Construction of a chromosome specific library of human MARs and mapping of matrix attachment regions on human chromosome 19

Lev G. Nikolaev*, Tsogtkhishig Tsevegiyn, Sergey B. Akopov, Linda K. Ashworth1 and Eugene D. Sverdlov

Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry, Russian Academy of Sciences, Miklukho-Maklaya 16/10, Moscow 117871, Russia and 1Human Genome Center, L-452, Lawrence Livermore National Laboratory, Livermore, CA 94550, USA

Received November 28, 1995; Revised and Accepted February 15, 1996 GenBank accession nos +

ABSTRACT

Using a novel procedure a representative human chromosome 19-specific library was constructed of short sequences, which bind preferentially to the nuclear matrix (matrix attachment regions, or MARs). Judging by 20 clones sequenced so far, the library contains > 50% of human inserts, about 90% of which are matrix-binding by the in vitro test. Computer analysis of sequences of eight human MARs did not reveal any significant homologies with the EMBL Nucleotide Data Base entries as well as between MARs themselves. Eight MARs were assigned to individual positions on the chromosome 19 physical map. The library constructed can serve as a good source of MAR sequences for comparative analysis and classification and for further chromosome mapping of MARs as well.

INTRODUCTION

Libraries of short sequences with certain specific functions is an important instrument for functional mapping of the human genome (1) considered to be the next step in the Human Genome Project once sufficiently high resolution physical maps of human chromosomes are compiled (2–6). The most advanced area of functional mapping is apparently the mapping of transcribed sequences, namely mRNA (7–11) and heterogeneous nuclear RNA (hnRNA) (12–14). This allows one to position protein coding and/or transcribed sequences on the chromosomes and is finally aimed at the chromosome localization of all estimated 50–100 thousand human genes (15). However, apart from the protein coding and transcribed sequences there exist many kinds of other unique sequences also having specific functions, such as promoters, enhancers, protein-recognizing sequences etc. Positioning of these sequences on the human chromosomes could significantly expand the scope of the human genome functional map.

The sequences preferably binding to the nuclear matrix or scaffold (matrix or scaffold attachment regions, MARs or SARs) are an example of such functional sequences. Being polymorphic and relatively evenly scattered over the human genome these sequences can serve as chromosome markers (16). Such MARs (SARs) were previously characterized as relatively short (100–1000 bp) DNA sequences involved in anchoring chromatin loops (domains) to the protein network termed nuclear matrix in interphase or chromosomal scaffold in mitosis (for review see 16–20). The general attributes assigned to MARs have been summarized (16). Among other elements, MARs often include potential origins of replication, relatively long A+T-rich stretches harbouring topoisomerase II binding sites and palindromic sequences. Some classes of MARs contain CT-rich stretches or, in some cases, are enriched in TG-motif. In addition, these sequences are rich in transcription factors binding sites and contain potentially curved or kinked DNA (16). Functionally, MARs are bringing together the presumed active components of the nuclear matrix (topoisomerase II and other enzymes of DNA and RNA metabolism, transcription factors, etc.) and their target chromatin regions which include, but are not limited to, MARs (17–20).

Here we report the construction of a human chromosome 19-specific library of DNA sequences preferably binding to the nuclear matrix in vitro as well as the early results of mapping some of these sequences on chromosome 19.

MATERIALS AND METHODS

Basic protocols

Growth and transformation of Escherichia coli cells, preparation of plasmids and λ DNA, gel electrophoresis, blot-hybridization, and other standard manipulations were performed as described (21,22).

Cell lines

Jurkat cells were grown in suspension using RPMI-1640 medium supplemented with 10% fetal calf serum. DMEM medium with 10% fetal calf serum was used for other cell lines.
Nuclear matrix preparation and *in vitro* binding of MARs

Nuclear matrices were prepared from Jurkat cells by a high salt extraction method (23) or, alternatively, using a lithium 3,5-diodosalicylate (LIS) extraction protocol (24,25). Binding of radioactively labelled MARs to the nuclear matrices *in vitro* was carried out as described (25).

**Preparation of chromosome 19 genomic DNA fragments**

A human chromosome 19 genomic library in a λ-vector Charon 40 (provided by A.V. Carrano, Lawrence Livermore National Laboratory, USA) with average insert of 16 kb was used as a source of chromosome 19 DNA.

Total λ-library DNA, 10 μg, was divided into halves and digested to completion with 20 U of either Sau3A (New England Biolabs) or Csp6I (Fermentas, Lithuania) in 100 μl of the corresponding manufacturer supplied buffer for 2 h at 37°C. The fragments obtained were extracted with phenol–chloroform, ethanol precipitated and dissolved in 30 μl TE buffer (10 mM Tris–HCl pH 8.0, 1 mM EDTA).

To the resulting sticky ends, the following PCR adapters were ligated:

| 5′-GATCCTGTGATG-3′ | adapter Sau3A |
| 3′-ACAGATTCACTTAGCTCTCAGTCA-5′ | library primer |
| 5′-TATGTTCTATG-3′ | adapter Csp6I |

The library primer was identical for both enzymes and contained an XhoI site (underlined).

For ligation, 10 μl of each digested DNA was mixed with 1 nmol of the library primer and 1 nmol of the corresponding adapter in a final volume of 20 μl 50 mM Tris–HCl, pH 7.5, 5 mM MgCl2, 5 mM dithiotreitol, 0.5 mM ATP and 50 μg/ml bovine serum albumin (BSA). The reaction was done with 8 U (Weiss) ligase for 24 h at 13°C for Sau3A digest and 4°C for Csp6I digest. After the ligation, the mixtures were diluted 100-fold with TE buffer, and 1 μl of each was pacarded in 100 μl 10 mM Tris–HCl, pH 8.3, 50 mM KCl, 3 mM MgCl2, 0.25 mM of each dCTP, dGTP and dTTP, 20 mM dATP, 50 μM each dCTP, dGTP and dTTP, 25 μM of the library primer with 2 U Taq DNA polymerase. The amplification profile was 94°C for 0.8 min; 50°C for 0.6 min; 72°C for 1 min for 30 cycles. The resulting DNA products were extracted with phenol–chloroform, precipitated with ethanol, washed, dried, dissolved in 50 μl TE buffer and pooled together.

**Selection of MARs**

Binding of the DNA to nuclear matrices was done according to ref. 23 with modifications. A suspension of nuclear matrices from 5 × 10⁶ cells was washed three times with 1 ml of buffer 1 (10 mM Tris–HCl pH 7.5, 50 mM NaCl, 2 mM EDTA, 0.25 M sucrose, 0.25 mg/ml BSA). Two times concentrated buffer 1 (10 mM Tris–HCl pH 7.5, 300 mM NaCl, 1 mM MgCl2, 0.25 M sucrose, 0.25 mg/ml BSA), then with the same buffer but with 2 M NaCl, and finally with the same buffer containing 50 mM NaCl. The matrices were then resuspended in 50 μl TE buffer, 0.5% SDS, proteinase K was added up to 50 μg/ml and the mixture was incubated for 1 h at 37°C to digest proteins. The mixture with 10 μl 10 M ammonium acetate added was extracted twice with phenol–chloroform and precipitated with 2 vol of ethanol overnight at –20°C. Matrix-bound DNA was then collected in a microcentrifuge for 10 min at 4°C, washed with 75% ethanol, dried and dissolved in 10 μl TE buffer.

This DNA, 5 μl, was again PCR-amplified and purified as described above and then used for the next round of binding to the nuclear matrices under the same conditions. This selection procedure was repeated five times.

**Cloning and arraying of the library**

The PCR-amplified mixture of the putative MARs after the fifth round of binding with nuclear matrices was digested with XhoI, taking advantage of the introduced site, and cloned into XhoI site of a pGEM7zf(+) phagemid vector (Promega). After transformation the XLI blue E. coli cells were plated on X-gal/IPTG agar plates and 384 white colonies were arrayed on 96-well clusters.

**Sequencing of MAR clones**

Single-stranded DNA of the pGEM7zf(+) phagemids containing MAR inserts was prepared using R408 helper bacteriophage according to a Promega protocol (26). The sequencing reaction was done using the M13 direct sequencing primer, [α-32P]dATP (Obninsk, Russia) and Sequenase version 2.0 DNA sequencing kit (US Biochemicals) according to the manufacturer recommendations.

**DNA labelling**

PCR labelling of MAR probes for high density filter hybridization was performed as described earlier (14). Probes for binding with nuclear matrices were prepared as follows. MAR-containing fragments were PCR-amplified in 50 μl reaction volume using the corresponding plasmid template and library primer. Negative control fragments were amplified from the corresponding phage DNA using the following primers (positions in the bacteriophage sequences are in parentheses).

<table>
<thead>
<tr>
<th>λ-phage (162 bp fragment)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Direct primer 5′-TCCGTGAGTTGATGTTGGTG-3′ (916–935)</td>
</tr>
<tr>
<td>Reverse primer 5′-TAGTCCGTCACGTTGGTT-3′ (1077–1058)</td>
</tr>
<tr>
<td>T7 phage (139 bp fragment)</td>
</tr>
<tr>
<td>Direct primer 5′-ACCAGAGAGAGACGTTGACC-3′ (34257–34275)</td>
</tr>
<tr>
<td>Reverse primer 5′-CGCCAGCAGATTCAATTAG-3′ (34395–34376)</td>
</tr>
</tbody>
</table>

A pUCMAR10 positive control was provided by I. Shatsky and PCR-amplified using the M13 direct and reverse 17 nt sequencing primers. The fragments were purified by low melting point agarose gel electrophoresis, their bands cut out from the gel, diluted 3-fold with TE buffer and used as templates for subsequent PCR labelling.

The PCR labelling was conducted in a final volume of 50 μl containing ~10 ng of the template fragment, 10 μM each primer (for controls) or 20 μM library primer (for MAR fragments), 200 μM each dCTP, dGTP and dTTP, 20 μM dATP, 50 μCi [α-32P]dATP (Obninsk, Russia), 3 mM MgCl2, 10 mM Tris–HCl pH 8.3 and 50 mM KCl. The reaction profile was 94°C for 0.8 min; 50°C for 0.6 min; 72°C for 1 min for 20 cycles. Reaction products were extracted twice with phenol–chloroform, precipitated with ethanol and dissolved in TE buffer.
DNA isolated from the nuclear matrices as described above was labelled using the Ready-To-Go labelling mixture (Pharmacia) according to the manufacturer’s protocol.

RESULTS

Library construction

The general scheme of the construction of a chromosome-specific MARs library is shown in Figure 1.

A pool of relatively short (200–1000 bp) chromosome 19 DNA fragments with the library primer on both ends, obtained as described in Materials and Methods, was used for selection of MARs by the in vitro matrix binding assay. Nuclear matrices were mixed with this pool, incubated, extensively washed to remove unbound DNA, and matrix-bound DNA fragments were isolated as described in Materials and Methods. The matrix-bound DNA was PCR-amplified with the library primer and used for the second round of matrix binding. The process was repeated five times. The result of each successive binding step is shown in Figure 2. The initial DNA gave a ladder of fragments mostly originated from the digested λ-vector arms (lane 1), as judged from strong hybridization of the majority of these fragments with the λ-probe (lane 2). After successive matrix binding–purification rounds the ethidium bromide stained ladder gradually disappeared being replaced by smear (lanes 3, 5, 7 and 9), and the hybridization signal disappeared as well (lanes 4, 6, 8 and 10). This indicated selective removal of λ-DNA from the mixture of fragments, indirectly supporting that the residual DNA fragments represented a matrix binding fraction of the total chromosome DNA. After the fifth round of binding with nuclear matrix the PCR-amplified mixture of fragments was digested with XhoI and cloned into XhoI site of a pGEM7zf(+) plasmid vector. After transformation the E.coli cells were plated on X-gal/IPTG agar plates and 384 white colonies were arrayed on 96-well clusters.

Selection of the chromosome 19-specific clones

To verify the absence of λ-derived clones in the library the colonies were transferred onto nitrocellulose filters and hybridized with the λ-probe. Only seven of 384 colonies were λ-positive, indicating that the extent of library enrichment in MARs was >50 times, given an ~1:1 initial ratio of the vector and insert DNA in the mixture (data not shown).

The presence and the size of inserts in all library clones were checked by PCR with the library primer. Twenty clones with 250–500 bp inserts were selected and sequenced. The nucleotide sequences obtained were compared to those of the EMBL Nucleotide Data Base using BLAST algorithm (27). No significant (>80%) homologies with the data bank entries were found. The sequences were analysed on mutual homology using PC/Gene package. Only two clones (M0A3 and M0A12) were >95% homologous. As 19 out of 20 clones were individual, it can be concluded that the library obtained is highly representative. Nevertheless, it could not be excluded that very long or short MARs are underrepresented in the library due to the PCR reaction bias.

The source λ-library of human chromosome 19-specific DNA was constructed by sorting chromosomes isolated from human–hamster hybrid UV5HL9-5B cells containing chromosome 19 as a sole human component. Due to limitations of the sorting, a considerable number of the library clones were of hamster origin. It was necessary, therefore, to check the human chromosome 19 origin of the clones before mapping. To this end internal pairs of primers specific for each clone were designed using Primer software (version 0.5, Whitehead Institute for Biomedical Research) and synthesized. Sequences of the primers are presented in Table 1. These primer pairs were subsequently used in PCR with DNA isolated from human, hamster and hybrid
Figure 3. An example of PCR amplification using DNA from human (1), hybrid (2) and hamster (3) cell lines as templates and primer pairs corresponding to clone M0B6. 1.5% agarose gel electrophoresis. M, Gibco-BRL 100 bp ladder. Arrow indicates position of the expected (326 bp) PCR band.

UV5HL9-5B cells (14). The detection of a PCR product of the expected length for a clone with only human and hybrid templates, but not with hamster ones was considered to be a proof that the clone did originate from human chromosome 19 (Fig. 3). Of 20 sequenced clones, 10 were found to be human chromosome 19-specific, eight were of hamster origin and two were human, but belonging to some other human chromosome. Only human chromosome 19-specific clones were used for subsequent analysis.

Table 1. Summary of MAR sequences mapped on chromosome 19

<table>
<thead>
<tr>
<th>Clone characteristics</th>
<th>BROWSER attribute</th>
<th>Primer structure</th>
<th>PCR fragment Length (bp)</th>
<th>Physical map position</th>
</tr>
</thead>
<tbody>
<tr>
<td>No.</td>
<td>No.</td>
<td>Clone</td>
<td>Primer structure</td>
<td>Length (bp)</td>
</tr>
<tr>
<td>1</td>
<td>M0A3</td>
<td>PCR3298</td>
<td>ACGTTGTCACCAGATTCTCTTC</td>
<td>175</td>
</tr>
<tr>
<td>2</td>
<td>M0A7</td>
<td>PCR3299</td>
<td>ACACGTAGTCCAGATCTCTGCA</td>
<td>253</td>
</tr>
<tr>
<td>3</td>
<td>M0A9</td>
<td>PCR3300</td>
<td>GAGTCACTGACATCCAGCG</td>
<td>251</td>
</tr>
<tr>
<td>4</td>
<td>M0A11</td>
<td>PCR3301</td>
<td>ATTTCGGAGGAGGTTGAGG</td>
<td>157</td>
</tr>
<tr>
<td>5</td>
<td>M0B1</td>
<td>PCR3302</td>
<td>AAGTCAGCTCAGCTACAA</td>
<td>103</td>
</tr>
<tr>
<td>6</td>
<td>M0B2</td>
<td>PCR3303</td>
<td>GCCAGAGCGGAGGACATGAGC</td>
<td>216</td>
</tr>
<tr>
<td>7</td>
<td>M0B6</td>
<td>PCR3304</td>
<td>GATCTCCTATTGGACGAGTC</td>
<td>326</td>
</tr>
<tr>
<td>8</td>
<td>M0B7</td>
<td>PCR3305</td>
<td>AAGGGAAGAAGAGGACG</td>
<td>264</td>
</tr>
</tbody>
</table>

aBROWSER attribute, an attribute assigned to the clone in the LLNL Human Genome Center database.
bThe distance in Mb of the clone chromosome position from the telomeric terminus of the p-arm. nd, not determined

Binding of the library clones to the nuclear matrix in vitro

To check the matrix binding properties, nine chromosome 19-specific inserts (except clone M0A12, a homolog of M0A3) were labelled with $^{32}$P using PCR with the library primer (see Materials and Methods). Fragments of phage λ DNA (nucleotides 916–1077) and phage T7 DNA (nucleotides 34257–34395) were used as negative controls, whereas pUCMAR10 insert served as a positive control. The pUCMAR10 plasmid, containing 250 bp insert of 10× repeated synthetic 25 bp sequence of MAR located 3′ to IgH enhancer in the pUC19 vector, was constructed as described in (28) and provided to us by I. Shatsky, Moscow State University. The control DNAs were PCR-labelled as above, using the corresponding 20-nt primers for phage templates and 17 nt direct and reverse M13 sequencing primers for pUCMAR10.

Labelled inserts of the library clones and pUCMAR10 were mixed with equal amounts of each negative control and nuclear matrices isolated by the lithium 3,5-diiodosalicylate (LIS) extraction method (24). The mixture was incubated and separated into pellet and supernatant fractions as described (25). The DNA purified from these fractions was analysed by denaturing polyacrylamide gel electrophoresis (Fig. 4). As clearly seen in Figure 4, all the inserts of the library clones except one (M0A6; ~90%) were preferentially bound to the nuclear matrix as compared with the negative controls, although the efficiency of binding varied considerably among the clones. Similar results were obtained with matrices isolated by the high salt method (23), their capability to bind DNA being however several times lower than that for LIS extracted matrices (not shown).

To obtain the quantitative estimation of the binding efficiency of the putative MARs relative to the negative controls, the corresponding bands were cut out from the gel, counted in a scintillation counter and binding coefficients were calculated using the equation:

$$B = \frac{S_c \times P_m}{P_c \times S_m}$$
where B represents a binding coefficient; Pm, c.p.m. of the corresponding MAR band in the pellet; Pc, c.p.m. of the control band in the pellet; Sm, c.p.m. of the corresponding MAR band in the supernatant; and Sc, c.p.m. of the control band in the supernatant.

The background was subtracted from all counts. The calculated coefficients are presented in Table 2.

**Relationship between MARs and tightly bound matrix DNA**

As all the MARs were identified here by binding to the matrices in vitro, the relation between them and tightly bound to the nuclear matrix residual DNA was investigated. To achieve this, the PCR-amplified inserts of the human-specific clones together with positive and negative controls were blotted onto nylon membrane and hybridized to labelled total matrix DNA, purified from the high salt extracted nuclear matrices. The hybridization results after a mild wash (5× SSC, 65°C) are presented in Figure 5. The figure reveals that all our clones (including poorly binding to the nuclear matrix clone M0A6) have homologs among the matrix DNA fragments, but intensities of their hybridization bands are different and do not correlate with their matrix binding ability (compare corresponding bands in Figs 4 and 5). Heterogeneity of the matrix DNA probe can provide at least two reasons for differential hybridization of the clones. For instance weak hybridization bands could be either a result of hybridization with highly homologous but not abundant matrix DNA fragments, or with loosely homologous but abundant fragments. To choose between these possibilities, we washed the hybridized filter, additionally, under highly stringent conditions in 0.5× SSC at 65°C (data not shown). As the relative intensities of the bands did not change, we concluded that all hybridization signals were due to highly homologous probes of different abundance.

**Table 2. Properties of human MAR clones**

<table>
<thead>
<tr>
<th>No.</th>
<th>Clone</th>
<th>Accession no. (EMBL)</th>
<th>Insert size (bp)</th>
<th>Matrix binding coefficient (%)</th>
<th>G + C (%)</th>
<th>A + G (%)</th>
<th>A + T-rich regions (bp)a</th>
<th>Inverted repeatsb</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>M0A3</td>
<td>Z35279</td>
<td>366</td>
<td>26</td>
<td>48</td>
<td>63</td>
<td>None</td>
<td>8c</td>
</tr>
<tr>
<td>2</td>
<td>M0A6</td>
<td>Z35288</td>
<td>225</td>
<td>1.4</td>
<td>52</td>
<td>75</td>
<td>None</td>
<td>Nonec</td>
</tr>
<tr>
<td>3</td>
<td>M0A7</td>
<td>Z35291</td>
<td>302</td>
<td>12</td>
<td>40</td>
<td>41</td>
<td>None</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>M0A9</td>
<td>Z35290</td>
<td>321</td>
<td>7.3</td>
<td>59</td>
<td>63</td>
<td>None</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>M0A11</td>
<td>Z54220</td>
<td>368</td>
<td>23</td>
<td>42</td>
<td>59</td>
<td>59</td>
<td>2</td>
</tr>
<tr>
<td>6</td>
<td>M0B1</td>
<td>Z54221</td>
<td>650</td>
<td>53</td>
<td>40</td>
<td>52</td>
<td>182</td>
<td>9</td>
</tr>
<tr>
<td>7</td>
<td>M0B2</td>
<td>Z54222</td>
<td>562</td>
<td>22</td>
<td>57</td>
<td>59</td>
<td>None</td>
<td>2</td>
</tr>
<tr>
<td>8</td>
<td>M0B6</td>
<td>Z54223</td>
<td>350</td>
<td>6.5</td>
<td>49</td>
<td>61</td>
<td>None</td>
<td>6</td>
</tr>
<tr>
<td>9</td>
<td>M0B7</td>
<td>Z54224</td>
<td>334</td>
<td>20</td>
<td>40</td>
<td>63</td>
<td>127</td>
<td>3</td>
</tr>
</tbody>
</table>

aThe length of continuous A + T-rich (>75% A + T) regions longer than 20 bp.
bThe number of perfect inverted repeats (including palindromes) longer than 6 bp.
cImperfect GA repeats.
Mapping of MARs on the human chromosome 19

Finally, we mapped the characterized MAR sequences on human chromosome 19 (see Table 1). The strategy of mapping was described in detail elsewhere (14). The F and R cosmid libraries of chromosome 19, arrayed on high density filters (5,6), were constructed at Lawrence Livermore National Laboratory (USA) and provided by A.V. Carrano. They were used for hybridization with the PCR-labelled MAR probes. To verify the hybridization results the hybridization-positive cosmids were used as templates for PCR with clone-specific internal primer pairs. The detection of a product of the expected length was considered to be a proof that the sequence of the corresponding MAR is located in a given cosmid (data not shown).

MAR positive cosmids were positioned on the chromosome 19 physical map by virtue of their membership in cosmid contigs generated by restriction fingerprinting, and subsequent mapping of these contigs using fluorescence in situ hybridization (FISH). A description of the physical map is available (5). Subsequent versions of this map can be found using the following URL: http://www-bio.llnl.gov/bbrp/genome/genome.html.

DISCUSSION

The nuclear matrix attachment regions (MARs) are structurally and functionally important elements of eukaryote genome (16–20). Being polymorphic and evenly scattered over the human genome (16), these sequences can serve not only as markers for physical mapping, but also provide an opportunity for functional characterization of the human genome, such as location of sites of initiation of DNA replication and transcription. For the functional mapping chromosome specific representative libraries of sequences serving particular functions (e.g., binding to the nuclear matrix) are superior to non-specific ones (discussed in detail in ref. 14). Such a library in our case is a source of hundreds of chromosome specific MARs also suitable for comparative analysis which might help to reveal the nature of the nuclear matrix binding sites and their interaction with MARs.

Assuming that the length of chromosome 19 DNA is ~6 × 10^7 bp and the average size of chromatin loops is 60 kb (18), the total number of the chromosome DNA fragments attached to the nuclear matrix in interphase should be ~1000. The number of different sequences, potentially capable of binding the nuclear matrix, however, can be significantly larger as a large proportion of the potential sites can be detached from the matrix during the cell cycle (29). Despite this ambiguity, it seems reasonable to assume that the number of MARs in the chromosome is around several thousands. Consequently, to be a good source of MAR sequences for mapping, a MARs library should contain at least several hundred independent clones.

Basically there are two possible approaches for MAR library construction: either to clone directly the residual matrix DNA or to clone the fragments capable of binding nuclear matrices in vitro. The first one has an obvious advantage in that all clones obtained are by definition a part of the nuclear matrix. This approach was employed recently by Boulikas and Kong (30). However, 90 of 150 clones of the resulting library contained the same insert. Another serious limitation of this kind of libraries if used for mapping is that they cannot be constructed in a chromosome specific manner, and therefore imply the laborious and time-consuming chromosome assignment of each clone.

We used the second approach which permitted us to obtain both a chromosome-specific and highly representative library. Of 20 sequenced inserts 19 were independent, 10 were human chromosome 19-specific and nine of them were bound specifically to the nuclear matrix in vitro. This indicates that our arrayed library potentially contains about 150 MARs specific for human chromosome 19, which is about five times more than the total number of eukaryotic MARs characterized so far (20,31). It should be noted that this number can be readily increased by arraying additional clones. Moreover, blot-hybridization (Fig. 5) showed that all human MARs from this library had close homologs among residual matrix DNA fragments. The only obvious disadvantage of this library is that it contains a significant (~50%) proportion of clones of hamster origin due to a high content of hamster DNA in the source λ-library.

Although the number of MARs mapped on chromosome 19 is not yet sufficient to draw general conclusions, the first mapping attempt indicated that MARs were evenly distributed on both chromosome 19 arms (Table 1). These findings are in accord with generally accepted loop models of interphase chromatin or metaphase chromosome organization (17–20). All eight MARs were mapped to one or several cosmids belonging to the same contig. Seven of these contigs were assigned to specific locations on the chromosome using high resolution FISH to localize one or more selected cosmids from each contig (32). The position of the contig containing M0B7 could be determined only with an accuracy of up to a single chromosome band.

Within the limits of the accuracy, only one position on the chromosome could be assigned to each individual MAR clone. However, we cannot exclude that the same or closely related MARs can be found on other human chromosomes or within one and the same cosmid. Moreover, MARs M0A3, M0A7, M0B1, M0B6 and M0B9 were also mapped to additional cosmids, which are either not assigned at all (‘orphan’ cosmids) or assigned to contigs with yet unknown locations. Even with the limitations above we can conclude that all characterized MARs are individual and do not belong to any family of repeated sequences. This finding is in contrast to the results obtained for the hncDNA clones where more than one-third of the mapped sequences belonged to various repeated families (14).

Using the library constructed we also compared some properties of the chromosome 19 MARs (summarized in Table 2). The only common characteristic of all (except one) MARs was that they were relatively enriched in purines (or pyrimidines in complementary chains). However, it should be kept in mind that the actual matrix-binding sequences are most likely only a part of the MAR inserts. Therefore integral characteristics like A + G content can be deceptive.

To characterize the MAR elements in more detail, we analysed their sequences using the PC/Gene package. Four out of eight MARs contained the A + T-rich regions (>75% A + T), which is characteristic of one of the major classes of these sequences (16,19,20). Nevertheless, four other MARs with high affinity for the nuclear matrix did not contain any A + T-rich sequences longer than 20 bp. Similarly, the number of inverted repeats (potential hairpin or cruciform structures) did not correlate with the matrix affinity: for example, clones M0B2 and M0B7 with high affinity contained many less inverted repeats than M0B6, a clone with the lowest affinity (Table 2). On the other hand, a clone M0B1 with the highest affinity behaves like a typical A + T-rich MAR with many inverted repeats. Significant sequence differ-
ences between MARs in combination with their ability to compete with each other for binding with the matrix (23,25) make a serious challenge to researchers trying to find out the structural basis for this kind of specific interactions. The sequence variability can be caused, for example, by multiple overlapping DNA–protein interaction sites each with its own sequence requirements, which can, in addition, be degenerated. The experiments aimed at location of the DNA–protein contact sites within the MAR sequences will hopefully clarify the nature of this specificity.

ACKNOWLEDGEMENTS

The authors are grateful to V. K. Potapov, Y. B. Lebedev, S. V. Volik and Y. O. Shevchenko for participating in some experiments, to A. A. Volodin for sequencing of some MARs, to B. O. Glotov and other colleagues from Lawrence Livermore National Laboratory for indispensable help in mapping of sequences on chromosome 19. Work in Russia was supported by grant of Human Genome State Project of Russia and International Science Foundation grant #MGT000. Work at LLNL was performed under the auspices of US Department of Energy by LLNL under Contract # W-7405-ENG-48.

REFERENCES