High resolution mapping DNAs by R-loop atomic force microscopy

Dmitry V. Klinov1, Irina V. Lagutina1, Valery V. Prokhorov1, Tatyana Neretina1, Pavel P. Khil1, Yuri. B. Lebedev1, Dmitry I. Cherny2, Victor V. Demin1 and Eugene D. Sverdlov1,2,*

1Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry and 2Institute of Molecular Genetics, Russian Academy of Sciences, Miklukho-Maklaya 16/10, Moscow 117871, Russia

Received July 2, 1998; Accepted August 10, 1998

ABSTRACT

R-loops formed by short RNA transcripts have been imaged by atomic force microscopy (AFM) at a constant force in the height mode. The technique was applied to mapping the human endogenous retrovirus K10 family (HERV-K10) long terminal repeats (LTR) within individual plasmids and cosmids. RNA probes specific for the U5 (384 nt) and U5 (375 nt) LTR regions separated by a span of 200 bp were used for R-loop formation with LTRs located within plasmid (3.8 kb) or cosmid (∼40 kb) DNAs. R-loops stabilized by glyoxal treatment and adsorbed onto the mica surface in the presence of magnesium ions looked like looped out segments of RNA:DNA hybrids. The total yield of R-loops was usually ∼95%. The RNA:DNA hybrids were found to be 12–15% shorter than the corresponding DNA duplex. The two regions of the LTR could be easily discerned in the AFM images as clearly separated loops. R-loop positions determined on cosmids by AFM were accurate to ∼0.5% of the cosmid length. This technique might be easily adapted for mapping various sequences such as gene exons or regulatory regions and for detecting insertions, deletions and rearrangements that cause human genetic diseases.

INTRODUCTION

A fast and reliable technique for mapping particular sequences and sequence-specific complexes on long DNA stretches is essential for structure-functional analysis of genomes. Such a technique applied to precise localization of defined sequences within long chromosome fragments cloned in cosmid or YAC vectors would facilitate the mapping of various regulatory elements relative to gene coding regions. It would also be helpful in search of mutations such as small deletions, insertions or rearrangements as well as in the analysis of the distribution of interspersed repeated elements like SINE, LINE or endogenous retroviral sequences within a DNA fragment.

Several approaches were proposed for physical mapping of relatively short (5–40 kb) DNA fragments representing typical inserts of standard plasmid or cosmid clones. Mapping of restriction sites over DNA segments is currently the major point of such assays. Typically, the digestion products are separated by electrophoresis and the distribution of bands is used to find the location of the restriction sites within the fragment under study. It provides resolution from tens to thousands of base pairs depending on the enzyme used and peculiarities of the analyzed DNA fragment. The technique is limited by the quality of electrophoretic separation, which is insufficient for long DNA fragments when time-consuming and low-resolution pulse-field electrophoresis is required. Moreover, only restriction sites can be mapped in this way. Physical mapping is also performed by traditional electron microscopy (EM) of individual DNA molecules with sequence-specific ligands bound to them (1). The resolution here depends on the ligand size (2–6) (for earlier works see 7) and is usually of the order of several hundred base pairs.

Recently, a new powerful tool for DNA visualization called atomic force microscopy (AFM) was developed (8,9). Similar to EM, it provides a view of individual DNA molecules and can be used for mapping complexes of DNA with sequence-specific ligands (8,10–17). Although the resolutions of AFM and EM are, at present, roughly the same, AFM seems to have several evident advantages over EM: the samples can be prepared directly from salt solutions which allows flexibility in the choice of salt concentration and incubation temperature, direct adsorption of molecules onto mica is more convenient than in standard procedures of EM sample preparation and a sample can be imaged under conditions where its native structure is retained. The basic AFM equipment is rather inexpensive and occupies much less laboratory space as compared to EM equipment. In addition, the resolution of AFM instruments is expected to be significantly improved in the future and this will probably solve the problem of fast mapping multitude of DNA features. This paper deals with AFM detection of R-loops formed by RNA molecules with their complementary strands in double-stranded DNA. The technique was widely used in EM, but previously never applied to AFM mapping. We developed a reliable procedure allowing one to readily visualize the loops by AFM and report here its successful application to mapping long terminal repeats (LTRs) of the human endogenous retrovirus (HERV) on plasmid or cosmid DNAs. HERVs of different families are

*To whom correspondence should be addressed: Institute of Bioorganic Chemistry, Russian Academy of Sciences, Miklukho-Maklaya 16/10, Moscow 117871, Russia. Tel: +7 095 330 6529; Fax: +7 095 330 6538; Email: eds@glasnet.ru

The authors wish it to be known that, in their opinion, the first three authors should be regarded as joint First Authors.
scattered throughout the human genome and occupy ~1% of its length (18,19). They are thought to have been inserted into the germline 10–60 million years ago (18,19) and could cause significant and evolutionary important changes in expression patterns of neighboring genes, due to the presence of various transcription regulatory elements in their structures, such as promoters, enhancers, hormone-responsive elements and polyadenylation signals. LTRs of endogenous retroviruses were shown to be involved in transcription regulation of cellular downstream genes (20–25). Precise mapping of the retroviral elements and their sequence analysis might be of help in revealing their origin and evolutionary history, whereas identification of the nearby genes could provide a deeper insight into their role in the genome evolution, function and pathology.

MATERIALS AND METHODS

Isolation and purification of DNAs

Cosmid clones 27877 and 30267 from a human chromosome 19-specific library (26,27) were provided by Dr L. Ashworth, and plasmids pGEM39 and pGEM26 (Fig. 1A and C, respectively) were donated by Dr L. Nikolaev. pGEM39 contained a 841 bp PCR fragment of HERV-K10 LTR, amplified with primers LTRfor (5′-TATAGAAAAGTAGACATAGAGACT-3′) and LTRrev (5′-AAAACACACAGACAAAGTATAGA-3′) and cloned into SmaI site of pGEM-7Zf (+) vector (Promega). In SmaI site of the same vector, pGEM26 contained a 328 bp PCR fragment from the U3 region of the LTR amplified with primers LTRfor and S90rev (5′-GCCATATTTCCATATCATCAGTGG-3′) using pGEM39 as a template. Plasmid and cosmid DNAs were routinely isolated from Escherichia coli strain XL1-blue using Qiagen plasmid midi kit (Qiagen). Prior to R-loop formation, plasmids pGEM39 and pGEM26 were digested with SacI (having a unique site in the vector), cosmid 27877 with EcoRI (6 sites) and cosmid 30267 with NruI (a unique site in the vector) restriction endonucleases. Digested DNAs were purified by phenol extraction/ethanol precipitation.

Cosmid restriction mapping

Cosmid 27877 was digested with EcoRI and/or BamHI fragments were separated on a 0.8% agarose gel and alkali blotted onto Zeta-probe nylon membrane (Bio-Rad) according to the manufacturer’s recommendations. Filters were hybridized with LTRfor-LTRrev PCR fragments of the HERV-K LTR, random-primer labeled using an [α-32P]dATP. Labeling was done with Prime-a-Gene labeling system (Promega) as recommended by the manufacturer.

Oligonucleotide synthesis

Oligonucleotides were synthesized with an ASM-102U DNA synthesizer (Russia) using commercial CED phosphoramidites and purified by electrophoresis in a 20% acrylamide gel.

RNA probe preparation

RNA probes were synthesized using the Riboprobe Combination System (Promega) according to the manufacturer’s recommendations. RNAs were generated by means of in vitro transcription with T7 RNA polymerase on PCR fragments containing T7 promoter sequences, and purified by acidic phenol extraction. Prior to the transcription, PCR-amplicons were purified by phenol extraction/ethanol precipitation. The following templates were used for RNA probe production: (i) PCR-amplicons obtained with primers T7 (5′-GAATAACGACTCACTATAGGG-3′) and Rfor577 (5′-GCTTAGGAGGTCAGGAGTG-3′) from plasmid pGEM39 as a template for U5 probe, (ii) PCR-amplicons prepared with primers T7 and LTRfor from plasmid pGEM26 as a template for U3 probe (Fig. 1) and (iii) PCR-amplicons obtained with the use of cosmid DNAs. This last probe was prepared in the following way: cosmid 30267 was PCR-amplified with phosphorylated primers LTRfor and LTRrev. The PCR product obtained was blunt-ended with T4 DNA polymerase (Promega) and purified by phenol/chloroform extraction followed by ethanol precipitation. The phosphorylated and blunt-ended PCR product was then ligated with T7Not1/Not2 adapter (T7Not1: 5′-GAATAACGACTCACTATAGGGCAGCTGTTGTCGGCC- CGAGGT-3′; Not2 5′-ACCTGCGG-3′; T7 promoter sequence is shown in bold). The ligation mixture contained 4 ng/µl of the PCR amplimcon, 0.2 µl of adapter, 1 µl ATP, 50 mM Tris–HCl pH 7.6, 10 mM MgCl2 and 0.25 U/µl T4 DNA ligase (MBI Fermentas). After overnight incubation at 16°C the mixture was diluted 10-fold and a 1 µl aliquot thereof was used for PCR amplification with primers T7 and Rfor577 (for U5 probe) or T7 and S90rev (for U3 probe).

Formation of R-loops and DNA:RNA hybrids

The hybridization mixture for R-loop formation on plasmid DNAs contained 50 ng of a linearized plasmid, 25 ng of the RNA probe (in the case of R-loop formation with two probes 25 ng of each), that corresponded to the probe:template molar ratio of ~10:1, in 20 µl of hybridization buffer (400 mM NaCl, 100 mM PIPES, pH 7.2 and 10 mM EDTA in 70% deionized formamide). The mixture was incubated at 56°C for 60 min and then slowly (90 min) cooled to 43°C. The hybridization mixture for R-loop formation on cosmid DNAs contained 100 ng of the linearized cosmid DNA, 10 ng each of the RNA probes (that gave the probe:template molar ratio of ~20:1) in 20 µl of the hybridization buffer. The mixture was incubated for 60 min at 56°C and then slowly (90 min) cooled to 37°C overnight. Short DNA:RNA hybrids for length measurements were prepared by incubation of 100–150 ng of the PCR fragment obtained with primers T7-Rfor577 or T7-LTRfor and 50 ng of the corresponding DNA probe (at the DNA/RNA molar ratio of 1:1 to 1.5:1) in 20 µl of the hybridization buffer for 2 h at 70°C and then 10 min at 56°C followed by slow (90 min) cooling to 43°C. The reaction mixtures were treated with glyoxal according to Kabak et al. (28). After completing the reaction the hybridization mixture was quickly chilled to 0°C and glyoxal was added to a final concentration of 1.0 M (17 volume of 40% glyoxal purified on the ion exchange resin Amberlite MB-1). The solution was further incubated at 12°C for 2 h. The hybridization mixture was precipitated with ethanol, resuspended in 20 µl of 20 mM NH4OAc and used for preparation of the samples for AFM. In the preliminary experiments glyoxal treatment was not used.

AFM

A 10 µl aliquot containing 20–40 ng DNA was placed on the freshly cleaved mica surface for 0.5–1 h in the buffer containing 20 mM NH4OAc and 7–9 mM MgCl2. After completing the adsorption, the mica sheet was washed with 0.1% aqueous uranyl acetate.
acetate, blotted with filter paper and dried with hot air. The samples were imaged with a NanoScope II (Digital Instruments, Santa Barbara, CA) operating at constant force in the height mode using commercial silicon nitride tips with the apex curvature radius of 50 nm. During operations the samples were constantly blown with hot nitrogen at a temperature over the scanning area within the range of 70–110°C. The scanning rate was 7 Hz with the scanning angles 180 and 90°. The images were filtered using the flatten filter of the Nanoscope II software. The analysis of the plasmid or cosmid images included determination of absolute (nm) and relative (%) positions of R-loop ends on the DNA. A series of length measurements of the PCR fragments and RNA:DNA hybrids formed by these fragments with the homologous RNA probes was implemented. The difference in lengths between DNA:RNA hybrids and corresponding DNA:DNA duplexes, determined by AFM under the conditions used in our experiments, were evaluated by measurements for ∼1000 fragments of each type. The length measurements were done using a software written and kindly provided by Dr G. Glushchenko (Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry), start- and endpoints for each molecule were entered manually. The reproducibility of the curvilinear measurements between the points was within 1% of the length.

To determine the positions of the R-loops, the nucleotide scale of the R-looped DNAs was calculated counting that 1 bp is equivalent to 0.35 nm in DNA segments, while for R-loops a 15% difference in lengths of DNA:DNA and DNA:RNA duplexes (Results) was taken into account. Asymmetrical location and high specificity of R-loops greatly facilitated the alignment of molecules.

RESULTS

AFM visualization of R-loops formed by LTRs within plasmid DNAs

To demonstrate the possibility of positioning short stretches of DNA within longer sequences by AFM of R-loops, we mapped LTR sequences of the human endogenous retrovirus HERV-K10 (29) in DNAs of plasmid size, i.e. ∼4 kb. For this purpose, a nearly
full-length 841 bp fragment of the LTR was cloned into SmaI site of a pGEM-7Zf(+) plasmid. Two probes corresponding to the U5 and U3 regions, respectively, and sufficiently long (≈300 bp) to permit the quantitative formation of R-loops (30,31), were prepared by a two-step procedure. At the first step the desired area was PCR amplified with two primers, one of them targeted at the LTR sequence selected while the other at the T7 promoter part of the vector. The positions of the selective primers are shown in Figure 1. The fragments amplified contained the T7 promoter and thus could be transcribed with T7 RNA polymerase. The transcription was performed at the second stage yielding U5 and U3 RNA probes of 375 and 384 nt in length, respectively. At their 5'-termini both transcripts contained 56 extra nucleotides corresponding to a part of the vector polylinker. These extra sequences could take part in the R-loop formation with a pGEM39 plasmid in the case of U5 but not U3 RNA probe (Fig. 1). Therefore the effective length of U3 probe was 328 nt. GC-content of the U3 and U5 regions is equal to 43.5 and 49%, respectively, which gives calculated T\text{m} = 55 and 57°C (30). The R-loops were prepared with a pGEM39 plasmid linearized with the restriction endonuclease ScaI. The LTR borders on the linearized plasmid corresponded to positions 1185–2025 (Fig. 1). The R-loops were prepared under hybridization conditions used for EM visualization (28,30,32) as described in Materials and Methods.

The positions of the 5' and 3' boundaries of the ‘bear ear’ R-loops for the U5 region of the LTR were determined. The resulting histogram for glyoxal treated samples is presented in Figure 3A. We found that the lengths of DNA:RNA hybrids were ~12–15% smaller than the lengths of the corresponding DNA:DNA duplexes under the conditions used for mounting the molecules for AFM viewing, i.e. in the presence of magnesium ions. The averaged values of R-loop boundaries were 1118 ± 20 and 1502 ± 20 bp, while the true values should be 1128 and 1503 bp, respectively. The R-loop 3' border was shifted towards the 5'-end, when the glyoxal treatment was omitted whereas the 5' boundary position remained unaffected (data not shown).

Simultaneous mapping of the U3 and U5 LTR regions

The DNA molecules with two R-loops also seen as blobs and/or ‘bear ear’ were formed by hybridization of pGEM39 plasmid DNA simultaneously with two RNA probes complementary to the U3 and U5 LTR regions (Fig. 2C). Again, the total yield of R-loops was close to 95% irrespective of the glyoxal treatment, but the latter essentially increased the yield of ‘bear ear’-like structures (Table 1). Only those molecules that had ‘bear ear’ R-loops were used for mapping (Fig. 3B). DNA:RNA hybrids treated with glyoxal were also in this case ~14% shorter than their counterpart DNA:DNA duplexes of the plasmid DNA. Finally, the averaged positions of the R-loop boundaries were determined as 1110 ± 20, 1468 ± 20, 1699 ± 20 and 2008 ± 20 bp, whereas according to the nucleotide sequence the borders are located at 1128, 1503, 1697 and 2025 bp, respectively.

A similar map for the molecules not treated with glyoxal was also obtained and revealed the same effect of the treatment on the U5 boundary position as for U5 RNA probe alone. On the contrary, the shift of the boundary observed in the non-treated samples for the U3 R-loop affected mostly the 5'-end whereas the 3'-terminus positions were the same irrespective of the treatment (data not shown). The reason for this phenomenon remains to be explained.

| Table 1. Yields of different types of R-loops depending on the sample preparation conditions |
|---------------------------------|-------------|----------------|-------------|----------------|---------------|---------------|
| DNA with:                       | number of   | with one       | with two     | with blobs   | without      | adsorption    |
|                                 | molecules   | R-loop (%)     | R-loops (%)  | (%)          | R-loops (%)  | time (h)      |
| U5 RNA probe without glyoxal    | 310         | 28.1           | –            | 65.2         | 6.8          | 0.5           |
| treatment                        |             |                |              |              |              |               |
| U5 and U3 RNA probes            | 178         | 8.4 (U3)       | 39.3         | 30.3         | 2.2          | 0.5           |
| without glyoxal treatment       |             |                |              |              |              |               |
| U5 RNA probe with glyoxal       | 86          | 63             | –            | 31           | 6            | 1             |
| treatment                        |             |                |              |              |              |               |
| U5 and U3 RNA probes with       | 127         | 19             | 73           | 4            | 4            | 1             |
| glyoxal treatment               |             |                |              |              |              |               |

condensed DNA segments (Fig. 2A and C, Table 1). The latter structures were termed ‘bear ear’. DNA length measurements revealed that both the blobs and the ‘bear ears’ seemed to be located at the positions expected for the U5 LTR on the plasmid DNA. It suggested that the observed DNA chain perturbations were the consequence of the R-loop formation.

Glyoxal treatment was used to increase the stability of the R-loops and therefore the yield of ‘bear ear’-like structures in subsequent experiments (Discussion). This treatment did not affect the total yield of R-loops determined as the sum of blobs and ‘bear ears’, usually close to 95%, but substantially increased the yield of the ‘bear ear’ structures (Table 1).
Figure 2. AFM images of R-loops. (A) R-loops formed by hybridization of pGEM39/ScaI plasmid DNA with U5 RNA probe treated with glyoxal. Insert, magnified view of an R-loop in the form of ‘bear ear’. Scale bar = 200 nm. (B) The same as in (A), but without glyoxal treatment. Insert, magnified view of a blob. Scale bar = 200 nm. (C) R-loops formed by hybridization of pGEM39/ScaI plasmid DNA with both U5 and U3 probes and treated with glyoxal. Scale bar = 200 nm.

Cosmid mapping by R-loop AFM

Two overlapping cosmids 27877 and 30267 from a human chromosome 19 specific library (26,27), mapped to 19q13.1 were used for LTR mapping by AFM. Both cosmids contained one copy of identical HERV-K LTRs (34,35). An EcoRI restriction map is available for the cosmids (Fig. 4C) as well as for a considerable portion of chromosome 19 (27, current version of chromosome 19 restriction map is freely available at http://www-bio.llnl.gov/rmap). Since differences between various LTR structures can be as high as 30% (35,36), to achieve the most efficient R-loop formation we produced RNA probes specific for the cosmid LTR. For this purpose a near full-length LTR fragment was amplified directly from the corresponding cosmid with primers LTRfor and LTRrev. The amplicon was subsequently ligated to the T7Not1 adapter, containing T7 promoter sequences. The tagged amplicon as a new template and primers T7 and 590rev, a subfragment containing a part of the U3 region was amplified to yield a template for subsequent transcription of U3 probe. U5 probe was prepared by transcription of another subfragment of the same amplicon, reamplified with primers T7 and Rfor577. In both cases the amplification of the whole fragment with the only T7 primer complementary to the adapter was precluded due to a special design of the T7Not1 adapter. When attached to a dsDNA fragment, the adapter forms 40 bp long inverted terminal repeats with high GC-content in its inner part. This feature provides the formation of pan-handle like structures with long double-stranded stems at single-stranded DNA termini during the PCR annealing stage that blocks the binding of the T7 primer (the PCR-suppression effect; 37,38) and the amplification with only this primer. In this system PCR is possible only after addition of the second primer targeted at an inner single-stranded part of the fragment. The probes obtained in such a way contained the LTR regions corresponding to those described above for plasmids.

R-loop mapping of the LTR was performed on both the 30267 and 27877 cosmids, the latter being digested with EcoRI. The R-loops were formed mainly as described for plasmids but at a somewhat higher probe:template ratio (20:1) and with overnight incubation of the hybridization mixture at 37°C. The yield of R-loops was reasonably high; as many as 90% of the molecules contained R-loops. The coefficient 0.35 nm/bp (determined from measurements of several molecules of known lengths) was used to convert the length into its base pair equivalent for calculations of the LTR positions within the cosmids.

For the cosmid 27877, ten EcoRI fragments with double ‘bear ear’ R-loops were analyzed. The lengths of these fragments were ~17.5 kb corresponding to the longest of the EcoRI fragments of
Figure 3. Histograms of R-loop boundary distribution. Histograms were calculated for AFM images of the R-loops formed by (A) U5 RNA probe and (B) both U5 and U3 RNA probes stabilized by glyoxal treatment. Positions of the R-loop boundaries were calculated according to physical lengths of molecules in percent. Corresponding positions in nucleotides are indicated below. Due to the difference in nucleotide density per unit length of DNA:DNA and DNA:RNA hybrids these figures should be considered as approximate. Vertical lines indicate averaged R-loop boundary positions (in bp from ScaI site). The numbering corresponds to that in Figure 1.

For the full-length linearized cosmid 30267, the analysis of 10 molecules (Fig. 4B) put the R-loop boundaries at 8400 ± 160/8720 ± 160 and 8920 ± 160/9190 ± 210 bp from the nearest cosmid terminus. Which of the two cosmid termini is proximal to the LTR was determined by comparison of these figures with those obtained for the R-looped EcoRI fragment of the 27877 cosmid. The longest EcoRI fragment of the 27877 cosmid bearing the LTR corresponds to the left terminus of the cosmid 30267, as shown in Figure 4C. The data allowed us to find the positions of the LTR borders with respect to the nearest EcoRI site in both cosmids. For the cosmid 30267 these positions were 3960 ± 90/4230 ± 70 and 4450 ± 80/4680 ± 80 bp from the nearest terminus of the fragment, respectively (Fig. 4A and C).

For the full-length linearized cosmid 30267, the analysis of 10 molecules (Fig. 4B) put the R-loop boundaries at 8400 ± 160/8720 ± 160 and 8920 ± 160/9190 ± 210 bp from the nearest cosmid terminus. Which of the two cosmid termini is proximal to the LTR was determined by comparison of these figures with those obtained for the R-looped EcoRI fragment of the 27877 cosmid. The longest EcoRI fragment of the 27877 cosmid bearing the LTR corresponds to the left terminus of the cosmid 30267, as shown in Figure 4C. The data allowed us to find the positions of the LTR borders with respect to the nearest EcoRI site in both cosmids. For the cosmid 30267 these positions were 3960 ± 90/4230 ± 70 and 4450 ± 80/4680 ± 80 bp from the nearest terminus of the fragment, respectively (Fig. 4A and C).

The LTR mapping on these cosmids performed by a standard Southern blotting technique showed that the LTR was located within the 17.5 kb EcoRI fragment of the cosmid 27877 (not shown), in agreement with the AFM data. Therefore, to improve resolution, we have constructed a BamHI restriction map of this 17.5 kb fragment. As far as BamHI site is conserved and placed asymmetrically in the subfamily II LTRs, to which this particular LTR belongs (35,36), it was possible to precisely map the LTR with respect to BamHI sites and to determine its orientation. The larger part of the LTR sequence turned to reside within the 4.5 kb BamHI–EcoRI fragment, while the rest was in the 1.3 kb BamHI–BamHI fragment (Fig. 4C; lower line). Accordingly, the U3 and R regions of the LTR should be in the 4.5 kb fragment, the U5 region being closer to the nearest terminus of the cosmid than the U3. The data also put the borders of the LTR area involved in the R-loop formation at 3750 and 4600 bp from the nearest EcoRI site with estimated accuracy of ~5%. Thus the data obtained with AFM technique and Southern blot analysis are in agreement.

DISCUSSION

The R-loop observations were proposed first for DNA:RNA heteroduplex EM analysis. An efficient method for R-loop formation was introduced by Thomas et al. (30). The authors demonstrated the ability of RNA to displace the identical DNA strand and quantitatively hybridize with the complementary strand in the presence of 70% formamide (28,32). Once formed, the R-loops display considerable stability in the presence of formamide, but its removal causes rapid RNA displacement from the R-loops. The displacement is more pronounced for the R-loops formed by short RNAs. The R-loops could be stabilized by glyoxal treatment (28). The yield of R-loops was reported to be ~100% for long RNAs. For shorter RNA molecules (of ~100 nt) the yield was less even at the excess of RNA, but could be enhanced up to 100% by means of glyoxal treatment (31). Having
Figure 4. AFM images of R-loops and restriction map of cosmids. (A) An AFM image of the R-loops formed by hybridization of 17.65 kb EcoRI-restriction fragment of the cosmid 27877 DNA with two RNA probes corresponding to the U3 and U5 regions of the LTR. R-loops in the form of ‘bear ear’ are indicated by arrows. Scale bar = 200 nm. (B) An AFM image of the R-loops formed by hybridization of the cosmid 30267 DNA linearized with NruI restriction enzyme with the same RNA probes. Insert, magnified view of R-loops in the form of ‘bear ear’. Scale bar = 200 nm. (C) A restriction map of the inserts of 27877 and 30267 cosmids. Positions of EcoRI (E) and two BamHI (B) sites are indicated by short vertical lines. The lengths of the corresponding restriction fragments in kb are indicated over the fragments. Distances (in bp) from the R-loop borders to EcoRI sites are indicated over corresponding double-arrows. The LTR location within the 17.65 kb fragment of the cosmid 27877, determined by restriction mapping, is shown in the lower part. The U5, R and U3 regions of the LTR, positions of BamHI sites and sizes of the corresponding restriction fragments are indicated.

This information as a starting point, we developed a procedure for preparation of DNA samples containing specific R-loops for AFM. As high as 95% yield of R-loops could be achieved in all experiments. The R-loops formed on DNA by short RNA probes (~300 nt in length) could be unambiguously discerned by AFM and mapped with the accuracy of ~0.5% of the DNA length.

Two types of clearly discernible R-loop structures were observed: ‘blob’ and ‘bear ear’. Their relative amounts were affected by treatment of AFM samples with glyoxal. An unusual structure of the ‘bear ear’ R-loop form might be explained by coiling of the displaced single-stranded DNA chain, giving rise to a short straight part of the ‘ear’. The arch-like part of the R-loop probably represents the RNA:DNA heteroduplexes. It was reported that DNA:RNA hybrids mounted by cyt-c method for electron microscopy at a high concentration of formamide were either equal in length to the counterpart DNA:DNA duplexes (39) or ~12–15% shorter (40). X-ray studies on DNA:RNA hybrid fibers also showed considerable shortening of DNA:RNA heteroduplexes with respect to their DNA:DNA counterparts (41). However, no data were available on the DNA:RNA hybrid lengths, when adsorbed onto the mica surface with magnesium ions. The control length measurements of specially prepared DNA:RNA hybrids revealed their ~15% shortening as compared to the corresponding DNA:DNA homoduplexes. Likewise, the heteroduplex part of the R-loops within plasmid DNA was also 12.5–14% shorter than its DNA:DNA analogue.

One of the reasons for the ‘blob’ formation might be a partial RNA displacement from the RNA–DNA heteroduplexes by the...
complementary DNA chain. The R-loop heteroduplex part shortened due to the displacement can take the shape of a blob. This correlates well with an increase in the blob content, when glyoxal fixation of the samples is omitted (Fig. 2B). Another effect which can be related to the partial RNA displacement is the observed difference in the positions of the R-loop boundaries depending on glyoxal treatment (Results).

We applied the technique to mapping of two partially characterized overlapping cosmide clones, 30267 and 27877, that contained DNA of human chromosome 19 with a HERV-K10 LTR therein. Recently, we mapped 72 HERV-K LTRs on this chromosome and demonstrated frequent coincidences of the LTR positions with those of some known genes (34,35). The neighborhood of an LTR can influence the gene expression. Therefore we are interested in more precise determination of relative positions of genes and LTRs than can be achieved in a standard restriction endonuclease analysis. In particular, the cosmid 27877 used in the present work was shown to contain the SUTP3H gene (GenBank accession no. U56402, current version of chromosome 19 restriction map is freely available at http://www-bio.llnl.gov/rmap) and a novel trachea-specific transcribed sequence (Khil et al., manuscript in preparation). By the time of this work we had determined that the LTR was located within the 17.5 kb EcoRI fragment of the cosmid. To find its precise position we used R-loop mapping. Figure 4 shows the R-loops visualized by AFM on the cosmid with the simultaneous use of U3- and U5-specific RNA probes. One can see that the loop borders could be precisely mapped on the cosmid. The accuracy of the measurement was confirmed by AFM analysis of the LTR position in two overlapping cosmids as well as by Southern blot analysis. Precision of the AFM R-loop mapping on large DNA fragments is good enough to enable rapid positioning of various structural genomic elements relative to each other. In particular, our preliminary data suggest that LTRs can be mapped simultaneously with cDNA using hybridization with the corresponding pair of RNA probes. The procedure can be adapted for mapping a variety of sequence features such as coding or regulatory regions isolated by means of a functional assay. This technique could be also used for detection of different kinds of insertions, deletions and rearrangements that cause human genetic diseases.

ACKNOWLEDGEMENTS

The authors thank N. Skaptsova and Dr V. Potapov for oligonucleotide synthesis, Dr L. Nikolaev for providing plasmids, Dr G. Glushchenko for the computer program used for image analysis and Dr B. Glotov for helpful discussion. The work was partially supported by a grant of Human Genome State Project of Russia, International Research Scholar award from HHMI (75195-544201), INTAS grant INTAS-96-1710.

REFERENCES