeling by Klenow in the presence of the four triphosphates can occur exclusively at termini with 5′-protruding configuration, TdT is less discriminative in the presence of Co^{++} ions [Roychoudhury and Wu, 1980; Takemoto et al., 1998]. Indeed, those fragments of the lambda Eco RI digest ending in AT-rich sequences could be labeled by TdT, as opposed to fragments ending in GC-rich sequences (as concluded from a comparison of labeling intensities shown in Fig. 2B with the sequences of the termini of the lambda Eco RI fragments). However, the ends of several ∼50 kb fragments showed no GC predominance (for the sequences analyzed, see their list in Methods), suggesting involvement of other factors.

The lack of labeling by TdT in solution is also unexpected in view of the fact that cytocentrifuged, fixed cells can become TUNEL-positive upon proteinase treatment [Gál et al., 2000]. However, a weak but well detectible end-labeling of ∼50 kb DNA prepared from previously ethanol-fixed cells by TdT was observed,

TABLE II. Examples Within the Breakpoint Sequences of the Palindromic Multilevel Consensus Motif CCAGGCTGG Detected by the MEME Motif Discovery Tool (at $P < 8.36e{-}5$; for the Program see Methods)

38	C	C	A	G	G	C	T	G	G
11 (3)	C	C	A	G	G	C	T	G	G
36 (3)	C	C	A	G	G	C	T	G	G
29 (2)	C	C	A	G	G	C	T	G	G
27 (2)	C	C	A	G	G	C	T	G	G
15 (3)	C	C	A	G	G	C	T	G	G
14	C	C	A	G	G	C	T	G	G
42	C	C	A	G	C	A	T	G	G
40	C	C	A	G	C	A	T	G	G
39	A	C	A	G	G	C	T	G	G
2	C	C	A	T	G	A	T	G	G
40	C	C	A	G	C	G	T	G	G
34	A	C	A	G	C	A	T	G	G
16	C	C	T	G	G	A	T	G	G
2	C	C	T	G	G	A	T	G	G

The code name of the sequences containing the motif is followed by the number of motif copies in that sequence. In the case of multiple occurrences, the sequence of the best motif is shown.

while parallel samples of non-fixed origin were completely negative (data not shown). Fixation may facilitate the proteolytic digestion and elimination of protein contaminants close to the 5′-protruding DNA termini that might otherwise prevent TdT labeling by hampering DNA melting at the ends.

DNA fragmentation may involve a series of enzymatic or chemical processes including steps generating 3′-hydroxyls, as well as others that eliminate them by generating 3′-phosphates. In *Caenorhabditis elegans* [Wu et al., 2000], apoptotic DNA fragmentation was shown to occur in two steps, with a DNase II-like (NUC-1) activity generating TdT-silent, 3′-phosphorylated termini following the initial cleavages that generate TdT-positive termini. DNase II generates nicks with 3′-phosphate and 5′-hydroxyl termini [Bernardi, 1971], just like topoisomerase I [Yang et al., 1996]. 3′-phosphate groups are also among the major strand-break modifications emerging in the wake of being exposed to ionizing radiation and some other genotoxic effects [Takemoto et al., 1998]. However, labeling of the ~50 kb fragment ends by Klenow exo$^-$ in our system argues against the presence of 3′-phosphorylated termini. This points to the first interpretation, assuming 5′-protruding ends with free 3′-hydroxyls as the more likely explanation for our end-labeling data. Such a structure may be indicative of an enzymatic mechanism of the cleavages, and suggests that our cloning strategy is probably unbiased.

Relationship With Apoptotic DNA Fragmentation

The size distribution of DNA fragments derived from normal and apoptotic cells may be strikingly similar depending on the extent of the proteolytic treatment, in line with the fact that DNA fragmentation in the caspase-dominated milieu of apoptotic cells is already observed at mild proteolytic circumstances [Szabó, 1995]. Although the ~50 kb fragments were shown to include blunt-ended molecules [Khodarev et al., 2000], their proportion in the cell lysates was not investigated. Furthermore, the biochemistry of the cleavages is probably complex (see Wu et al., 2000, above; Jin et al., 1999), therefore, the blunt [Khodarev et al., 2000] or 1-base 5′- overhangs with 5′-phosphate and 3′-hydroxyl [Widlak et al., 2000] end configuration reported for apoptotic fragmentation should not be considered decisive evidence against the relationship between the two phenomena.

Sequence Characteristics of the Breakpoints

It is possible that the fragile/hypersensitive regions involved belong to several classes reflecting different mechanisms acting in parallel, blurring perhaps more significant individual regularities. Notwithstanding, consensus motives, with potential transcription factor binding capability, have been identified. Their frequency appears to exceed the level of average genomic occurrence (based on real, rather than calculated, frequency data). Repetitive elements, expected to account for ~30% of randomly fragmented human DNA [Moyzis et al., 1989; Stallings et al., 1990], have been observed in the majority of the breakpoint sequences analyzed. The architecture of the wider sequence environment of the breakpoints is, in general, moderately curved. These general features collectively suggest that fragmentation was non-random, rather, it occurred at special predilection points. This impression has been further substantiated by some features of the consensus motives identified in the breakpoint sequences.

Although the AAAAAAACAAAA motif resembles one of the signal sequences of VDJ recombination, no example of true heptamer–nonamer signal-pair has been detected within the sequences analyzed. Another short motif, CCAGCCTGG is strongly curved (Fig. 3B), just like the similar but octameric CCAGCTGG, a palindromic sequence suggested to be involved in rearrangements regularly inflicting the MLL gene in childhood acute leukemia and posttherapeutic leukemia [Aplan et al., 1996; Stanulla et al., 1997; Stanulla et al., 1998; Strissel et al., 1998]. Cleavage at MLL appears as part of the ~50 kb fragmentation of chromatin in apoptosis [Stanulla et al., 1997]. Treatment of human lymphoid cells with topoisomerase II inhibitors, or just exposing them to mild stress, leads to cleavage within a 8.3 kb region of MLL, implicating a topoisomerase II binding motif containing the above octamer [Aplan et al., 1996]. The cells which are not able to repair these discontinuities are thought to undergo apoptosis, while repair may occasionally lead to rearrangements at this locus. The cleavages producing ~50 kb fragments upon DNA isolation may involve fragile sites exemplified by the MLL breakpoint.

We have compared two rearrangement-prone 500 bp-long subregions (between nucleotides 74,501–75,000 and 125,000–125,500) of the ~150 kb BCR gene and found that six clones from our 25 showed extensive homologies with the two subregions, in ALU repetitive elements. Although the issue of ALU involvement in the rearrangements leading to t(9;22) translocation has been controversial, the non-random arrangement of these interspersed repeats in the BCRs is highly impressive [Papadopoulos et al., 1990; Jeffs et al., 1998]. We also note that ALU sequences have been recently implicated as binding sites for a cohesin-chromatin-remodelling complex [Hakimi et al., 2002]. It is interesting in this context that the more exposed, GC-rich chromatin regions flanked by cohesin-binding AT-rich DNA segments fluctuating along the chromosomal DNA with loop-size periodicity have been suggested to correspond to recombinational hotspots [Filipski and Mucha, 2002]. The ALU elements found in six of our sequences share the short motives with breakpoints devoid of ALU and this repetitive element was mapped immediately adjacent to the breakpoint itself in some clones (see Fig. 3C).

The same parameters that exhibit sharp transitions exactly at the breakpoints of our sequences behave in an analogous fashion at some, but not all, of the known BCR breakpoints (data not shown). The fluctuation of one of these parameters (twist angle) has been suggested to be indicative of fragile, high-flexibility regions [Mimori et al., 1999]. Although the immediate sequence environment of the breakpoints analyzed here exhibit no preference for AT or GC, a peculiar, symmetrical behavior of CpG dinucleotide distribution was observed around several (8/25) of the breakpoints (see Fig. 4E,F). Interestingly, such patterns of CpG distribution can also be observed in some hot-spots of BCR rearrangement (not shown). These observations may be related to instances when the GC and/or CpG content of the DNA appeared to influence the propensity for the formation of nuclease hypersensitive regions affecting higher order chromatin organization [Mucha et al., 2000].

In summary, the ~50 kb DNA fragments obtained upon isolation of DNA from mammalian chromatin by conventional procedures appear to involve non-random sequences reminiscent of certain gene regions frequently implicated in chromosomal rearrangements. Our data support the possibility that the fragile/hypersensitive regions involved in loop-size fragmentation of chromatin derived from normal cells may be the substrates for gene rearrangements in a pathological scenario.

ACKNOWLEDGMENTS

We thank Dr. János Sümegi (Omaha, Nebraska), Dr. Andor Udvardy (Biological Research Center, Szeged, Hungary), and Dr. Dieter Werner (DKFZ, Heidelberg, Germany) for valuable advice and helpful discussions, Dr. János Sümegi for the sequencing of several breakpoint clones and Dr. Terumi Kohwi-Shigematsu (Lawrence Berkeley Laboratory, Berkeley, CA) for generous help and advice in the evaluation of certain sequence characteristics. The authors thank Daniel Benediktsson (University of Debrecen) for the linguistic corrections.

REFERENCES

Aplan PD, Chervinsky DS, Stanulla M, Burhans WC. 1996. Site-specific DNA cleavage within the MLL breakpoint cluster region induced by topoisomerase II inhibitors. Blood 87:2649–2658.

Bailey TL, Elkan C. 1994. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. Proc Int Conf Intell Syst Mol Biol 2:28–36.

Bernardi G. 1971. The enzymes. In: Boyer PD, editor. New York: Academic Press. pp 271–287.

De Santis P, Palleschi A, Savino M, Scipioni A. 1990. Validity of the nearest-neighbor approximation in the evaluation of the electrophoretic manifestations of DNA curvature. Biochemistry 29:9269–9273.

Filipski J, Mucha M. 2002. Structure, function, and DNA composition of *Saccharomyces cerevisiae* chromatin loops. Gene 300:63–68.

Filipski J, Leblanc J, Youdale T, Sikorska M, Walker PR. 1990. Periodicity of DNA folding in higher order chromatin structures. EMBO J 9:1319–1327.

Gál I, Varga T, Szilagyi I, Balazs M, Schlammadinger J, Szabó G, Jr. 2000. Protease-elicited TUNEL positivity of non-apoptotic fixed cells. J Histochem Cytochem 48:963–970.

Gotoh O, Tagashira Y. 1981. Locations of frequently opening regions on natural DNAs and their relation to functional loci. Biopolymers 20:1043–1058.

Gromova II, Thomsen B, Razin SV. 1995. Different topoisomerase II antitumor drugs direct similar specific long-range fragmentation of an amplified *c-MYC* gene locus in living cells and in high-salt-extracted nuclei. Proc Natl Acad Sci USA 92:102–106.

Hakimi MA, Bochar DA, Schmiesing JA, Dong Y, Barak OG, Speicher DW, Yokomori K, Shiekhattar R. 2002. A chromatin remodelling complex that loads cohesin onto human chromosomes. Nature 418:994–998.

Jeffs AR, Benjes SM, Smith TL, Sowerby SJ, Morris CM. 1998. The *BCR* gene recombines preferentially with Alu

elements in complex BCR–ABL translocations of chronic myeloid leukaemia. Hum Mol Genet 7:767–776.
Jin K, Chen J, Nagayama T, Chen M, Sinclair J, Graham SH, Simon RP. 1999. In situ detection of neuronal DNA strand breaks using the Klenow fragment of DNA polymerase I reveals different mechanisms of neuron death after global cerebral ischemia. J Neurochem 72: 1204–1214.
Khodarev NN, Bennett T, Shearing N, Sokolova I, Koudelik J, Walter S, Villalobos M, Vaughan AT. 2000. LINE L1 retrotransposable element is targeted during the initial stages of apoptotic DNA fragmentation. J Cell Biochem 79:486–495.
Kohwi-Shigematsu T, Kohwi Y. 1990. Torsional stress stabilizes extended base unpairing in suppressor sites flanking immunoglobulin heavy chain enhancer. Biochemistry 29:9551–9560.
Kokileva L. 1988. Endogenous degradation of rat liver chromatin studied by agar gel electrophoresis of nuclei. Mol Biol Rep 13:139–143.
Lagarkova MA, Iarovaia OV, Razin SV. 1995. Large-scale fragmentation of mammalian DNA in the course of apoptosis proceeds via excision of chromosomal DNA loops and their oligomers. J Biol Chem 270:20239–20241.
Larsen F, Gundersen G, Lopez R, Prydz H. 1992. CpG islands as gene markers in the human genome. Genomics 13:1095–1107.
Liebich I, Bode J, Reuter I, Wingender E. 2002. Evaluation sequence motifs found in scaffold/matrix-attached regions (S/MARs). Nucleic Acids Res 30:3433–3442.
Liu QY, Pandey S, Singh RK, Lin W, Ribecco M, Borowy-Borowski H, Smith B, LeBlanc J, Walker PR, Sikorska M. 1998. DNaseY: A rat DNaseI-like gene coding for a constitutively expressed chromatin-bound endonuclease. Biochemistry 37:10134–10143.
Liu QY, Ribecco M, Pandey S, Walker PR, Sikorska M. 1999. Apoptosis-related functional features of the DNaseI-like family of nucleases. Ann NY Acad Sci 887:60–76.
Materna T, Weber S, Kofler-Mongold V, Phares W. 1998. Electrophoretic separation of both single- and double-stranded nucleic acids in the same urea-containing agarose gel. Anal Biochem 255:161–163.
Mimori K, Druck T, Inoue H, Alder H, Berk L, Mori M, Huebner K, Croce CM. 1999. Cancer-specific chromosome alterations in the constitutive fragile region FRA3B. Proc Natl Acad Sci USA 96:7456–7461.
Moyzis RK, Torney DC, Meyne J, Buckingham JM, Wu JR, Burks C, Sirotkin KM, Goad WB. 1989. The distribution of interspersed repetitive DNA sequences in the human genome. Genomics 4:273–2789.
Mucha M, Lisowska K, Goc A, Filipski J. 2000. Nuclease-hypersensitive chromatin formed by a CpG island in human DNA cloned as an artificial chromosome in yeast. J Biol Chem 275:1275–1278.
Munteanu MG, Vlahovicek K, Parthasarathy S, Simon I, Pongor S. 1998. Rod models of DNA: Sequence-dependent anisotropic elastic modelling of local bending phenomena. Trends Biochem Sci 23:341–347.
Nagata S. 2000. Apoptotic DNA fragmentation. Exp Cell Res 256:12–18.
Oberhammer F, Wilson JW, Dive C, Morris ID, Hickman JA, Wakeling AE, Walker PR, Sikorska M. 1993. Apoptotic death in epithelial cells: Cleavage of DNA to 300 and/or 50 kb fragments prior to or in the absence of internucleosomal fragmentation. EMBO J 12:3679–3684.
Pandey S, Walker PR, Sikorska M. 1994. Separate pools of endonuclease activity are responsible for internucleosomal and high molecular mass DNA fragmentation during apoptosis. Biochem Cell Biol 72:625–629.
Pandey S, Walker PR, Sikorska M. 1997. Identification of a novel 97 kDa endonuclease capable of internucleosomal DNA cleavage. Biochemistry 36:711–720.
Papadopoulos PC, Greenstein AM, Gaffney RA, Westbrook CA, Wiedemann LM. 1990. Characterization of the translocation breakpoint sequences in Philadelphia-positive acute lymphoblastic leukemia. Genes Chromosomes Cancer 1:233–239.
Razin SV, Petrov P, Hancock R. 1991. Precise localization of the alpha-globin gene cluster within one of the 20- to 300-kilobase DNA fragments released by cleavage of chicken chromosomal DNA at topoisomerase II sites in vivo: Evidence that the fragments are DNA loops or domains. Proc Natl Acad Sci USA 88:8515–8519.
Roychoudhury R, Wu R. 1980. Terminal transferase-catalyzed addition of nucleotides to the 3′ termini of DNA. Methods Enzymol 65:43–62.
Sakahira H, Enari M, Nagata S. 1998. Cleavage of CAD inhibitor in CAD activation and DNA degradation during apoptosis. Nature 391:96–99.
Sarai A, Mazur J, Nussinov R, Jernigan RL. 1989. Sequence dependence of DNA conformational flexibility. Biochemistry 28:7842–7849.
Stallings RL, Torney DC, Hildebrand CE, Longmire JL, Deaven LL, Jett JH, Doggett NA, Moyzis RK. 1990. Physical mapping of human chromosomes by repetitive sequence fingerprinting. Proc Natl Acad Sci USA 87: 6218–6222.
Stanulla M, Wang J, Chervinsky DS, Thandla S, Aplan PD. 1997. DNA cleavage within the MLL breakpoint cluster region is a specific event which occurs as part of higher-order chromatin fragmentation during the initial stages of apoptosis. Mol Cell Biol 17:4070–4079.
Stanulla M, Schunemann HJ, Thandla S, Brecher ML, Aplan PD. 1998. Pseudo-rearrangement of the *MLL* gene at chromosome 11q23: A cautionary note on genotype analysis of leukaemia patients. Mol Pathol 51:85–89.
Strissel PL, Strick R, Rowley JD, Zeleznik-Le NJ. 1998. An in vivo topoisomerase II cleavage site and a DNase I hypersensitive site colocalize near exon 9 in the MLL breakpoint cluster region. Blood 92:3793–3803.
Szabó G, Jr. 1995. 50-kb chromatin fragmentation in the absence of apoptosis. Exp Cell Res 221:320–325.
Szabó GJ, Bacsó Z. 1996. Chromatin isolated from viable human PBLs contains DNA fragmented to >50 kb. Cell Death Differ 3:237–241.
Szabó G, Jr., Boldog F, Wikonkal N. 1990. Disassembly of chromatin into approximately equal to 50 kb units by detergent. Biochem Biophys Res Commun 169:706–712.
Takemoto T, Zhang QM, Matsumoto Y, Mito S, Izumi T, Ikehata H, Yonei S. 1998. 3′-blocking damage of DNA as a mutagenic lesion caused by hydrogen peroxide in *Escherichia coli*. J Radiat Res (Tokyo) 39:137–144.
Ulyanov NB, James TL. 1995. Statistical analysis of DNA duplex structural features. Methods Enzymol 261:90–120.
Varga T, Szilagyi I, Szabó G, Jr. 1999. Single-strand breaks in agarose-embedded chromatin of nonapoptotic cells. Biochem Biophys Res Commun 264:388–394.

Vlahovicek K, Pongor S. 2000. Model.it: Building three dimensional DNA models from sequence data. Bioinformatics 16:1044–1045.

Widlak P, Li P, Wang X, Garrard WT. 2000. Cleavage preferences of the apoptotic endonuclease DFF40 (caspase-activated DNase or nuclease) on naked DNA and chromatin substrates. J Biol Chem 275:8226–8232.

Wu YC, Stanfield GM, Horvitz HR. 2000. NUC-1, a *Caenorhabditis elegans* DNase II homolog, functions in an intermediate step of DNA degradation during apoptosis. Genes Dev 14:536–548.

Yang SW, Burgin AB, Jr., Huizenga BN, Robertson CA, Yao KC, Nash HA. 1996. A eukaryotic enzyme that can disjoin dead-end covalent complexes between DNA and type I topoisomerases. Proc Natl Acad Sci USA 93: 11534–11539.