Incorporating Mouse Genome
Mammalian Genome
Genes and Phenotypes

Sequence organization and matrix attachment regions of the human serine protease inhibitor gene cluster at 14q32.1

Stephanie J. Namciu,* Richard D. Friedman,* Mark D. Marsden, Lourdes M. Sarausad, Christine L. Jasoni, R. E. K. Fournier

Division of Basic Sciences, Fred Hutchinson Cancer Research Center, 1100 Fairview Ave. N., A2-O25, Seattle, Washington 98109-1024, USA

Received: 8 July 2003 / Accepted: 5 November 2003

Abstract

The human serine protease inhibitor (serpin) gene cluster at 14q32.1 is a useful model system for studying the regulation of gene activity and chromatin structure. We demonstrated previously that the six known serpin genes in this region were organized into two subclusters of three genes each that occupied ~370 kb of DNA. To more fully understand the genomic organization of this region, we annotated a 1-Mb sequence contig from data from the Genoscope sequencing consortium (http://www.genoscope.cns.fr/). We report that 11 different serpin genes reside within the 14q32.1 cluster, including two novel α1-antiproteinase-like gene sequences, a kallistatin-like sequence, and two recently identified serpins that had not been mapped previously to 14q32.1. The genomic regions proximal and distal to the serpin cluster contain a variety of unrelated gene sequences of diverse function. To gain insight into the chromatin organization of the region, sequences with putative nuclear matrix-binding potential were identified by using the MAR-Wiz algorithm, and these MAR-Wiz candidate sequences were tested for nuclear matrix-binding activity *in vitro.* Several differences between the MAR-Wiz predictions and the results of biochemical tests were observed. The genomic organization of the serpin gene cluster is discussed.

Serine protease inhibitors (serpins) are a group of structurally related proteins that have evolved to perform diverse physiological functions *in vivo* (for reviews see Gettins 2000; Marshall 1993; Carrell et al. 1987). These proteins are encoded by a family of genes that are distributed over several different human chromosomes (Irving et al. 2000); however, some serpin genes are organized into discrete gene clusters. For example, an ~370-kb segment of human Chr 14q32.1 contains six different serpin genes, and these genes are organized into two distinct subclusters of three genes each (Rollini and Fournier 1997a, 1997b; Billingsley et al. 1993). The proximal subcluster contains the *α1-antitrypsin (α1AT)* and *corticosteroid-binding globulin (CBG)* genes, as well as an *antitrypsin-related (ATR)* pseudogene. The distal subcluster includes the *kallistatin (KAL)*, protein *C-inhibitor (PCI)*, and *α1-antichymotrypsin (AACT)* genes. Expression of these genes is cell-specific, with different genes being expressed in different cell types (Hammond et al. l991; Rollini & Fournier, 1997a).

We previously characterized the DNA and chromatin organizations of an ~150-kb segment of 14q32.1 that contains the proximal serpin subcluster, that is, *α1AT*, *ATR*, and *CBG*, using a variety of approaches. These included restriction endonuclease mapping, DNA sequencing, promoter mutagenesis, DNase I-hypersensitive site (DHS) mapping, and analyses of matrix attachment regions (MARs) (Rollini, 1997; Rollini, 1999; Rollini, 2000; Rollini, 2000; Rollini, 1999; Rollini, 1997). These maps have been useful guides for functional analysis of the locus by targeted mutagenesis (Marsden & Fournier 2003). To gain further insight into the genomic organization of the serpin gene cluster and neighboring regions, we analyzed ~1 Mb of 14q32.1 sequence from the Genoscope sequencing consortium (http://www.genoscope.cns.fr/). This report describes our analyses of the genomic organization of the serpin

*These authors contributed equally to this work.

Correspondence to: R.E.K. Fournier; E-mail: kfournie@fhcrc.org

 DOI: 10.1007/s00335-003-2311-y • Volume 15, 162–178 (2004)

cluster and neighboring DNA segments, and it includes both *in silico* and *in vitro* analyses of MARs throughout the region.

Materials and methods

Sequence assembly. The initial assembly was made from the June 2000 (version 6.0) draft sequences of seven BACs (R-179A9, R-986E7, R-262P9, R-349L1, R-304M6, R-2068J7, and R-1089B7) from the Genoscope web site (http://www.genoscope.cns.fr/) subsequently updated through the January 2003 freeze. The average base calls with a phred value >40 (Richterich 1998) was 96.5% for the seven BACs. The BACs were oriented, aligned, and assembled into a single contig using Sequencher™ 4.1. Sequencher-generated gaps were eliminated by manual alignment. The sequence data were imported into a Gene Construction Kit™ 2.0.13 (GCK) file for gene mapping and translation of mapped exons. Position Zero on this map is defined as the transcription start site from the hepatic promoter of the *α1-antitrypsin* gene (*PI*).

Sequence motifs. The University of Wisconsin GCG version 10.0 Window program was used to determine the percentage AT content and the locations of CpG islands in 100-bp windows shifted in 50-bp steps. The results were graphed by using Microsoft® Excel 98.

Repetitive elements. Repetitive DNA elements were mapped using RepeatMasker (http://ftp.genome.washington.edu/cgi-bin/RepeatMasker). Sequence data were submitted in 100-kb segments with 10-kb overlaps to assist alignment. RepeatMasker reports were sorted into text files based on the different types of repetitive elements. A PERL script was used to generate a color-coded graphical representation of LINES, SINES, retrotransposon-like elements, and fossil DNA/MER2 transposable elements.

Matrix-attachment regions. Putative MARs were mapped by using MAR-Wiz 1.0 (http://www.futuresoft.org/MAR-Wiz/). MAR-Wiz uses six criteria—AT-richness, kinked DNA, origins of replication, TG-richness, curved DNA, and topoisomerase II recognition sites—to predict matrix association potential. We submitted unmasked sequence data in 100-kb segments with 50-kb overlaps to assist in alignment. The default settings and rules for generating matrix association potential were used. The graphical outputs of the matrix association potential were color coded and aligned to form a single graph in Adobe Photoshop™ 5.0. The color coding aided in differentiating the peaks from the two overlapping graphs. Peaks with matrix association potentials >0.6 have generally been considered MARs.

Gene identification. The unmasked sequence data were submitted to the *ab initio* gene identification programs GenScan (http://genome.dkfz-heidelberg.de/cgi-bin/GENSCAN/genscan.cgi) and FGENES 1.5 (http://searchlauncher.bcm.tmc.edu/gene-finder/gfb.html) with the standard settings for human sequences. We obtained text and graphical outputs indicating the positions and extents of candidate genes, their orientations, putative intron-exon structures, promoters, polyadenylation sites, and predicted peptide sequences. The positions and orientations of candidate genes were plotted in the GCK sequence file.

BLAST searches. The predicted peptide sequences from the GenScan and FGENES analyses were submitted to the NCBI BLAST sequence comparison algorithms (http://www.ncbi.nlm.nih.gov/BLAST/) for comparison with the non-redundant protein sequence database (BLASTP) or with the DNA sequence database translated in all six reading frames (TBLASTN), either at the NIH website or by using the tentative human consensus (THC) database at TIGR (http://www.tigr.org/tdb/hgi/index.html). An advantage of the TIGR database is that overlapping ESTs have been assembled into single THC files, which facilitates non-redundant identifications. We also performed BLAST searches of the EST database at the NIH. Scores for BLASTP and TBLASTN analyses of >400 with E-probabilities of <e–100 for either GenScan or FGENES were initially considered sufficient for candidate gene assignment. The scores for the analysis from the TIGR and NCBI EST databases were generally set at >1000 with E-values <e–100. The scores and probabilities for the best hits were tabulated and analyzed. As we progressed, we noted that candidate genes encoding short polypeptides or those with regions of low complexity (repeating units of one to three amino acids) tended be assigned lower scores. These low-scoring gene candidates were reexamined by BLASTN analysis, even if their scores were below the conservative criteria we set initially, in order to determine whether the putative gene identification was plausible. BLAST2 alignments of the contig sequences to the GenBank sequences for the best hits

were made as an additional confirmation of the identification. We also used BLAST2 to compare the protein sequence of various serpins with the another.

The GenBank records were examined to determine whether additional information was available indicating the chromosomal location of each candidate gene. We submitted the DNA sequence data for each candidate gene and the GenBank genomic DNA or cDNA sequence records to BLAST2 for alignment. Those alignments that generated high scores provided an additional means for confirming the gene identification. The promoters, coding sequences, and poly-adenylation sites were mapped and translated in the GCK sequence file.

Construction of test plasmids for matrix-binding assays. Thirty-nine different test DNA fragments from six MAR-Wiz candidate sequences were amplified by PCR (5 cycles at 94°C for 20 s, 55°C for 30 s, and 72°C for 90 s, followed by 30 cycles at 94°C for 20 s, 68°C for 30 s and 72°C for 110 s) by using sequence specific primers with *Nco*I, *Sal*I, or *Sac*II linkers at their 5′ ends. The sequences of the primer pairs used for the PCR amplifications can be found at http://www.fhcrc.org/labs/fournier/mammalian-genome-supplemental-data/Primers.html.

Probe #10/11 from the +228–247 kb region was generated by digesting BAC R-2068J7 with *Eco*RI and *Xho*I, which produced an ~6.3-kb fragment containing the #10/11 sequence. The ~6.3-kb DNA fragment was purified, digested with *Nco*I to generate an ~4.1-kb fragment, which was subcloned as described below.

The various probe fragments were cut with *Nco*I, *Sal*I, or *Sac*II and ligated into the *Nco*I, *Sal*I, or *Sac*II sites of one of two pGEM 5fZ(+) plasmids that contained either the 850-bp *apoB* 3′ MAR or the 4.1-kb *ATR* MAR. Isolated test plasmids were digested with either *Sal*I/*Not*I (*ATR*), *Nco*I/*Not*I or *Sac*II/*Not*I (*apoB* 3′) to generate three DNA fragments: the plasmid backbone (negative control), the *apoB* 3′ MAR or *ATR* MAR (positive control), and the specific test fragment from the serpin locus (experimental). These mixtures were end-labeled with α-[^{32}P]-dCTP and α-[^{32}P]-dATP by using Klenow polymerase and were used as probes in *in vitro* MAR assays.

In vitro MAR assay. Nuclei were isolated from human HepG2 (hepatocellular carcinoma) cells that had been grown in 1:1 Ham's F12/Dulbecco's modified Eagle's medium (F/DV) supplemented with 10% fetal bovine serum (Gibco). 10^7 cells were washed twice with a hypotonic Wash buffer [10 mM Tris-HCl pH 7.4, 40 mM KCl, 0.25 mM spermidine, 0.1 mM spermine, 1 mM EDTA/KOH pH 7.4, 20 mg/mL Aprotinin (Sigma), 1% thiodiglycol], resuspended in homogenization buffer (Wash buffer plus 0.5% Digitonin), and homogenized with approximately 20 strokes in a 15-mL Wheaton Dounce homogenizer and a tight pestle to liberate the nuclei. The nuclei were washed twice in homogenization buffer, resuspended at 20 $OD_{260nm/ml}$ in freezing buffer (homogenization buffer containing 50% glycerol), and stored at –20°C

The *in vitro* MAR assay utilized histone-depleted nuclear matrices as first reported by Mirkovitch et al. [Mirkovitch et al. 1984] and described in detail in Rollini et al. (1999). Briefly, nuclei were washed twice in Wash buffer, resuspended in 50 μL of Wash buffer, and stabilized by incubation at 42°C for 20 min. Chromatin proteins were extracted by incubating the nuclei in high salt buffer (2 M NaCl, 5 mM HEPES pH 7.4, 0.25 mM spermidine, 2 mM EDTA/KOH pH 7.4, 0.2 mM KC1, and 0.1% digitonin) for 10 min at room temperature. The DNA ''halos'' were pelleted by centrifugation in a microcentrifuge at 10K rpm for 15 min. The pellet was washed five times in digestion buffer (20 mM Tris-HCl pH 7.4, 20 mM KC1, 70 mM NaCl, 10 mM $MgCl_2$, 0.125 mM spermidine, 0.05 mM spermine, 20 mM aprotinin, 0.1% digitonin), and incubated with *Nco*I, *Spe*I, *Eco*RI, *Xho*I, and *Xba*I for 5–6 h at 37°C. Final concentrations were adjusted to 20 mM EDTA; 7.5 ng of α-[^{32}P] end-labeled probe was added to the digested halos; and the hybridization reaction was incubated at 37°C for 14–16 h. The supernatant containing digested DNA and the pellet containing nuclear matrix proteins and associated DNA were separated by centrifugation, and proteins were removed by a 3-h incubation at 55°C with proteinase K, followed by phenol/chloroform extraction and ethanol precipitation. Six μg of supernatant and pellet DNA were separated by gel electrophoresis, and audoradiography was performed. Each experiment was repeated two to four times.

Results

Assembly and analysis of the contig. Sequence data for seven BACs from the HGP contributor, Genoscope, were assembled into a single contig by using DNA Sequencher 4.1™. Computer-generated misalignments were eliminated by manual alignment. Of the assembled 1,009,375 bp, 206 bp were unidentified bases submitted as placeholders in the BACs,

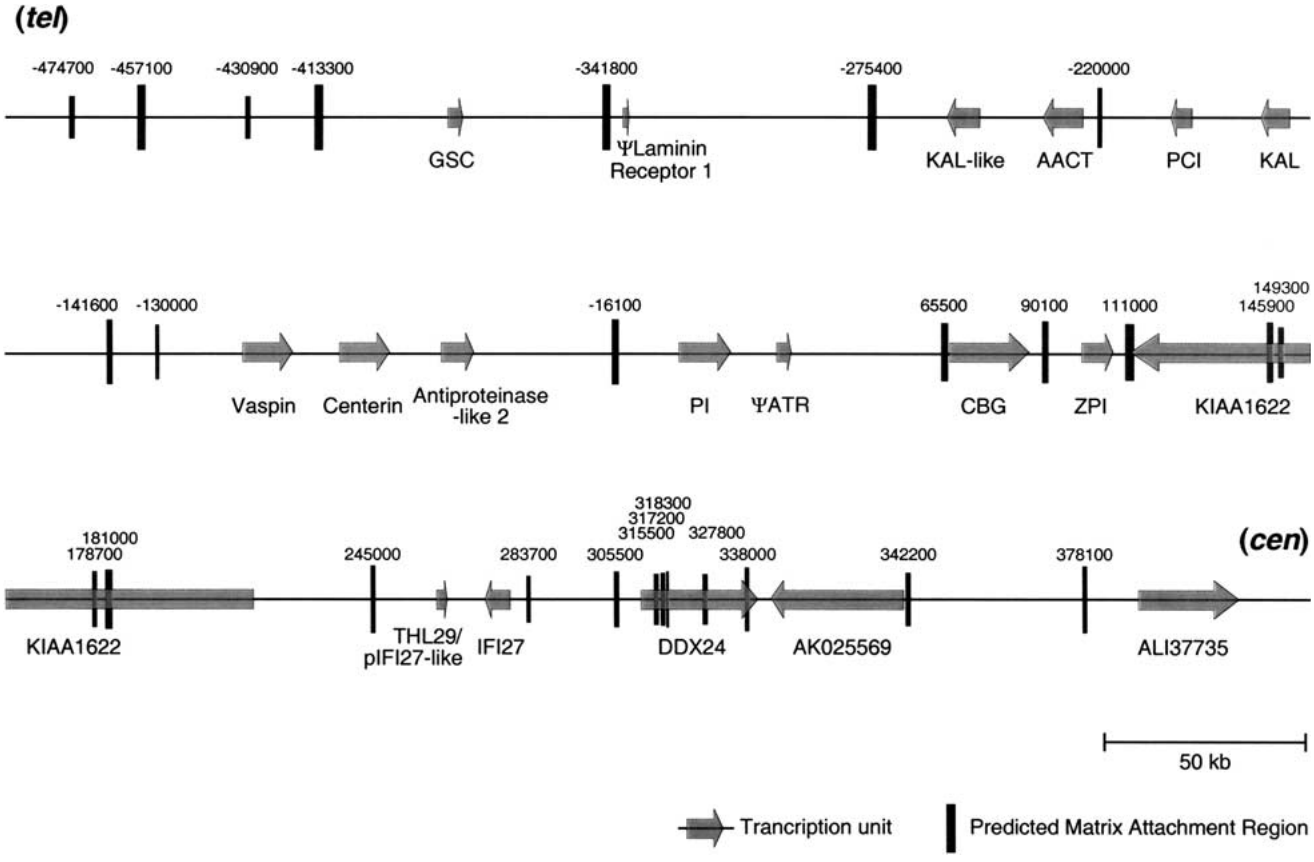


Fig. 1. Sequence annotation of the 14q32.1 serpin gene cluster and surrounding regions. The illustration displays a l-Mb sequence contig from distal (left) to proximal. Position zero on the map is defined as the transcription start site from the hepatic promoter of the α1-antitrypsin gene. Genes are denoted by light grey arrows, which indicate their transcriptional orientations. MAR-wiz-predicted matrix attachment regions are indicated by dark grey bars; the heights and widths of the bars indicate the MAR-wiz potential score and size of the putative MARs. A more detailed version of this map can be found at http://www.fhcrc.org/labs/fournier/mammalian-genome-supplemental-data/Figure_l.html.

and 77 bp were uncalled bases owing to sequencing ambiguities or alignments that could not be corrected manually.

Gene content. The assembled sequence was analyzed *in silico* to map simple sequence motifs, repetitive DNA elements, and matrix-attachment regions (MARs), and the *ab initio* gene prediction programs GenScan and FGENES were used to determine the locations and coding potentials of candidate genes (Burge and Karlin 1997; Murakami and Takagi 1998; Salamov & Solovyev 2000). A detailed map of the region can be found at http://www.fhcrc.org/labs/fournier/mammalian-genome-supplemental-data/Figure_1.html; a simplified version of this map showing coding sequences and putative matrix-attachment regions is presented in Fig. 1. Initially, FGENES identified 76 potential genes, and GenScan identified 33 potential genes in the 1-Mb interval (http://www.fhcrc.org/labs/fournier/mammalian-genome-supplemental-data/Table_1.html) We compared these candidate sequences with known genes in the public databases by using BLASTP and TBLASTN, and with those in the TIGR human EST database. This analysis resulted in the identification of 13 genes from the GenScan predictions and 14 genes from the FGENES predictions that were highly related to known gene sequences. All of these genes were identified by both gene prediction programs, but FGENES correctly scored the *α1-antitrypsin* (*PI*) gene and the linked *ATR* pseudogene as separate hits, while GenScan fused the two sequences into a single transcription unit. All six known serpin genes in the region were identified by both GenScan and FGENES. Thirteen of the 14 candidate genes had been assigned previously to human Chr 14, including the recently identified *centerin and protein Z-dependent protease inhibitor* genes [Rasmussen et al. 1993; Zhao et al. 2000; Nagase et al. 2000; Rollini & Fournier 1997a; Blum et al. 1994; Billingsley et al. 1993; Han et al. 1999, 2000; Frazer et al. 2000]. Candidate FGENES16/GenScanI was highly homologous to the *laminin*

Fig. 2. Matrix-binding activities of DNA fragments in and around a putative MAR at +88 kb. **Panel A.** Matrix-binding potential for the +88 kb region as calculated by MAR-Wiz; a peak with a matrix-binding potential of ~0.75 is located at +88 kb. The six numbered boxes below the graph represent DNA fragments that were isolated and tested for matrix-binding activity *in vitro*; + indicates matrix association and – indicates no matrix association as assessed by the nuclear matrix-binding assays shown in Panel C. The boxes below the map represent repetitive DNA elements: LTR elements are shown in light grey, LINEs and SINES are shown in dark grey. A color version of this figure can be found at http://www.fhcrc.org/labs/fournier/mammalian-genome-supplemental-data/Figure_2.html. **Panel B.** The six sequence motif rules used by MAR-WIZ to calculate matrix binding potentials, and the locations of specific motifs within the +88 kb region: Ori = origin of replication motif, TG = TG-rich region, Kinked = kinked DNA motif, Curved = curved DNA motif, AT = AT-rich region, and Topo II = topoisomerase II consensus site. Each hatch-mark represents one rule match. **Panel C.** *In vitro* MAR assay. Genomic DNA fragments from the +88 kb region were amplified by PCR and subcloned into a plasmid vector that contained either the *apoB* 3′ MAR (fragments 3 and 5) or the ATR MAR (fragments 1, 2, 4, and 6) as described in Materials and methods. Isolated test plasmids were digested with restriction enzymes that liberated the plasmid backbone (negative control), the *apoB* 3′ MAR or ATR MAR (positive control), and the genomic test fragment (experimental). These mixtures were end-labeled with α-[^{32}P]-dCTP and α-[^{32}P]-dATP and incubated with histone-depleted nuclear matrices as described (Rollini et al. 1999). The partitioning of specific DNA fragments into matrix-associated (P, pellet) and non-associated (S, supernatant) fractions was assessed after centrifugation, and specific DNA fragments in each fraction were identified by electrophoresis and autoradiography. The autoradiograph shows the partitioning of DNA fragments 1 through 6 in total DNA (T), supernatant (S), and pellet (P) fractions. The positive controls for matrix-association are either the 4.1-kb ATR MAR. indicated by an arrow on the left side of the figure, or the 0.85-kb apo B 3′ MAR, labeled with a white cross on the figure itself. The negative control is the plasmid DNA vector (pDK101). Specific enrichment of fragment 4, but not other test fragments, in the pellet fraction was observed.

receptor-1 gene, which is located on human Chr 3. However, the 14q32.1 genomic sequence comprised a single exon with a poly(A)-tract at its 3′ end; thus, it appears to be a member of the previously described family of processed laminin receptor pseudogenes (Jackers et al. 1996).

To determine whether any other candidate sequences might represent *bona fide* human genes, we searched for additional candidates that had been identified by both gene identification programs. Only five genes among the remaining candidates fulfilled this criterion. Three of these sequences exhibited homology to known serpin genes. FGENES24/GenScanM was homologous to the human *kallistatin* (*KAL*) gene, which is located ~70 kb more proximal on 14q32.1. This *KAL-like* sequence was ~16 kb downstream of *AACT*, and it was in the same transcriptional orientation. Transcripts from the KAL-like sequence have been detected in human liver RNA (data not shown). Two other putative serpin genes were located in the 170-kb interval between the distal and proximal serpin subclusters. These candidate genes, FGENES37/GenScanR and FGENES41/GenScanT, were homologous to the *α1-antiproteinase* genes of several vertebrate species. Recently, a cDNA clone of FGENES37/GenScanR was isolated and designated visceral adipose-specific serpin (vaspin) (Strausberg et al. 2002). A cDNA clone for FGENES41/GenScanT has been designated BX248259 (http://www.genoscope.cns.fr/). The remaining two candidate genes (FGENES62/GenScanBB and FGENES64/GenScanCC) encode human *THL29* and *pIFI27*, respectively.

Thus, the human serpin gene cluster at 14q32.1 contains 11 different serpin genes in a genomic interval that extends from map positions –259 kb to +106 kb, and there are no non-serpin genes in this interval (Fig. 1). The ~270-kb region distal to the serpin gene cluster contains only two genes, *Goosecoid* (*GSC*) and a *laminin receptor-1* pseudogene. In contrast, the ~370-kb region proximal to the serpin cluster is relatively gene rich: it contains six different human genes, including the ~110-kb *KIAA1622* transcription unit, THL29, pIFI27, DDX24 (a DEAD/H box-containing protein), AK025569, and ASB2. Our observations coincide with the HGP and Genoscope databases for this region.

Simple sequence motifs. The average AT content of the region was 54.7%, although there were variations over the length of the contig. Most notably, the region around *GSC* (–385 to –375 kb) was relatively AT-poor (39.9% AT), whereas the entire *KIAA1622* transcription unit (+110 to +210 kb) was relatively AT-rich (64.3% AT). CpG islands were located upstream of *GSC* (~–380 kb) and *KIAA1622* (~+215 kb) and between the *interferon* α-inducible *p27* gene and *DDX24* (~+310 kb). There were also small CpG-enriched regions within AK025569 (~+362 kb) and AL137735 (~+453 kb), as shown in detail at http://www.fhcrc.org/labs/fournier/mammalian-genome-supplemental-data/Figure_1.html.

Repetitive elements. Interspersed repetitive DNA elements make up ~40% of the human genome [Heilig et al. 2003]. We mapped LINES, SINES,

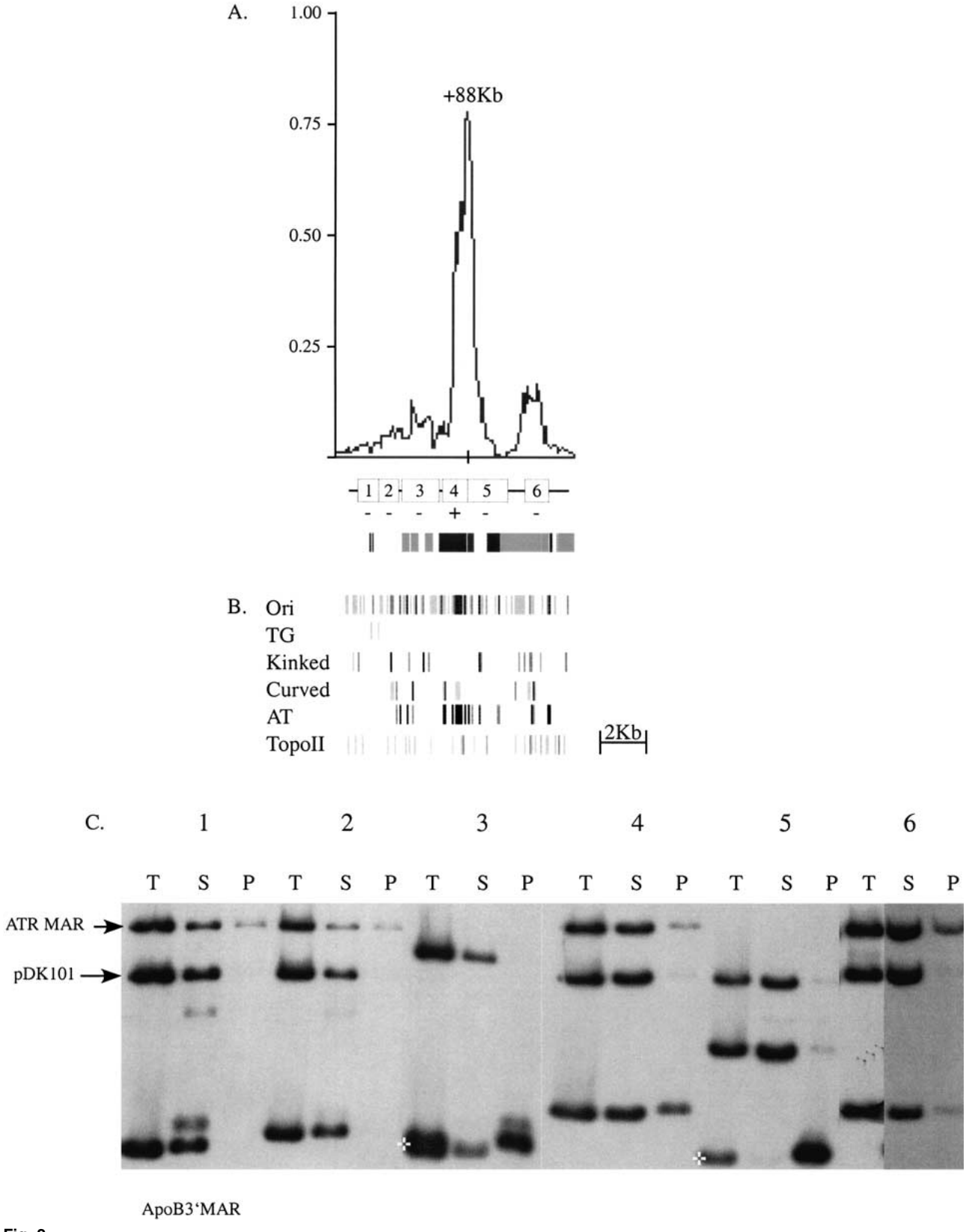

A.
1.00
0.75
0.50
0.25
+88Kb
1 2 3 4 5 6
- - - + - -
B.
Ori
TG
Kinked
Curved
AT
TopoII
2Kb
C.
1 2 3 4 5 6
T S P T S P T S P T S P T S P T S P
ATR MAR
pDK101
ApoB3'MAR

Fig. 2.

Fig. 3. Matrix-binding activities of DNA fragments in and around a putative MAR at +109 kb. **Panel A.** Matrix-binding potential for the +109-kb region as calculated by MAR-Wiz; a peak with a matrix-binding potential of ~0.75 is located at +109 kb. The seven numbered boxes below the graph represent DNA fragments that were isolated and tested for matrix-binding activity. The remainder of the Panel A is organized as described in the legend to Fig. 2. Repetitive elements within fragments 1–7 were either DNA-type repeats (medium grey) or SINES (dark grey). A color version of this figure can be found at http://www.fhcrc.org/labs/fournier/mammalian-genome-supplemental-data/Figure_3.html. **Panel B.** Sequence motifs used by MAR-WIZ to calculate matrix-binding potentials, and the locations of specific motifs within the +109-kb region, as described in the legend to Fig. 2. **Panel C.** *In vitro* MAR assay. Matrix-binding activity of DNA fragments 1–7 was assessed as described in the legend to Fig. 2. Fragments 4, 5, and 6 displayed nuclear matrix-binding activity.

retrotransposon-related elements, and DNA-type repetitive elements using RepeatMasker (http://ftp.genome.washington.edu/cgi-bin/RepeatMasker). Repetitive DNA elements were distributed throughout the 1-Mb region, but the coding regions of many genes in the interval were largely devoid of repeats. Interestingly, the distal half of the region was relatively rich in SINEs, but poor in LINEs, and the LINEs that did occur in this genomic segment were generally only ~1-2 kb. Conversely, the proximal half of the contig contained many LINEs of 5–10 kb (http://www.fhcrc.org/labs/fournier/mammalian-genome-supplemental-data/Figure_1.html).

Matrix-attachment regions. Matrix-attachment regions (MARs) are DNA sequences that are defined by their abilities to bind to nuclear matrix preparations *in vitro* (Mirkovitch et al. 1984; Cockerill and Garrard 1986; Izaurralde et al. 1988) MARs are generally found in intergenic regions, suggesting that they may function to organize eukaryotic chromatin into individual loops or domains (Gasser and Laemmli 1986; Phi-Van and Stratling 1996; Kellum and Elgin 1998). As such, MARs may provide insight into the chromatin domain organization of large genomic regions.

We used MAR-Wiz (http://www.futuresoft.org/MAR-Wiz/) to identify elements with putative matrix-binding potential in the 1-Mb region. MAR-Wiz identified 27 different elements with matrix-binding potentials >0.6 (Fig. 1). Sixteen of these putative MARs were identified in both overlays, but 11 scored >0.6 in only one trace. This indicates that the matrix-binding potentials calculated by MAR-Wiz are sequence context-dependent; that is, the values are calculated relative to other sequences in the 100-kb window.

The pattern of matrix-binding potential across the 1-Mb region varied considerably (Fig. 1 and http://www.fhcrc.org/labs/fournier/mammalian-genome-supplemental-data/Figure_1.html). The distal half of the contig, which contained few genes, displayed discrete peaks of high matrix-binding potential separated by long stretches of DNA with little or no predicted matrix-binding activity. These inter-MAR segments varied in length from ~20 kb to ~100 kb. Some of the inter-MAR segments were apparently gene free; for example, the three small segments (~18, 27, and 18 kb, respectively) at the distal end of the contig did not contain any genes (Fig. 1). Other inter-MAR segments contained only a single gene, for example, the −413 to −342 kb region, which contained *GSC*. Other segments contained multiple human genes.

The proximal part of the 1-Mb segment contained an ~160-kb region just downstream of *ATR* that was rich in matrix-binding potential. This was by far the highest concentration of potential matrix-binding elements in the 1-Mb region, with peaks of matrix-binding activity occurring both within and between genes (Fig. 1). This region was relatively AT-rich. Further proximal on the chromosome (+280 to +340 kb), another ~60-kb gene-rich region with high matrix-binding potential was encountered. In both of these regions an increase in matrix-binding potential was correlated with an increase in gene density.

Two different kinds of tests were performed to determine whether the DNA elements identified by MAR-Wiz corresponded to *bona fide* matrix-attachment regions. First, we analyzed MAR-Wiz predictions within the proximal serpin subcluster, which has been mapped for nuclear matrix-binding activity in detail (Rollini et al. 1999). Second, we tested MAR-Wiz candidate sequences from throughout the interval for matrix-binding activity *in vitro*.

Matrix-attachment regions in the proximal serpin subcluster. DNA sequences in the proximal serpin subcluster between -20 and +83 kb have been assayed for nuclear matrix-binding activity *in vitro* using three different assays (Rollini et al. 1999). This region contains five MARs that are matrix-associated in both hepatic and non-hepatic cells (Rollini et al. 1999). Therefore, we examined MAR-Wiz predictions for this region.

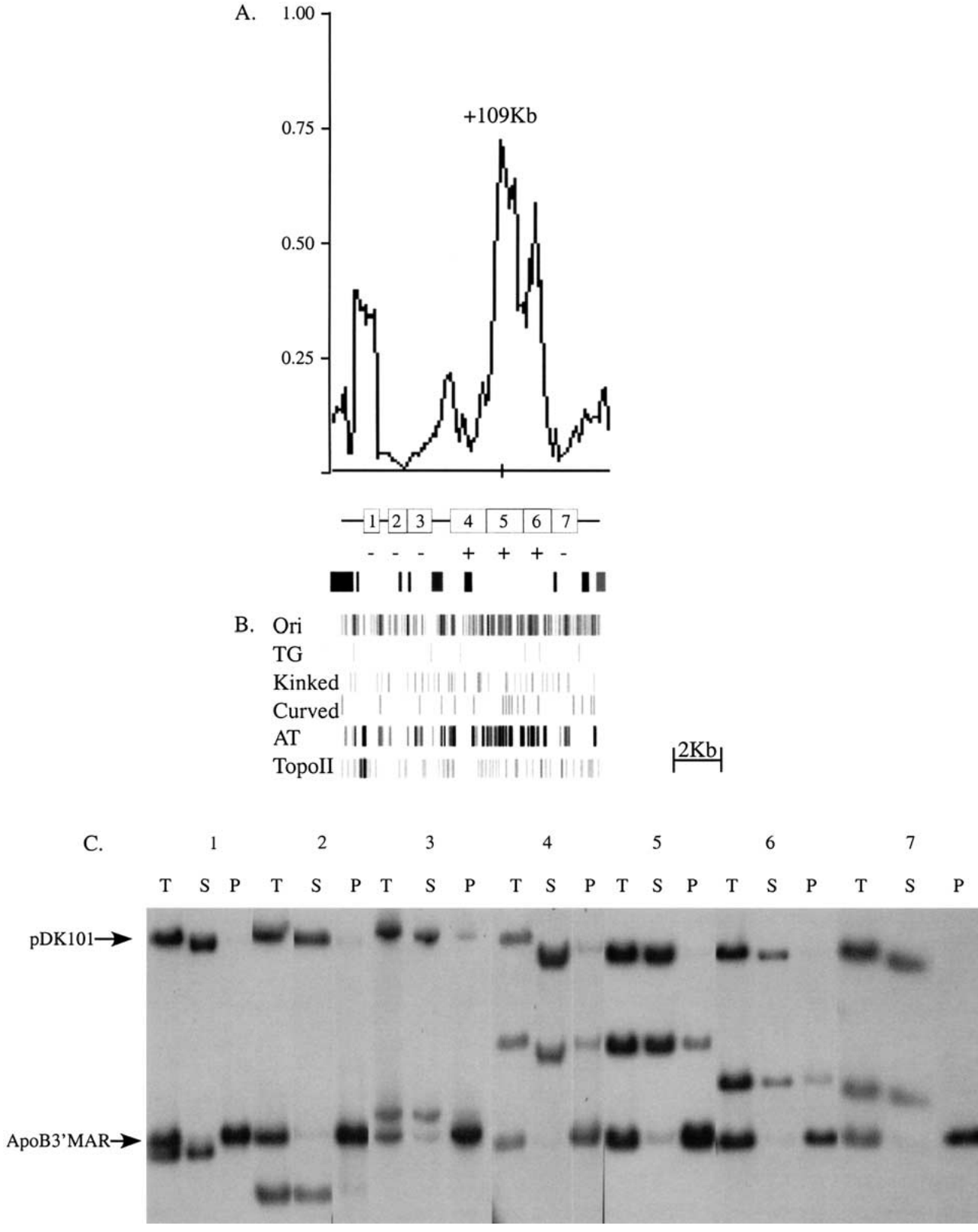

A.
1.00
0.75
0.50
0.25
+109Kb
1 2 3 4 5 6 7
- - - + + + -
B.
Ori
TG
Kinked
Curved
AT
TopoII
2Kb
C.
1 2 3 4 5 6 7
T S P T S P T S P T S P T S P T S P T S P
pDK101
ApoB3'MAR

Fig. 3.

Fig. 4. Matrix-binding activities of DNA fragments in and around a putative MAR at –143 kb. **Panel A.** Matrix-binding potential for the –143 kb region as calculated by MAR-Wiz; a peak with a matrix-binding potential of ~75 is located at –143 kb. The four numbered boxes below the graph represent DNA fragments that were isolated and tested for matrix-binding activity. The remainder of the Panel A is organized as described in the legend to Fig. 2. Repetitive elements in the region included LTR elements (light grey), DNA-type repeats (medium grey), and LINES and SINES (dark grey). A color version of this figure can be found at http://www.fhcrc.org/labs/fournier/mammalian-genome-supplemental-data/Figure_4.html. **Panel B.** Sequence motifs used by MAR-WIZ to calculate matrix-binding potentials, and the locations of specific motifs within the –143-kb region, as described in the legend to Fig. 2. **Panel C.** *In vitro* MAR assay. Matrix-binding activity of DNA fragments 1–4 was assessed as described in the legend to Fig. 2. Fragment 2 displayed nuclear matrix-binding activity.

The α1AT MAR is an ~2-kb element located ~16 kb upstream of the α1AT gene. This MAR is composed of ~85% repetitive DNA of the LTR retrotransposon type, and it is highly AT-rich (~75% AT) (Rollini 1999). MAR-Wiz correctly identified the α1AT MAR as an element with high (~0.9) matrix-binding potential (Fig. 1).

The *CBG* promoter MAR (CBGp MAR) is an ~1-kb element just upstream of the *CBG* gene that is composed of ~90% repetitive DNA. About one-third of this element is composed of AT-poor SINE sequences (~45% AT), and the remainder of the element consists of AT-rich LINE sequences (~75% AT). This element was recognized by MAR-Wiz as an element with significant matrix-binding potential (~0.8) in one of the two traces.

The other three MARs in the proximal serpin subcluster are relatively large elements with strong matrix-binding activity *in vitro* (Rollini et al. 1999). The *ATR* MAR is an ~4-kb element downstream of *ATR*; the *CBG* 5′ MAR is a large (~8-kb) matrix-binding sequence between *ATR* and *CBG*, and the *CBG* intron MAR (*CBGi* MAR) is an ~5-kb MAR in *CBG* intron I (Rollini et al., 1999 and Fig. 1). All of these elements are composed of AT-rich repetitive DNA. Although MAR-Wiz indicated that each of these MARs possessed some matrix-binding potential, none of them scored >0.6, the value generally used to identify matrix-binding elements. As such, these three elements were not scored as MARs by the MAR-Wiz algorithm. Therefore, analysis of five previously identified MARs in the serpin locus indicates that MAR-Wiz generally underestimates the matrix-binding potential of human DNA sequences.

Matrix-binding potentials of MAR-Wiz candidate sequences. As a further test of the algorithm, we tested MAR-Wiz candidate sequences for matrix-binding activity *in vitro*, using a well-established biochemical assay (Mirkovitch et al. 1984; Rollini et al. 1999). To do this, DNA fragments (0.5–2.0 kb) corresponding to sequences in and around various peaks of matrix-binding potential, as assessed by MAR-Wiz, were generated by PCR amplification. The test fragments were subcloned into plasmid vectors containing a *bona fide* MAR—either the ~850-bp *apoB 3′* MAR (Levy-Wilson and Fortier 1989) or the ~4-kb *ATR* MAR (Rollini et al. 1999). Isolated plasmid subclones were digested with restriction enzymes that liberated the plasmid backbone (negative control), the *apoB* or *ATR* MAR (positive control), and the test sequence. This mixture of DNA fragments was labeled and incubated with restriction-digested, histone-depleted nuclear matrices. The matrix-association of the various DNA fragments was assessed after centrifugation of the mixtures into supernatant (non-associated) and pellet (matrix-associated) fractions.

Figure 2 shows the results of such an analysis for DNA sequences from a putative matrix-binding element in the proximal serpin sub-cluster at +88 kb, between *CBG* and *ZPI* (Fig. 1). Six different DNA fragments, representing ~8 kb of sequence between +84 and +92 kb, were isolated and tested (Fig. 2A). Only one fragment, fragment 4, displayed matrix-binding activity (Fig. 2C). This repetitive element-rich DNA fragment (~50% repetitive DN) corresponded to a region that MAR-Wiz predicted to have a matrix-binding potential of 0.75 (Fig. 2A). The six rules that MAR-Wiz uses to calculate MAR-binding potentials of test DNA sequences are shown in Fig. 2B. Each hatch mark represents one match with the rule set. Fragment 4 had significantly more matches for the replication origin and AT-richness rules than adjacent DNA fragments. Thus, the MAR-Wiz prediction that the sequence located at +88 kb has matrix-binding activity, but surrounding DNA fragments do not, was verified by *in vitro* assays.

Another putative MAR in the proximal serpin subcluster was located at ~+109 kb, just downstream of *ZPI* (Fig. 1). Seven different DNA fragments from this region, representing ~8 kb of sequence between +104 and +112 kb, were generated and tested for matrix-binding activity (Fig. 3A). These included major and minor peaks with matrix-

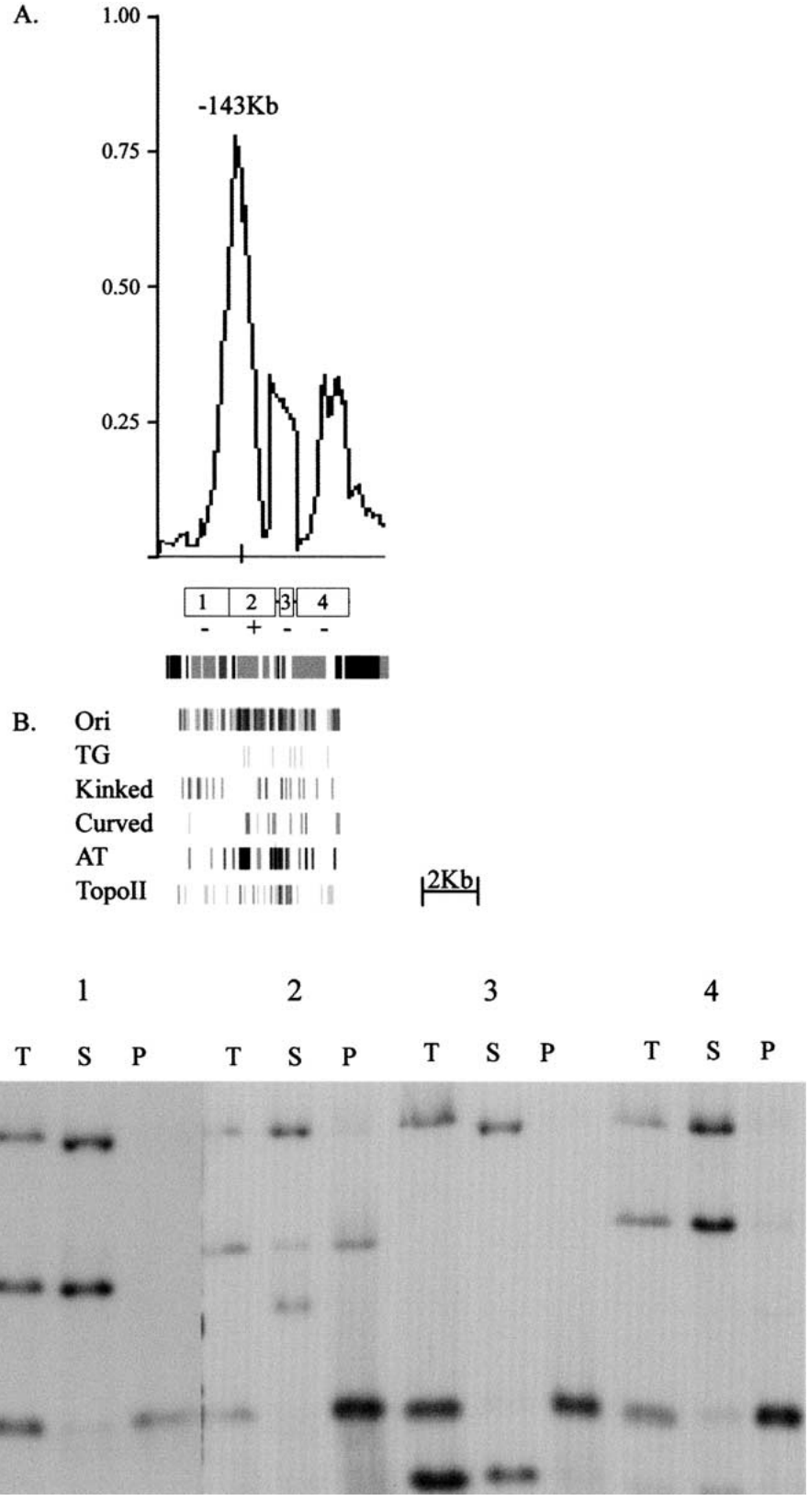

A.
1.00
0.75
0.50
0.25
-143Kb
1
2
3
4
- + - -
B.
Ori
TG
Kinked
Curved
AT
TopoII
2Kb
C.
1
2
3
4
T S P T S P T S P T S P
pDK101
ApoB3'MAR

Fig. 4.

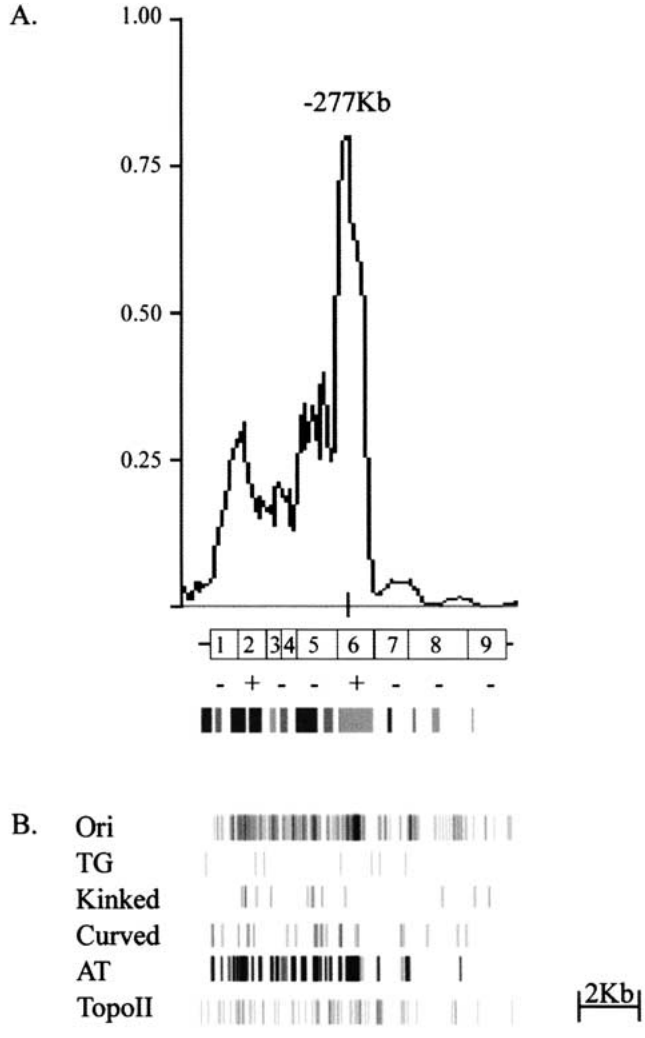
Fig. 5. Martix-binding activities of DNA fragments in and around a putative MAR at –277 kb. **Panel A.** Matrix-binding potential for the –277-kb region as calculated by MAR-Wiz; a peak with a matrix binding potential of ~0.75 is located at –277 kb. The nine numbered boxes below the graph represent DNA fragments that were isolated and tested for matrix-binding activity. The remainder of the Panel A is organized as described in the legend to Fig. 2. Repetitive elements in the region included an LTR element in fragment 8 (light grey), DNA-type repeats in fragments 3 and 6 (medium grey), and LINES and SINES in other fragments (dark grey). A color version of this figure can be found at http://www.fhcrc.org/labs/fournier/mammalisn-genome-supplemental-data/Figure_5.html. **Panel B.** Sequence motifs used by MAR-WIZ to calculate matrix-binding potentials, and the locations of specific motifs within the –277 kb region, as described in the legend to Fig. 2. **Panel C.** *In vitro* MAR assay. Matrix-binding activity of DNA fragments 1–9 was assessed as described in the legend to Fig. 2. Fragments 2 and 6 displayed nuclear matrix-binding activity, but other DNA fragments in the region did not. Note that matrix-binding fragment 2 was assigned a matrix-binding potential of only ~0.3 by MAR-Wiz, well below the threshold for putative MARs as assessed by the algorithm.

binding potentials of 0.75 and 0.40, respectively. Three of the seven DNA fragments had significant matrix-binding activities *in vitro* (Fig. 3C); these contiguous DNA fragments coincided with the major peak of matrix-binding potential predicted by MAR-Wiz (Fig. 3A). Fragment 1, which represented sequences from the minor peak of matrix-binding potential, did not display matrix-binding activity *in vitro*. Thus, these data define an ~4-kb MAR between the most proximal serpin gene in the cluster, *ZPI*, and its immediately adjacent non-serpin neighbor, *KIAA1622* (Fig. 1). This AT-rich MAR (66%) has a high density of sequences related to replication origins (Fig. 3B). However, this DNA segment is largely devoid of mid-repetitive DNA (Fig. 3A), a property that distinguishes the +109-kb MAR from all other matrix-binding elements in the region. Another unusual feature of this MAR is that it contains the terminal exon and 3′UTR of *KIAA1622.* The presence of transcribed sequences within a matrix-attachment region has been described previously (Broeker et al. 1996).

MAR-Wiz predicted the existence of two MARs between the central and distal serpin subclusters (Fig. 1). One of these elements, at −143 kb, had a large peak with significant (~0.7) matrix-binding potential, plus two smaller peaks with matrix-binding potentials of ~0.3 (Fig. 4A). Four DNA fragments from this region were tested for matrix-binding activity, and only one fragment, fragment 2, displayed matrix-binding activity *in vitro* (Fig. 4C). This DNA fragment corresponded to the large peak of matrix-binding potential identified by MAR-Wiz (Fig. 4A). The smaller peaks of matrix-binding potential did not bind to nuclear matrices *in vitro.* This MAR was composed primarily of MER1 DNA-type repetitive sequences, and it had a significantly greater number of replication origin consensus sequences and AT-richness rule matches than adjacent DNA sequences (Fig. 4B). Thus, MAR-Wiz accurately predicted the matrix-binding potentials of DNA fragments in the −143 kb region.

That was not the case in the region around −277 kb, another region that consisted of a large peak with high (~0.8) matrix-binding potential and a smaller peak with lower (~0.3) matrix-binding potential (Fig. 5A). Nine different DNA fragments from this ~9-kb region were isolated and tested for matrix-binding activity *in vitro.* Two of these fragments, fragments 2 and 6, displayed strong matrix-binding activity *in vitro* (Fig. 5C). Fragment 6 corresponded to the large peak of matrix-binding potential identified by MAR Wiz (Fig. 5A). Fragment 2 corresponded to the shoulder peak, a DNA segment ~2 kb away from fragment 6, that had a matrix-binding potential of only ~0.3. However, fragment 2 displayed strong matrix-binding activity *in vitro* (Fig. 5C). Thus, MAR-Wiz failed to identify this *bona fide* matrix-associated region. None of the DNA fragments between fragments 2 and 6 displayed matrix-binding activity *in vitro.* The two matrix-binding elements in this region had high densities of AT-richness rule matches, but fragment 6 had a higher number of replication origins (Fig. 5B). Both of these matrix-binding regions were rich in repetitive element DNA: ~50% of fragment 2 consisted of LINE 1 and MER1 DNA, whereas fragment 6 was composed exclusively of repetitive DNA—~92% DNA-type elements (AcHobo) and ~7% *Alu* DNA (Fig. 5A).

Most of the MARs in the serpin cluster identified in this and a previous (Rollini et al., 1999) report consisted largely of mid-repetitive DNA. The seven MARs in the proximal serpin cluster are particularly enriched for LINE and SINE sequences (Fig. 1). To determine whether a large, uninterrupted block of LINE sequence was sufficient for matrix-binding activity, we tested ~7 kb of sequence at −69 kb for matrix-binding activity (Fig. 6A). This region included an ~4-kb segment of uninterrupted LINE sequence with a peak of matrix-binding potential of 0.4–0.5, values like those of the *ATR, CBG* 5′, and *CBG*i MARs, which were MAR-Wiz false negatives. However, none of the six DNA fragments from this region displayed matrix-binding activity *in vitro*


A.
1.00
0.75
0.50
0.25
-277Kb
1 2 3 4 5 6 7 8 9
- + - - + - - -
B.
Ori
TG
Kinked
Curved
AT
TopoII
2Kb

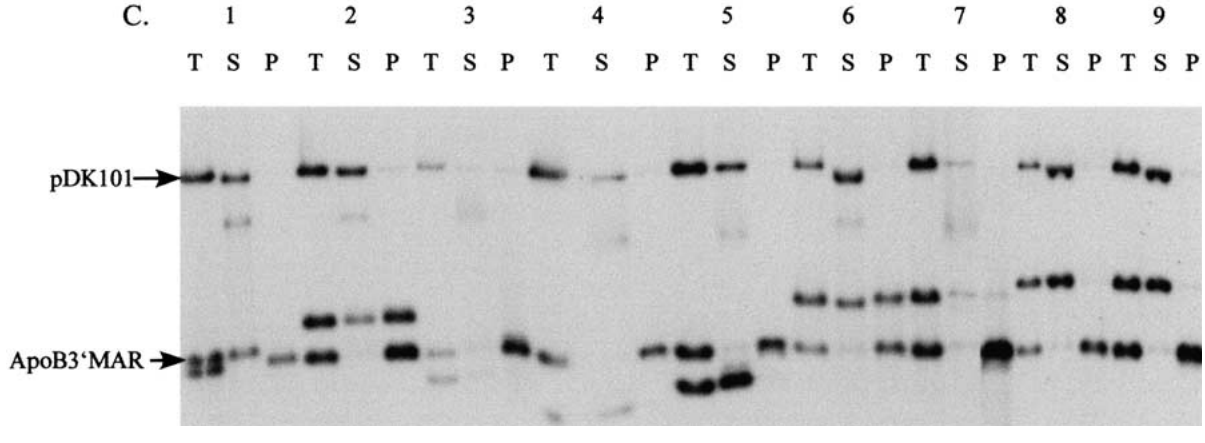

C.
1 2 3 4 5 6 7 8 9
T S P T S P T S P T S P T S P T S P T S P T S P T S P
pDK101
ApoB3'MAR

Fig. 5.

(Fig. 6C), so LINE sequences *per se* do not confer matrix-binding activity. One notable difference between DNA sequences at −69 kb and the other putative MARs in the region is the number of AT-richness rule matches. The –69 kb sequence had very few AT-rule matches compared with MARs located at +88, +109, −143, and −277 kb (Figs. 2B, 3B, 4B, 5B, and 6B).

Finally, a putative MAR located proximal of *KIAA1622* (+243 kb) had an unusual sequence composition: it did not consist of mid-repetitive DNA, but was dense with *Alu* elements and simple sequence repeats. Interestingly, this putative MAR was assigned the second highest score for matrix-binding potential among the 27 sequences identified by MAR-Wiz. This was likely owing to the presence of two blocks of sequence that contained extremely high densities of AT-richness, Topo II, curved DNA, and origin-of-replication consensus motifs (Fig. 7B). The larger (~250-bp) block of sequence was ~96% AT, whereas a smaller (~110-bp) block, ~900 distal, was ~98% AT. We divided the ~20-kb region between +228 and +248 kb into 11 DNA fragments and tested each fragment for matrix-binding activity *in vitro* (Fig. 7A). None of these DNA fragments had matrix-binding activity (Fig. 7C). Thus, MAR-Wiz predictions include both false negatives and false positives.

Discussion

The regulation of gene expression in specialized cells is a fundamental biological process, and mechanisms that control cell type-specific gene expression are intimately related to distinct ways of organizing the genome, at the level of chromatin, in different cell types. These regulatory processes often involve controls that affect genomic intervals that are hundreds of kilobases in length. It seems clear, then, that understanding the mechanisms that regulate gene expression and chromatin structure in mammalian cells will require that we are able to manipulate and analyze large genomic regions. Our approach to this problem has been to study the regulation of human chromosomal loci that have been modified specifically by targeted mutagenesis (Dieken, 1996), and the serpin gene cluster at 14q32.1 has been a useful model system for this analysis (Marsden & Fournier 2003)

Previously, the 14q32.1 serpin gene cluster was shown to comprise an ~370-kb region that included six different serpin genes (Rollini & Fournier 1997a; Rollini & Fournier 1997b; Billingsley et al. l993) Sequence annotation of the region indicates that the locus is actually considerably more complex. Our analysis of the 14q32.1 region indicates that there are actually 11 different serpin gene sequences in the interval, and a number of non-serpin sequences of unrelated function are located both proximally and distally. This genomic organization correlates well with that described in the January 2003 version of the Genoscope map (Heilig et al. 2003) and the November 2002 version of the UCSC Genome Browser map (Kent et al. 2002). There are, however, some differences in detail. For example, the *KAL-like* sequence in the distal subcluster is described as a pseudogene in the HGP map, but it appears to be transcribed in human liver (unpublished). From the perspective of serpin gene regulation, however, perhaps the most important conclusions from these studies are the identification of three subclusters within the serpin locus and the clear organization of the region into serpin and non-serpin segments. It will interesting to determine whether this organization is reflected in the chromatin structure of the locus as established by structural and functional tests.

One aspect of chromatin structure whose contribution to domain organization has yet to be established is the nuclear matrix association of specific DNA sequences. To begin to address this issue, we mapped putative matrix-attachment regions in the interval using the MAR-Wiz algorithm (Singh et al. 1997). MAR-Wiz identified 27 putative MARs in the region, but the distributions of matrix-binding potentials in the distal and proximal halves of the contig were strikingly different. The distal half of the contig contained 10 putative MARs separated by genomic DNA segments with little or no MAR activity. These inter-MAR segments were generally ~20–100 kb, in accord with previous estimates [Paulson and Laemmli 1977). In contrast, the proximal half of the contig contained 17 putative MARs, 10 of which were within genes. Most of these MARs were located in introns, but four of them contained *KIAA1622* exons. This observation is in contrast to the usual assertion that MARs commonly lie in intergenic DNA segments.

To determine the extent to which MAR-Wiz predictions reflect the true matrix-binding activities of specific DNA sequences, we compared the matrix-binding potentials of MAR-Wiz candidate sequences with their matrix-binding activities *in vitro.* Only two of five MARs previously identified in the proximal serpin subcluster (Rollini et al. 1999) were correctly identified by MAR-Wiz. This finding is in agreement with a previous study that suggested that MAR-Wiz underestimates matrix-binding activity (Glazko et al. 2001) Furthermore, we tested various MAR-Wiz candidate sequences for matrix-binding activity *in vitro.* Five of seven MAR-Wiz predictions tested were accurate, but we identified one false-

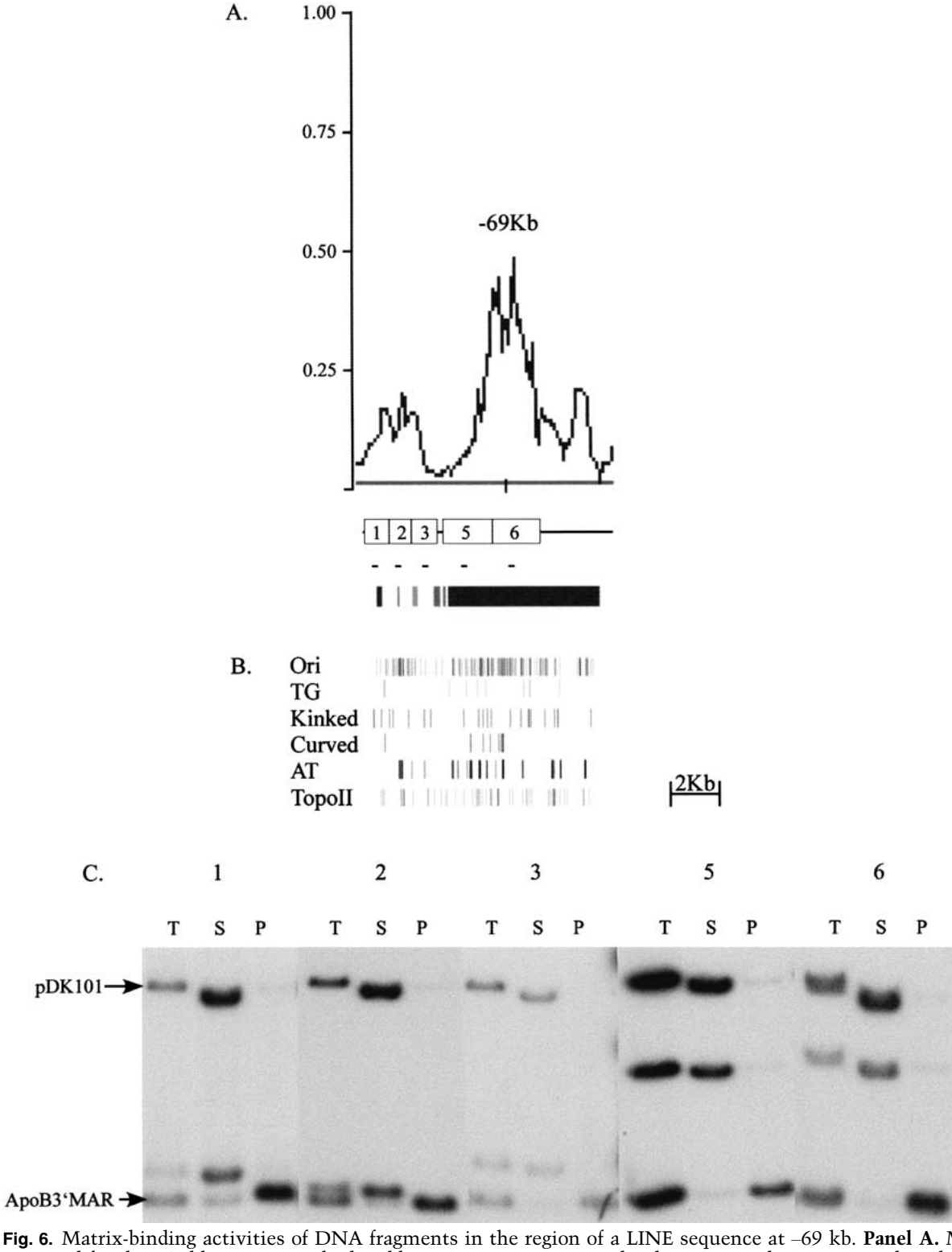


Fig. 6. Matrix-binding activities of DNA fragments in the region of a LINE sequence at –69 kb. **Panel A.** Matrix-binding potential for the –69 kb region as calculated by MAR-Wiz; no matrix-binding potential >0.5 was predicted for this region. The five numbered boxes below the graph represent DNA fragments that were isolated and tested for matrix-binding activity. The remainder of the Panel A is organized as described in the legend to Fig. 2. A large LINE sequence (dark grey) was present in fragments 5 and 6. A color version of this figure can be found at http://www.fhcrc.org/labs/fournier/mammalian-genome-supplemental-daia/Figure_6.html. **Panel B.** Sequence motifs used by MAR-WIZ to calculate matrix binding potentials, and the locations of specific motifs within the –69-kb region, as described in the legend to Fig. 2. **Panel C.** *In vitro* MAR assay. Matrix-binding activity of DNA fragments 1–3 and 5–6 was assessed as described in the legend to Fig. 2. None of the DNA fragments displayed matrix-binding activity.

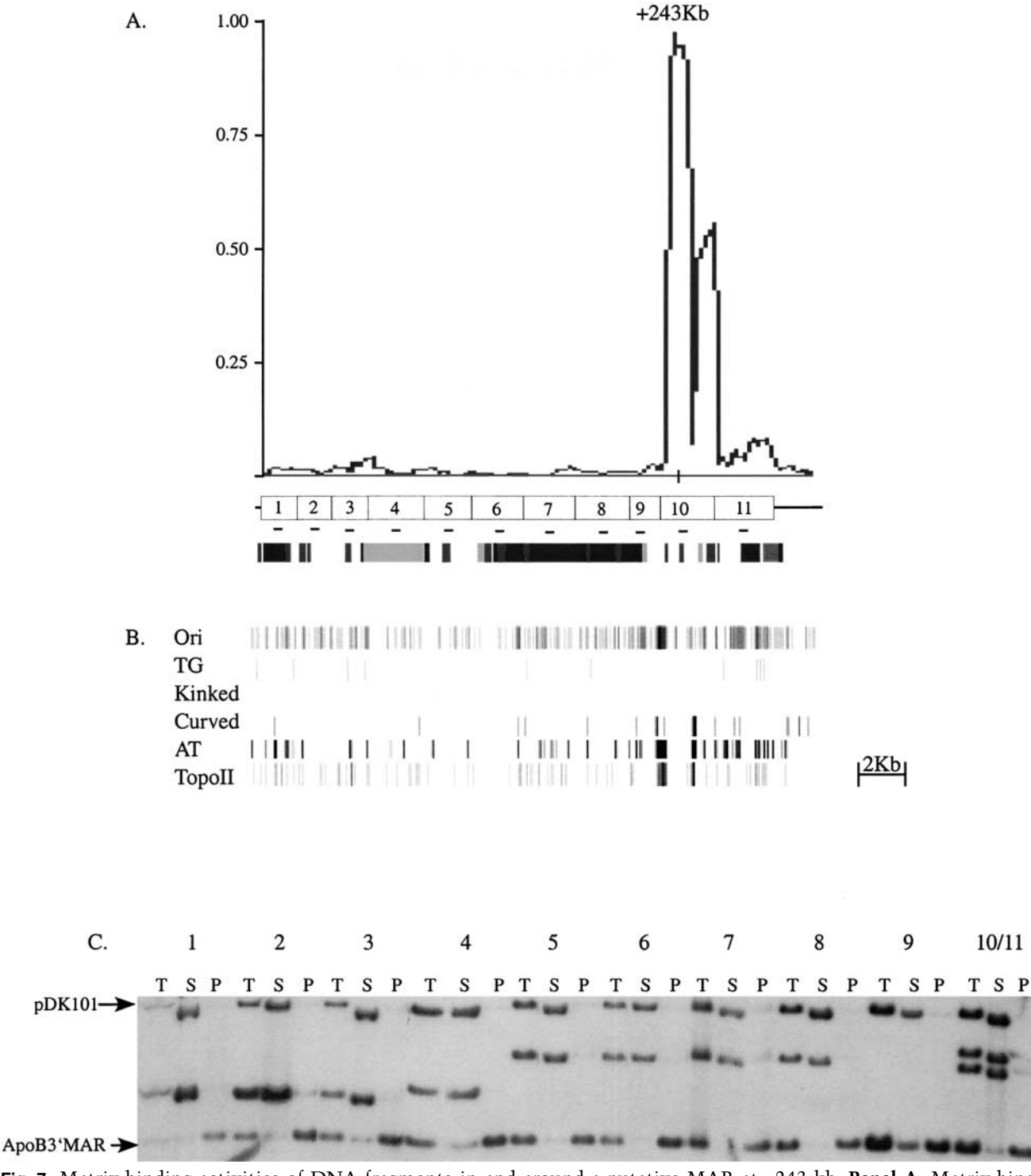


Fig. 7. Matrix-binding activities of DNA fragments in and around a putative MAR at +243 kb. **Panel A.** Matrix-binding potential for the +243-kb region as calculated by MAR-Wiz. A peak with a matrix-binding potential of ~0.9 is located at +243 kb; this was the second-highest peak of matrix-binding potential in the entire ~1-Mb region. The 11 numbered boxes below the graph represent DNA fragments that were isolated and tested for matrix-binding activity. The remainder of the Panel A is organized as described in the legend to Fig. 2. Repetitive elements to the region included LTR elements (light grey), DNA-type repeats (medium grey), and LINES and SINES (dark grey). A color version of this figure can be found at http://www.fhcrc.org/1abs/fournier/mammalian-genome-supplemental-data/Figure_7.html. **Panel B.** Sequence motifs used by MAR-WIZ to calculate matrix-binding potentials, and the locations of specific motifs within the +243-kb region, as described in the legend to Fig. 2. **Panel C.** *In vitro* MAR assay. Matrix-binding activity of DNA fragments 1–11 was assessed as described in the legend to Fig. 2. None of the DNA fragments displayed matrix-binding activity. Note that fragment 10, assigned a matrix-binding potential of ~0.9 by MAR-Wiz, did not display matrix-binding activity *in vitro*.

negative and one false-positive among the test sequences. On incorporation of results from the analyses of *bona fide* MARs in the proximal subcluster, the overall success rate of MAR-Wiz was 7 of 12 sequences tested, with four false-negatives and one false-positive. It will be important to test many more putative matrix-binding elements for matrix-binding activity *in vitro* to establish the generality of these observations. Nonetheless, it seems clear that, although MAR-Wiz is a useful tool for the *in silico* analysis of matrix-binding potential, *in vitro* assays are required to establish the true matrix-binding activities of specific DNA sequences.

DNA and chromatin maps of the 14q32.1 serpin gene cluster form an important conceptual framework for functional studies of the locus. The maps described in this report should further enhance our ability to dissect regulation within the serpin locus with molecular and genetic tests.

Acknowledgments

We thank Mark Groudine, Elaine Ostrander, and Meng-Chao Yao for their comments on the manuscript. These studies were supported by grant GM26449 from the National Institute of General Medical Sciences.

Website References

http://www.genoscope.cns.fr/, Genoscope Home Page
http://ftp.genome.washington.edu/cgi-bin/RepeatMasker, RepeatMasker
http://genome.dkfz-heidelberg.de/cgi-bin/GENSCAN/genscan.cgi, GenScan Home Page
http://searchlauncher.bcm.tmc.edu/gene-finder/gfb.html, Baylor College of Medicine Gene Finders
http://www.ncbi.nlm.nih.gov/BLAST/, NCBI BLAST Search
http://www.tigr.org/tdb/hgi/index.html, TIGR est database
http://www.futuresoft.org/MAR-Wiz/, MAR-Wiz Home Page
http://www.fhcrc.org/labs/fournier/mammalian-genome-supplemental-data/Figure_1.html
http://www.fhcrc.org/labs/fournier/mammalian-genome-supplemental-data/Figure_2.html
http://www.fhcrc.org/labs/fournier/mammalian-genome-supplemental-data/Figure_3.html
http://www.fhcrc.org/labs/fournier/mammalian-genome-supplemental-data/Figure_4.html
http://www.fhcrc.org/labs/fournier/mammalian-genome-supplemental-data/Figure_5.html
http://www.fhcrc.org/labs/fournier/mammalian-genome-supplemental-data/Figure_6.html
http://www.fhcrc.org/labs/fournier/mammalian-genome-supplemental-data/Figure_7.html
http://www.fhcrc.org/labs/fournier/mammalian-genome-supplemental-data/Primers.html
http://www.fhcrc.org/labs/fournier/mammalian-genome-supplemental-data/Table_1.html

References

1. Billingsley GD, Walter MA, Hammond GL, Cox DW (1993) Physical mapping of four serpin genes: alpha 1-antitrypsin, alpha 1-antichymotrypsin, corticosteroid-binding globulin, and protein C inhibitor, within a 280-kb region on chromosome 14q32.1. Am J Hum Gene 52, 343–353
2. Blum M, De Robertis EM, Kojis T, Heinzmann C, Klisak I, et al. (1994) Molecular cloning of the human homeobox gene goosecoid (GSC) and mapping of the gene to human chromosome 14q32.1. Genomics 21, 388–393
3. Broeker PL, Super HB, Thirman MJ, Pomykala H, Yonebayashi Y, et al. (1996) Distribution of 11q23 breakpoints within the MLL breakpoint cluster region in de novo acute leukemia and in treatment-related acute myeloid leukemia: correlation with scaffold attachment regions and topoisomerase II consenus binding sites. Blood 87 1912–1922
4. Burge C, Karlin S (1997) Prediction of complete gene structures in human genomic DNA. J Mol Biol 268, 78–94
5. Carrell RW, Pemberton PA, Boswell DR (1987) The serpins: evolution and adaptation in a family of protease inhibitors. Cold Spring Harbor Symp Quanti Biol 52, 527–535
6. Cockerill PN, Garrard WT (1986) Chromosomal loop anchorage of the kappa immunoglobulin gene occurs next to the enhancer in a region containing topoisomerase II sites. Cell 44, 273–282
7. Dieken ES, Epner EM, Fiering S, Fournier REK, Groudine M (1996) Efficient modification of human chromosomal alleles using recombination-proficient chicken/human microcell hybrids. Nature Genet 12, 174–182
8. Frazer JK, Jackson DG, Gaillard JP, Lutter M, Liu YJ, et al. (2000) Identification of *centerin:* a novel human germinal center B cell-restricted serpin. Europ J Immuno 30, 3039–3048
9. Gasser SM, Laemmli UK (1986) Cohabitation of scaffold binding regions with upstream/enhancer elements of three developmentally regulated genes of *D. melanogaster*. Cell 46, 521–530
10. Gettins PG (2000) Keeping the serpin machine running smoothly. Genome Res 10, 1833–1835
11. Glazko GV, Rogozin IB, Glazkov MV (2001) Comparative study and prediction of DNA fragments associ-

ated with various elements of the nuclear matrix. Biochim Biophy Acta 1517, 351–364
12. Hammond GL, Smith CL, Underhill DA (1991) Molecular studies of corticosteroid binding globulin structure, biosynthesis and function. J Steroid Biochem Mol Biol 40, 755–762
13. Han X, Huang ZF, Fiehler R, Broze Jr GJ (1999) The protein Z-dependent protease inhibitor is a serpin. Biochemistry 38, 11073–11078
14. Han X, Fiehler R, Broze GJ (2000) Characterization of the protein Z-dependent protease inhibitor. Blood 96, 3049–3055
15. Heilig R, Eckenberg R, Petit JL, Fonknechten N, Da Silva C, et al. (2003) The DNA sequence and analysis of human chromosome 14. Nature. 421, 601–607
16. Irving JA, Pike RN, Lesk AM, Whisstock JC (2000) Phylogeny of the serpin superfamily: implications of patterns of amino acid conservation for structure and function. Genome Res 10, 1845–l864
17. Izaurralde E, Mirkovitch J, Laemmli UK (1988) Interaction of DNA with nuclear scaffolds *in vitro*. J Mol Biol 200, 111–125
18. Jackers P, Clausse N, Fernandez M, Berti A, Princen F, et al. (1996) Seventeen copies of the human 37 kDa laminin receptor precursor/p40 ribosome-associated protein gene are processed pseudogenes arisen from retropositional events. Biochim Biophys Acta 1305, 98–104
19. Kellum R, Elgin SC (1998) Chromatin boundaries: punctuating the genome. Curr Biol 8, R521–R524
20. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, et al. (2002) The human genome browser at UCSC. Genome Res 12, 656–664
21. Levy-Wilson B, Fortier C (1989) The limits of the DNase I-sensitive domain of the human apolipoprotein B gene coincide with the locations of chromosomal anchorage loops and define the 5′ and 3′ boundaries of the gene. J Biol Chem 264, 21196–21204
22. Marsden MD, Fournier REK (2003) Chromosomal elements regulate gene activity and chromatin structure of the human serpin gene cluster at 14q32.1. Mol Cell Biol 23, 3516–3526
23. Marshall CJ (1993) Evolutionary relationships among the serpins. Phil Trans R Soc London—Ser B: Biol Sci 342, 101–119
24. Mirkovitch J, Mirault ME, Laemmli UK (1984) Organization of the higher-order chromatin loop: specific DNA attachment sites on nuclear scaffold. Cell 39, 223–232
25. Murakami K, Takagi T (1998) Gene recognition by combination of several gene-finding programs. Bioinformatics. 14, 665–675
26. Nagase T, Kikuno R, Ishikawa K, Hirosawa M, Ohara O (2000) Prediction of the coding sequences of unidentified human genes. XVII. The complete sequences of 100 new cDNA clones from brain which code for large proteins *in vitro*. DNA Res 7, 143–150
27. Paulson JR, Laemmli UK (1977) The structure of histone-depleted metaphase chromosomes. Cell 12, 817–828
28. Phi-Van L, Stratling WH (1996) Dissection of the ability of the chicken lysozyme gene 5′ matrix attachment region to stimulate transgene expression and to dampen position effects. Biochemistry 35, 10735– 10742
29. Rasmussen UB, Wolf C, Mattei MG, Chenard MP, Bellocq JP, et al. (1993) Identification of a new interferon-alpha-inducible gene (p27) on human chromosome 14q32 and its expression in breast carcinoma. Cancer Res 53, 4096–4101
30. Richterich P (1998) Estimation of errors in ''raw'' DNA sequences: a validation study. Genome Res 8, 251–259
31. Rollini P, Fournier REK (1997a) A 370-kb cosmid contig of the serpin gene cluster on human chromosome 14q32.1: molecular linkage of the genes encoding alpha 1-antichymotrypsin, protein C inhibitor, kallistatin, alpha 1-antitrypsin, and corticosteroid-binding globulin. Genomics 46, 409–415
32. Rollini P, Fournier REK (1997b) Molecular linkage of the human alpha 1-antitrypsin and corticosteroid-binding globulin genes on chromosome 14q32.1. Mamm Genome 8, 913–916
33. Rollini P, Fournier REK (1999) Long-range chromatin reorganization of the human serpin gene cluster at 14q32.1 accompanies gene activation and extinction in microcell hybrids. Genomics 56, 22–30
34. Rollini P, Fournier REK (2000) Differential regulation of gene activity and chromatin structure within the human serpin gene cluster at 14q32.1 in macrophage microcell hybrids. Nucleic Acids Res 28, 1767–1777
35. Rollini P, Namciu SJ, Marsden MD, Fournier REK (1999) Identification and characterization of nuclear matrix-attachment regions in the human serpin gene cluster at 14q32.1. Nucleic Acids Res 27, 3779– 3791
36. Rollini P, Xu L, Fournier REK (2000) Stable expression and cell-specific chromatin structure of human alpha 1-antitrypsin cosmid transgenes in rat hepatoma cells. Nucleic Acids Res 28, 3605–3614
37. Salamov AA, Solovyev VV (2000) *Ab initio* gene finding in *Drosophila* genomic DNA. Genome Res 10, 516–522
38. Singh GB, Kramer JA, Krawetz SA (1997) Mathematical model to predict regions of chromatin attachment to the nuclear matrix. Nucleic Acids Res 25, 1419–1425
39. Strausberg RL, Feingold EA, Grouse LH, Darge JG, Klausner RD, et al. (2002) Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences. Proc Nat Acad Sci USA 99, 16899–16903
40. Zhao Y, Yu L, Fu Q, Chen W, Jiang J, et al. (2000) Cloning and characterization of human DDX24 and mouse Ddx24, two novel putative DEAD-Box proteins, and mapping DDX24 to human chromosome 14q32. Genomics 67, 351–355