Nonrandom Gene Organization: Structural Arrangements of Specific Pre-mRNA Transcription and Splicing with SC-35 Domains

Yigong Xing, Carol V. Johnson, Phillip T. Moen, Jr., John A. McNeil, and Jeanne Bentley Lawrence
Department of Cell Biology, University of Massachusetts Medical School, Worcester, Massachusetts 01655

Abstract. This work demonstrates a highly nonrandom distribution of specific genes relative to nuclear domains enriched in splicing factors and poly(A)⁺ RNA, and provides evidence for the direct involvement of these in pre-mRNA metabolism. As investigated in hundreds of diploid fibroblasts, human collagen Iα1 and β-actin DNA/RNA showed a very high degree of spatial association with SC-35 domains, whereas three nontranscribed genes, myosin heavy chain, neurotensin, and albumin, showed no such preferential association. Collagen Iα1 RNA accumulates within the more central region of the domain, whereas β-actin RNA localizes at the periphery. A novel approach revealed that collagen RNA tracks are polarized, with the entire gene at one end, on the edge of the domain, and the RNA extending into the domain. Intron 26 is spliced within the RNA track at the domain periphery. Transcriptional inhibition studies show both the structure of the domain and the gene's relationship to it are not dependent upon the continued presence of accumulated collagen RNA, and that domains remaining after inhibition are not just storage sites. Results support a model reconciling light and electron microscopic observations which proposes that transcription of some specific genes occurs at the border of domains, which may also function in the assembly or distribution of RNA metabolic components. In contrast to the apparently random dispersal of total undefined hnRNA synthesis through interdomain space, transcription and splicing for some genes occurs preferentially at specific sites, and a high degree of individual pre-mRNA metabolism is compartmentalized with discrete SC-35 domains.

Despite the remarkable complexity of critical functions the nucleus performs, a simplified view of the extra-nucleolar nucleoplasm has persisted for many years. The idea that there exists some higher-level organization of the nucleoplasm that facilitates basic nuclear functions has been proposed for some time (see for example Comings, 1980; Blobel, 1985; Jackson, 1991; Lawrence et al., 1993), but the extent to which it exists is still largely unknown. It is known that the distributions of heterochromatin and satellite sequences are cell type specific (Manuelidis, 1984), that chromosomes occupy discrete nuclear “territories” (Cremer, 1982; Lichter, 1988; Pinkel et al., 1988), and that “chromosome position effects” can impact the expression of transgenes for largely unknown reasons (for example Al-Shawi et al., 1990). Relative to overall nuclear space, individual genes do not localize to precise coordinates, but distribute within preferred nuclear regions (Ward, W. S., J. A. McNeil, and J. B. Lawrence, manuscript submitted for publication; Lawrence et al., 1993). However, as explored in this work, greater order may become apparent when sequences are localized relative to defined internal reference points.

While some studies of nuclear structure investigate chromosomes or DNA within the nucleus, others focus on the localization of RNA or RNA metabolic components. These two broad aspects of nuclear structure are generally studied as separate entities and may in fact be independent of one another. Alternatively, these two compartments may be integrated, such that there is a nonrandom spatial arrangement of specific DNA sequences relative to nuclear regions devoted to the assembly, storage, transport, or activity of RNA metabolic components. This work investigates the possibility that there is nonrandom distribution of protein coding genes and/or their RNAs relative to discrete domains highly enriched in pre-mRNA splicing components and poly(A)⁺ RNA.

It is now well established that many pre-mRNA splicing components (Spector 1990; Fu and Maniatis, 1990) as well as poly(A)⁺ RNA (Carter et al., 1991; Visa et al., 1993) are highly concentrated at ~30-40 major nuclear sites or domains, in addition to a less concentrated signal dispersed throughout the nucleoplasm (reviewed in Lawrence et al., 1993; Spector, 1993; Carter, 1994; Fakan, 1994). Recent evidence indicates the presence of other nuclear matrix and SR proteins involved in pre-mRNA processing (Blencowe...
with the prominent domains highly enriched in poly (A)
mains. Earlier electron microscopic studies had described
hard, 1969; Fakan and Bernhard, 1971; Fakan and Puvion,
et al., 1994), poly(A) ÷ RNA binding protein (Krause et
The finding that poly(A) ÷ RNA concentrated in these re-
provides internal reference points with respect to which the
dependence now indicates that the IGC are largely coincident
though these may be heterogeneous categories, much evi-
dence now indicates that the IGC are largely coincident
with the prominent domains highly enriched in poly (A)
RNA or spliceosome assembly factor, SC-35 (Spector et
al., 1991; Visa et al., 1993). This work reflects a twofold in-
terest in this nucleoplasmic compartment. First, irrespec-
tive of their precise function, discrete domains detected by
anti-SC-35 immunofluorescence (Fu and Maniatis, 1990),
provide internal reference points with respect to which the
possibility of higher-level genomic organization may be
explored. Second, the potential function of this nuclear
compartiment in the biogenesis of RNA processing compo-
nents or in RNA processing itself is of obvious interest.
The finding that poly(A) ÷ RNA concentrated in these re-
gions led us to raise the question of a role in pre-mRNA
metabolism (Carter et al., 1991), an idea which would seem
to be discounted by earlier observations that short
pulses of 3H-uridine label primarily the PF, rather than the
IGC (reviewed in Fakan and Puvion, 1980; Fakan, 1994).
Biochemical studies indicate that the bulk of poly(A) ÷ RNA
gives rise to mRNA (for example, Scherrer et al.,
1970; Salditt-Geoffroy et al., 1981; reviewed in Brawer-
man, 1981; Lewin, 1990). However, recently XIST RNA pro-
vides a precedent for a poly(A) ÷ RNA that functions in
nuclear structure and appears to be long-lived in the nu-
cleus (Brown et al., 1992; Clemson et al., in press). Hence,
the finding of poly(A) ÷ RNA in domains does not ne-
equivalently show the presence of pre-mRNA (mRNA),
and the question of whether mRNA metabolism is associ-
ated with this nuclear compartment remains a subject of
interest and debate.
Studies which investigate broad classes of RNA, rely on
drug inhibition, or introduce exogenous sequences in a for-
gren structural context yield readily detectable signals which
can provide a general view, but may yield ambiguous con-
cusions. Furthermore, such studies cannot address the
fundamental question of whether specific genes localize to
and are transcribed in specific places. This requires an ap-
proach to identify directly the site of transcription and
RNA processing for specific protein coding genes. Such
methods showed the transcription and splicing site for fi-
bronectin RNA to exhibit a high degree of association,
mostly at the SC-35 domain periphery, leading us to sug-
gest that there may be a nonrandom arrangement of some
genetic loci relative to these regions (Lawrence et al.,
1993; Xing et al., 1993). Although not interpreted in this
way, the prior observation of c-fos RNA near snRNP “speckles” (Huang and Spector, 1991) is consistent with
this hypothesis.

To investigate the relationship of other endogenous genes
to this nuclear compartment, and to address key points
concerning the function(s) of domains, we applied and fur-
ther developed powerful molecular cytogenetic methods
for precise cellular localization of specific genes, RNAs,
and proteins (Langer-Safer et al., 1982; Lawrence et al.,
1989; Carter et al., 1991; Xing et al., 1993). The resulting
evidence demonstrates a nonrandom organization of spe-
cific sequences relative to these internal nuclear land-
marks, and clearly demonstrates that the transcription and
splicing of some pre-mRNAs is directly associated with
them. Results further demonstrate that collagen RNA
tracks are polar structures which extend beyond the gene
and show a reproducible orientation with SC-35 domains.

Materials and Methods

Cell Culture and Fixation
Human diploid cultures either from lung (WI-38, CCL 75, American Type
Culture Collection [ATCC], Rockville, MD) or from foreskin (Detroit
551, CCL 110) were grown in Dulbecco’s Modified Eagle’s medium, high
glucose (DME-high) supplemented with 10% FCS (GIBCO BRL, Gaith-
ersburg, MD), and 10 µg/ml gentamicin (GIBCO BRL). Before fixation,
cells were treated with cytoskeletal (CSK) buffer (100 mM NaCl, 300 mM
sucrose, 3 mM MgCl₂, 10 mM Pipes, pH 6.8; Fey et al., 1986) containing
0.5% Triton X-100 and 2 mM vanadyl ribonucleoside complex (VRC;
GIBCO BRL), for 0.5-5 min on ice. Cells were then immediately fixed in
4% paraformaldehyde (PFA) in 1× PBS (pH 7.4) for 10 min and stored at
4°C in 70% EtOH until used.

Probes and Antibodies
A genomic probe for the human neurotensin gene, hNTT14, which encom-
passes exons 1–3 was obtained from Dr. P. Dobner (University of Massa-
chusetts Medical School, Worcester, MA). For detection of the α-cardiac
myosin heavy chain gene, a genomic clone containing a 12.3-kb fragment
encompassing the 5' end of the gene was obtained from Dr. L.
Leinwand (University of Colorado, Boulder, CO). A 20-kb genomic clone
containing the human albumin gene, HSA-10, was obtained from ATCC.
Specific localization of the β-actin locus was accomplished using a 14-kb
genomic probe (pl4T B-17) encompassing the 6-kb human β-actin gene
(Leavitt et al., 1984; Ng et al., 1985). It was provided by Dr. Janet Stein
(University of Massachusetts Medical School). Studies of the collagen Iα1
and/or RNA utilized several probes. A 40-kb cosmId clone contain-
ing the entire 18-kb human type Iα1 collagen gene (CG103) was obtained
from ATCC. A 1.8-kb human collagen cDNA clone, pH F677, and a 24 kb
genomic mouse collagen Iα1 gene probe encompassing and consisting al-
most entirely of the homologous mouse collagen gene were provided by
Dr. William Strauss (Whitehead Institute for Biomedical Research, Cam-
bridge, MA). Human collagen Iα1 intron 26-specific probes were gener-
ated by PCR amplification of intron 26 sequences only, and was provided by
Dr. David Rowe (University of Connecticut, Farmington, CT). Hybrid-
ization of specific probes to metaphase chromosomes was used to confirm
probe specificity for a defined genetic locus, as described elsewhere
(Johnson et al., 1991, 1993).
A mouse monoclonal antibody against the spliceosome assembly factor
SC-35, provided by Dr. T. Maniatis (Harvard University, Cambridge,
MA) was used to delineate SC-35 domains by immunofluorescence. Anti-
SC-35 antibody detects a non-snRNP spliceosome assembly factor (Fu
and Maniatis, 1990) and may detect another structural protein which colo-
ralizes to the same nuclear domains (X.-D. Fu, personal communication).

Fluorescent In Situ Hybridization and
Immunofluorescence Staining
Probes were nick-translation using either biotin-16-dUTP or digoxigenin-
11-dUTP as the substituted nucleotide. Briefly, for hybridization to DNA,
cells were heat denatured in 70% formamide, 2× SSC at 70°C for 2 min,
followed by hybridization with labeled probe in 50% formamide 2× SSC
for 3 h to overnight at 37°C, as previously described (Johnson et al., 1991;
We previously reported the colocalization of the fibronectin gene and its RNA by targeting a nontranscribed 3' flanking sequence as a probe for DNA, and an expressed sequence probe for detection of RNA (Xing et al., 1993). This method was supplemented by the development of a sequential hybridization protocol that makes it possible to detect DNA and RNA using the identical probe detected with two different fluorochromes. Collagen Icd RNA and the gene were detected using the 24-kb genomic probe. The bound fluorescent antibody was then cross-linked by fixation in 4% PAF, rendering it impermeable to alkaline hydrolysis. The cells were then denatured in 0.07 N NaOH and then hybridized with the same sequence labeled with biotin to detect DNA specifically. Biotin was detected using FITC-avidin.

For all experiments, coverslips were rinsed in 4× SSC after the final detection incubation, when appropriate stained for 30 s in 0.1 μg/ml 4,6-diamidino-2-phenylindole (DAPI) in PBS and mounted on microscope slides in 90% glycerol/10% PBS containing 1 ng/ml phenylene diamine as an antioxidant. Slides were stored at ~20°C until analysis.

**Microscopy and Image Analysis**

A Zeiss Axioplan microscope equipped with a triple-bandpass epifluorescence filter (Chroma Technology, Inc., Brattleboro, VT) was used for sample analysis by both standard and digital imaging microscopy. Digital images were captured with a Photometrics series 200 CCD camera (Photometrics, Ltd., Tucson, AZ) equipped with a custom-made color filter wheel, such that three colors can be captured and viewed simultaneously with no optical shift (J. McNeil, personal communication). Care was taken to score cells randomly, with the only selection criteria being clear quality of the hybridization signal and the SC-35 staining, which was evaluated first using separate filter sets. Placement of the signals relative to domains was then visualized through the standard microscope using a dual-band filter and scored by two investigators independently.

Scoring was done using a 100×1.4 NA objective, which provides high-resolution and shallow depth of field, such that objects more than ~0.5 microns apart in the Z-axis will appear in a different focal plane (Carter et al., 1993). Hence, a significant degree of three-dimensional (3-D) information was discernible as cells were scored. As documented earlier by a thorough 3-D analysis using deconvolution and 3-D rendering, the SC-35 domains lie in a single plane in fibroblasts in the lower half of the nucleus (Carter et al., 1993), and hence are almost always all in focus in a single photograph. Optical sections of nuclei were obtained in 0.2 micron steps using a Z-axis stage controller, as detailed elsewhere (Carter et al., 1993).

As determined by computerized microfluorometry of ten cells, the surface area occupied by SC-35 domains averaged 22%. However, correcting for the fact that domains have a planar arrangement and appear in the focal plane in only 1/3 to 1/2 of the Z-axis using standard microscopy (Carter et al., 1993; J. McNeil, unpublished data), the expected random distribution of spots which would associate with domains would be ~8-12%. This estimate is close to and consistent with the empirical observations presented here.

**Results**

The distribution of five single-copy genes and/or their nuclear RNAs was investigated using fluorescence in situ hybridization to nonsynchronized human diploid fibroblasts, using a combination of previously described (Lawrence et al., 1989; Johnson et al., 1991; Xing et al., 1993) and new techniques (Materials and Methods). Nuclear domains enriched in the spliceosome assembly protein SC-35 were detected using immunofluorescence with anti-SC-35 antibody (Fu and Maniatis, 1990). For clarity we will refer to these regions here as SC-35 domains to distinguish them from the more diffuse component seen for a subset of SC-35 signal. The SC-35 domains colocalize with discrete nuclear domains that have previously been termed transcript domains, which are regions enriched in poly(A)+ RNA (Carter et al., 1991, 1993), and with the subset of Sm speckles that excludes coiled bodies (Lerner et al., 1981; Spector et al., 1983; Nyman et al., 1986; Fu and Maniatis, 1990). Generally 100–200 signals were examined by two investigators using 2-D analysis, and where necessary 3-D sectioning was done on a smaller cell sample. As previously described (Xing et al., 1993), “associated” denotes that the signal appears to contact the domain with no visible separation between the SC-35 domain and the gene/RNA, as opposed to “separate” in which a space between them is visible by fluorescence microscopy.

**Localization of Five Genetic Loci Relative to SC-35 Domains**

The first set of experiments addressed the question of whether there was a nonrandom organization of specific genetic loci relative to SC-35 domains. To investigate this method we examined quantitatively the distribution of five different sequences, including two transcriptionally active and three inactive genes. Since DNA signals are not detected unless cellular DNA is denatured (Lawrence et al., 1989; Xing et al., 1993), the presence of signal in the absence of denaturation confirms the expression of nuclear RNA from the gene. In all cases below, the in situ analysis of RNA expression was consistent with the predicted expression pattern of that gene in human fibroblasts. As shown in other work from our lab and confirmed below, specific pre-mRNA and gene signals overlap, and were therefore only separately visualized where necessary.

**Albumin.** The albumin gene, expressed only in liver cells (see Hawkins and Dugaiczyk, 1982), is transcriptionally inactive in human diploid fibroblasts. Hybridization generally produced two small fluorescent spots in samples where cellular DNA was denatured, in keeping with single-copy gene detection. As illustrated in Fig. 1F, and summarized in Table I, analysis of 100 signals for the spatial relationship of this gene to SC-35 domains showed 86% of genes clearly separate from domains. The remaining signals (14%) showed no visible separation by fluorescence microscopy, and hence were scored as associated. Interestingly, in most cells the gene localized near the nuclear periphery (Fig. 1F), in a region devoid of poly(A)+ RNA/splicing factor-rich domains in human fibroblasts (Carter et al., 1991, 1993), and an area in which heterochromatin is known to concentrate in many cell types (Lentz, 1971).

In human fibroblasts, poly(A)+ RNA/SC-35 domains occupy ~5% of nuclear volume, as measured in deconvoluted and 3-D restored digital images (Carter et al., 1993). However, as considered in Materials and Methods, based on both average surface area and Z-axis distribution of domains using the optics applied here, a spot randomly distributed in the nucleus would appear associated with SC-35 domains on the order of 8–12% of the time (see Materials and Methods). We conclude, therefore, that the distribution of the albumin gene with respect to SC-35 domains is augmented by the SC-35 domains.
Figure 1. Localization of genes and/or their transcripts relative to SC-35 domains. Genes were probed in normal human diploid fibroblasts using fluorescence in situ hybridization in conjunction with immunofluorescence staining using an antibody to SC-35. (A) Collagen Iα1 transcripts (red), visualized using a full length genomic probe, colocalize with SC-35 domains (green, overlap of red and green appears yellow) (2,500×). (B) Collagen Iα1 transcripts (green) hybridized with a cDNA probe (2,500×). (C) Collagen Iα1 gene (red) localizes at the edge of SC-35 domains (green) (2,500×). (D and E) Optical sections along the Z-axis of a single SC-35 domain (green) illustrating the location of (D) collagen Iα1 transcripts (red) that are coincident in all planes with the domain, and (E) the collagen Iα1
Table I. Localization of Genes and/or RNAs Relative to SC-35 Domains

<table>
<thead>
<tr>
<th>Gene</th>
<th>No. scored</th>
<th>% Associated</th>
<th>% Separate</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Albumin</td>
<td>160</td>
<td>14</td>
<td>86</td>
<td>Nuclear periphery</td>
</tr>
<tr>
<td>c-Myosin Heavy Chain</td>
<td>217</td>
<td>22</td>
<td>78</td>
<td>Near nucleolus</td>
</tr>
<tr>
<td>Actin Total</td>
<td>290</td>
<td>89</td>
<td>11</td>
<td>Nuclear periphery</td>
</tr>
<tr>
<td>Neurotensin</td>
<td>135</td>
<td>83</td>
<td>17</td>
<td>Edge of SC-35 domain</td>
</tr>
<tr>
<td>Collagen Total</td>
<td>104</td>
<td>98</td>
<td>2</td>
<td>Edge of domain</td>
</tr>
</tbody>
</table>

in keeping with a random distribution or a preferential separation from domains that falls within the limitations of our light microscopic technique (Materials and Methods).

Cardiac MHC. The human cardiac myosin heavy chain (cMHC) gene is expressed exclusively in cardiac or skeletal muscle and thus is inactive in fibroblasts (Lompre et al., 1984; Saez et al., 1987). Unlike the more peripheral albumin gene, the MHC gene distributed more internally and frequently near the nucleolus, consistent with its chromosomal location close to the ribosomal RNA genes on chromosome 14. As illustrated in Fig. 1 G, despite the fact that this gene is in a nuclear region where there is a preponderance of poly(A)^+ RNA/SC-35-rich domains in human fibroblasts (Carter et al., 1991, 1993), the vast majority (78%) of cMHC genes were not associated with any SC-35 domain. Under the optics used, 22% appeared to be in contact with a domain, and were hence scored as associated (Table I). While this frequency is slightly elevated compared to albumin, it does not differ clearly from random, especially given the localization of the gene in a region enriched in domains. Localization of cMHC near the nucleolus is likely related to its linkage to the rRNA genes, but is also consistent with a localization with extra-nuclear heterochromatin.

Neurotensin. The neurotensin gene, transcribed only in certain types of neural cells (Kislauskis et al., 1988), is transcriptionally inactive in human diploid fibroblasts. Analysis of over 100 cells showed 89% of neurotensin gene loci visibly separate from the SC-35 nuclear domains (Table I), with only 11% appearing to contact domains. Most neurotensin signals were localized to the extreme periphery of the nucleus, close to the nuclear envelope. Hence the inactive neurotensin gene is not preferentially associated with SC-35 domains, and may reside in or near peripheral heterochromatin.

β-actin. The location of the human β-actin gene and RNA, which is expressed in fibroblasts, was examined using a 14-kb genomic probe that specifically detects the locus encompassing the 6-kb β-actin gene as confirmed by hybridization to metaphase chromosomes (Fig. 1 H). Nuclear hybridization to the β-actin gene produced two small spots, and hybridization to RNA produced two small foci only slightly larger than the DNA. Detection of DNA and RNA simultaneously showed the same number of signals as when RNA was detected alone, confirming that the DNA and RNA signals overlapped. Initial analysis of 155 β-actin gene or gene and RNA signals showed a very high degree of association (94%). Subsequent hybridizations detecting β-actin RNA alone (non-denatured samples) confirmed that the gene was transcriptionally active and that the foci of nuclear RNA showed a high degree of association. Of 135 RNA signals scored, 83% were associated. In total 290 actin signals were scored by three investigators, with an average of 89% association with SC-35 domains (Table I).

For several reasons the details of the spatial association are important (see Discussion). All associated signals were categorized as either within, meaning the signal overlapped the interior or central SC-35 domain, or border, indicating it localized at the extreme periphery or edge of the domain (Xing et al., 1993). As illustrated in Fig. 1, I-L, ~94% of the associated actin gene or RNA signals localized at the border, directly adjacent or abutting the boundary of the SC-35 rich region, whereas only 6% appeared inside the domains (Table I). Most signals associated with a large domain, with ~16% associated with a small but discrete SC-35 domain (see Fig. 1, I and J, inset). The gene or RNA signals from homologous chromosomes generally did not associate with the same domain but with two different domains, and while the two alleles generally were both associated, in...
Figure 2. (A) Collagen Iα1 intron 26 probe (red) and collagen Iα1 cDNA (green, overlap of red and green appears yellow) simultaneously hybridized in human fibroblasts (2,500×). Insets show additional examples of intron sequences (red) and cDNA sequences (green) from other cells (4,000×). Note that intron sequences are generally localized to the edge of the track. (B) Colocalization of collagen Iα1 intron 26 sequences (red) and SC-35 domains (green) (2,500×). Inset shows enlarged example from a different cell, illustrating intron signal at the inner periphery of the domain (4,000×). (C and D) Collagen Iα1 gene (blue-green) remains associated with SC-35...
occasional cells one allele associated and the other did not (see bottom cell, Fig. 1 I).

**Collagen Iα1.** The distribution of the transcriptionally active collagen type Iα1 gene/RNA was investigated using a genomic probe encompassing the transcription unit for this 18 kb gene (Barsh et al., 1984), which detected two specific DNA or concentrated RNA signals in most nuclei. Simultaneous hybridization to DNA and RNA, versus DNA or RNA alone, confirmed that the gene signal was a small spot coincident with a larger accumulation of RNA. Although not highly elongated tracks, the RNA accumulations were substantially larger than the small foci observed for actin RNA, and were track-like in that they often had some longitudinal axis. As illustrated in Fig. 1, A and B and quantitated in Table I, essentially all collagen RNA signals spatially associated with large distinct SC-35 domains. In contrast to actin RNA, collagen RNA consistently appeared to occupy the inner regions of the domains.

Using high resolution optics which allow discrimination of objects more than 0.5 microns apart in the Z-axis, the collagen RNA signal consistently appeared to be within the SC-35 domain. This conclusion was confirmed by optical sectioning of several cells by digital imaging microscopy (Fig. 1 D) that showed the collagen RNA occupied the same focal planes as the SC-35 enriched region, and therefore overlapped it in 3-D space. The domains with which collagen RNA associated were among the largest within the nucleus (~1-3 microns; Carter et al., 1993). Some collagen RNA foci or tracks had a similar size and shape to their associated domains (compare Fig. 1 B), although sometimes (in some experiments as high as 80%) the collagen nuclear RNA appeared to concentrate in one part of the SC-35 domain (Fig. 1, A and D).

It is important that sequence distributions be determined in many cells (see Discussion). However, even in nonsynchronized cultures, the gene distributions were reproducible from experiment to experiment and distinct distributions were obvious from just brief viewing blind through the microscope. Irrespective of the function of domains, these results demonstrate a nonrandom organization of genomic loci relative to splicing factor rich domains. While we do not suggest that all transcriptionally active genes or pre-mRNAs associate with domains (see Discussion), these and other results support that the association of the two active genes with SC-35 domains, as contrasted with the three inactive genes, is not coincidental but related to their production or maturation of pre-mRNA.

**Structural Relationship of Collagen Iα1 Transcription and Splicing with SC-35 Domains and RNA Tracks**

The collagen Iα1 gene is one of the cell's most highly active genes, accounting for approximately 4% of total mRNA in fibroblasts and producing an RNA with 51 introns (Genovese et al., 1989). From some perspectives it might be considered surprising to find pre-mRNA from such a highly active gene associated with SC-35 domains (see Discussion). The collagen Iα1 gene and RNA were analyzed in much greater detail to address several key points as to the relationship of structure and function between RNA foci or tracks and the SC-35 domains.

**Transcription at the Border of SC-35 Domains**

To determine the precise spatial relationship of the collagen gene relative to the SC-35 domain, hybridizations were performed under conditions which remove RNA and denature cellular DNA, using the same full-length genomic probe. As shown directly in Fig. 1 C, the collagen gene signal was associated with the prominent SC-35 domains in ~99% of cells. But unlike the RNA that localized within the domain, the DNA signals localized as a spot at the domain border, with no visible separation between the gene and the edge of the SC-35 region. It was not possible to discern unequivocally on which side of the border the gene was positioned, since it sometimes appeared on the outer border, but frequently appeared just within the border. Interestingly, when the domain with which the gene was associated was elongated, in most cases the gene positioned at one end of the linear axis (see Fig. 1 C).

A curious and interesting aspect of our results is that in the vast majority of cells the collagen gene appeared adjacent to the domain in the X-Y plane, such that the gene was in focus with the boundary of the domain. The gene only rarely appeared to be above or below the domain. This impression was confirmed by optical sectioning of nu-
nuclei, as illustrated in Fig. 1E which showed the gene next to the domain, as opposed to the RNA that lay within the domain (Fig. 1D). These results clearly demonstrate that the transcription site of collagen is not only associated with SC-35 domains but displays a reproducible spatial position at the boundary or periphery of the SC-35 rich region. This point is pivotal to building a model of the structure and function of SC-35 domains (Discussion and Fig. 2f).

**Reproducible Polarity and Orientation of Collagen RNA Tracks**

Since the collagen gene signal positions at the edge of the domain and the collagen RNA concentrates within it, this strongly implies a polarity of the gene position relative to the RNA accumulation. The question of whether the RNA accumulation extends beyond the dimensions of the gene is important to understanding whether it comprises strictly nascent transcripts or a posttranscriptional accumulation of RNA. The ~18-kb collagen gene provides an ideal model to examine this, since overwhelming evidence indicates that sequences of this relatively small size are below the resolution of the light microscope, and produce just a single round spot of signal in unfractionated interphase nuclei (Lawrence et al., 1988, 1990; Trask et al., 1991; Wiegant et al., 1992; Gerdes et al., 1994). As expected, in numerous experiments the genomic probe encompassing the entire transcribed region of collagen 1α1 consistently appeared as a small round spot (Figs. 1C and 2E and F). That the gene was not packaged as a large extended loop was further confirmed by experiments in which hybridization of a 5.6-kb sequence near the middle of the gene produced overlapping signal with hybridization to the rest of the sequences from the full-length genomic probe (not shown).

To compare the DNA and RNA signals by the most direct means possible, a novel procedure involving sequential hybridization was developed which allowed simultaneous visualization of DNA and RNA in different colors using the same sequence as a probe (see Materials and Methods). A 24-kb mouse genomic probe encompassing almost exclusively the full transcription unit of the collagen gene (~20 kb) was used for this purpose in mouse fibroblasts. Results were striking. As shown in Fig. 2, E and F, this revealed that the collagen gene was clearly positioned at or very near one end of the RNA "track." This was the case in 124 out of 145 cases examined. As illustrated in the polyploid cell shown in Fig. 2F, even in a cell where the RNA signals are just small foci with no apparent linear axis, a prominent accumulation of RNA is most frequently observed to the side of the gene signal, producing an appearance of a comet consisting of a gene and a trailing track of RNA.

This clear polarity of the gene relative to the RNA track is consistent with and confirms the above observations which independently showed the collagen gene at the boundary of an SC-35 domain and the RNA within the domain. These experiments support that collagen RNA tracks or foci are formed by an accumulation of RNA which extends from and beyond the dimensions of the gene; hence even small foci may be RNA accumulations at some posttranscriptional step, rather than solely nascent transcripts directly on the gene. These RNA tracks extend vectorially from the gene into the SC-35 domain.

**Collagen 1α1 RNA Splicing in the Domain Periphery**

Why would there be an accumulation of RNA adjacent to the gene? Does the RNA track contain both introns and exons, spliced mRNA, or might this be an accumulation of excised introns? To address these questions and determine directly whether RNA splicing localized with the domain, a series of experiments was done using a combination of genomic, cDNA and intron probes. Sequences detected with a cDNA probe showed high concentrations throughout much of the SC-35 domain, similar to RNA detected with the genomic probe (Fig. 1B). Direct comparison of the signal from the cDNA and full-length genomic probes in two colors showed that these completely overlap (not shown). This demonstrates that no appreciable portion of the collagen 1α1 RNA tracks consisted of accumulated introns alone, ruling out the possibility that RNA within the domain consisted of excised introns.

Analysis was next done using an intron-only probe which detected sequences from intron 26 (143 bp). As illustrated in Fig. 2B, colocalization of intron sequences relative to SC-35 demonstrated that, in contrast to results with the cDNA and genomic probes, the intron was detected as a small focal signal in the periphery of the domain, rather than broadly throughout the central domain. This peripheral concentration was observed in the vast majority of cells; only infrequently was there a weaker more dispersed signal within the domain. Comparison of intron signal directly with cDNA showed that intron 26 sequences occupied only a portion of the track defined by the cDNA (Fig. 2A), with intron consistently concentrated towards one end.

The absence of intron from a portion of the cDNA track indicates that splicing of this intron is occurring within the track. Since collagen RNA tracks associate with SC-35 domains, both splicing and transcription of collagen RNA occurs in intimate association with the domain, generally near the periphery, as schematically illustrated in Fig. 2H. Two versions of the model are presented, illustrating cotranscriptional splicing of introns restricted to nascent transcripts on the gene (top) or posttranscriptional splicing at the periphery of the domain (bottom). Given that both gene and intron RNA localize near the domain periphery, this indirectly indicates that splicing of intron 26 is spatially proximal to the gene, if not at the gene itself. Cotranscriptional splicing is known to occur for some RNAs (Beyer and Osheim, 1988); however because the intron may extend slightly further into the domain, and because it is possible that the RNA's entry into a splicing factor-rich region is related to its splicing, from our data either version of the model is plausible. In both versions, the bulk of the RNA throughout the domain has been spliced, specifically with respect to intron 26.

**RNA Tracks Extend into the Domain but Not to the Nuclear Envelope**

The polar relationship between the collagen gene and its RNA accumulation suggests that there is movement of the RNA and gene with respect to one another, such that
the RNA extends from the gene into the domain in a vectorial manner. Consistent with the observation that SC-35 domains do not position adjacent to the nuclear envelope (Carter et al., 1993), collagen RNA tracks, which occupy SC-35 domains, do not appear to extend to the nuclear envelope. As shown in Fig. 2 G, dispersed collagen RNA signal is sometimes observed which may represent nuclear RNA in transport towards the nuclear envelope. However, no one clear track is seen to the envelope, suggesting that this step may occur very quickly and/or the RNA could disperse in multiple directions, consistent with previous results for endogenous fibronectin RNA (Xing and Lawrence, 1993; Xing et al., 1993). Fig. 2 G shows a ring of collagen RNA studding and encircling the nuclear envelope (inset); to the extent that this may reflect RNA as, or shortly after, it exits the nucleus, this would suggest that collagen RNA exits many or all of the nuclear pores and not just a specific region. While it was occasionally observed that there was more RNA in the cytoplasm nearest the nuclear RNA track, this was not commonly observed, and the distribution of the RNA in the cytoplasm was most consistent with random association of the mRNA with the endoplasmic reticulum (Fig. 2 G). As viewed in two dimensions, no specific orientation of the gene and RNA tracks relative to the nuclear envelope was noted, although more analysis would be required to demonstrate this rigorously.

In summary our results indicate that visible nuclear collagen RNA tracks lead into SC-35 domains, but do not directly extend to the nuclear envelope.

**The Collagen Iα1 Gene Remains Associated with a Prominent SC-35 Domain after Transcriptional Inhibition**

Finally we investigated whether the collagen gene remained associated with SC-35 domains in cells treated to inhibit transcription. As considered under discussion, such studies have produced variable and enigmatic effects on domain structure (Lawrence et al., 1993). Several studies report that IGC become fewer and larger and correlate these with the round prominent SC-35 domains observed after inhibition (Davis et al., 1993; Visa et al., 1993; Huang et al., 1994). In one case these studies have interpreted to support that the large prominent SC-35 domains or IGC which exist in uninhibited cells are unassociated with pre-mRNA metabolism, but are sites devoted to storage of splicing factors (Huang et al., 1994; reviewed in Mat-taj, 1994). To provide a critical test of this interpretation of inhibition studies, we colocalized the collagen gene and SC-35 after transcriptional inhibition. This experiment simultaneously addresses two other key questions: (a) Is the domain of SC-35 an accumulation of splicing factors dependent on the continued presence of accumulated collagen RNA? (b) Is the association of the collagen gene with SC-35 domains dependent on its continued transcription?

As shown in Fig. 2, C and D, transcriptional inhibition for 3 h resulted in prominent round SC-35 domains in many though not all cells. Under these conditions, collagen RNA tracks are generally undetectable or extremely reduced (Clemson et al., in press; and C. Clemson, data not shown). Hybridization to the collagen gene in denatured samples revealed that the gene remained precisely positioned at the edge of these round large structures, as shown in Fig. 2, C and D. In the vast majority of cells the collagen genes remained directly abutting the SC-35 domain. Neither the presence of the domain nor the gene’s relationship to it is dependent upon the continued presence of accumulated collagen RNA. These results also show that inhibition studies cannot be reliably interpreted to show that domains which remain or rearrange after inhibition are solely storage sites, as our results demonstrate that they correspond to the transcription and splicing site of one of the cell’s most active genes.

**Discussion**

This work details many new observations which substantially enhance an emerging view of a compartmentalized and structured nucleoplasm. We demonstrate that individual genes display distinct and nonrandom distributions relative to domains enriched in SC-35 and poly(A)+ RNA, with two transcribed sequences specifically associating with discrete domains. Three transcriptionally inactive sequences do not. β-actin RNA associated at the periphery of the domain, whereas collagen Iα1 RNA accumulates within the central region of the domain. The focus of collagen RNA contains exon sequences, is not an accumulation of excised introns, and has undergone splicing, as shown for intron 26. Although collagen RNA accumulates within the SC-35 domain, gene transcription occurs just at its border. RNA splicing of intron 26 occurs at the periphery of the domain, likely proximal to the gene. Two-color DNA/RNA hybridization with a full-length genomic probe shows that collagen RNA tracks or foci are polar accumulations of RNA which extend beyond the dimensions of the gene, suggesting a posttranscriptional RNA accumulation beyond the nascent transcript “tree.” This accumulation extends vectorially from the gene into the domain, and therefore may involve transport into the domain, but the track does not appear to extend to the nuclear envelope. The results presented also clearly show that neither the presence of the SC-35 domain nor the gene’s association with it is dependent on the continued transcription or accumulation of collagen RNA. The prominent domains that remain after inhibition with actinomycin D correspond to the transcription and processing site of one of the cell’s most highly active genes, and therefore cannot reliably be interpreted as solely storage sites unassociated with pre-mRNA metabolism.

Although gene position in overall nuclear space is more variable (Ward, W. S., J. A. McNeil, and J. B. Lawrence, manuscript submitted for publication; Lawrence et al., 1993), when viewed relative to SC-35 domains a statistically nonrandom distribution of specific sequences is clearly demonstrated. Irrespective of the relationship to function, these clear differences in gene distribution point to a previously unappreciated aspect of interphase genome packaging and a structural integration of two distinct nuclear compartments. We do not suggest that all active genes are associated with SC-35 domains (see below); however the two examined here, collagen and β-actin clearly show a specific preferential association. In contrast, three transcriptionally inactive genes (cMHC, neurotensin, and...
Transcription and Splicing of Some pre-mRNAs Is Directly Associated with SC-35 Domains

While it has been thought that prominent domains, which correspond largely to IGC, were not involved in short lived pre-mRNA metabolism (Fakan and Bernhard, 1971; Fakan and Puvin, 1980; Spector, 1993; Mattaj, 1994; reviewed in Fakan, 1994), the finding of poly(A)+ RNA (Carter et al., 1991, 1993; Visa et al., 1993) and microinjected intron-containing globin RNA (Wang et al., 1991) in all of these domains justified a reevaluation of this view. In this work, we show that pre-mRNA metabolism for two of the most active genes in fibroblasts, collagen Iα1 and β-actin, is directly associated with SC-35 domains. These results, together with the preferential association of fibronectin transcription/splicing with domains (Xing et al., 1993), and the proximity of induced c-fos nuclear RNA to domains (Huang and Spector, 1991) argues strongly that there is pre-mRNA metabolism associated with these regions, with transcription at the periphery. Localization to the central domain is not unique to COL1α1 RNA, as thus far two other pre-mRNAs have been found which show similar localization (J. Coleman, C. Johnson, P. Moen, and J. Lawrence, unpublished data).

While we believe the association is related to function of these specific genes, we do not suggest this to be a property of all transcriptionally active DNA, or even of all intron-containing protein-coding RNAs (Lawrence et al., 1993). Of 10 active genes examined thus far in our lab, 7 have been found associated and 3 have not; in contrast, none of several inactive genes have yet been found associated (this work and Carter et al., 1991; Xing et al., 1993; Clemson et al., in press; J. Coleman and P. Moen, unpublished data). A cell cycle-regulated histone gene showed a mixed result (Xing, 1993), which is under investigation in synchronized cells. Further support that association is related to expression is provided by work showing that two genes which associate with domains in expressing muscle cells do not associate with domains in nonexpressing myoblasts (P. Moen, unpublished data).

Recently, a report focusing on Adenovirus localization included what appears to be a cursory analysis of the β-actin RNA transcription site relative to SC-35 domains in five cells (Zhang et al., 1994). Reporting 4 of 10 signals associated, the authors conclude that the distribution is random, using a random expected frequency that apparently does not correct for Z-axis considerations. Based on our analysis of hundreds of signals from the identical gene/RNA relative to the same structures detected with the same SC-35 antibody (Xing, 1993 and this work), the very small sample size of Zhang et al. (1994) is insufficient to draw a conclusion, and may not be representative. Most importantly, the general conclusion these authors draw from their data, that there is no compartmentalization of gene transcription and splicing with nuclear domains or speckles, is only not contradicted directly by results presented here but argued against by a variety of earlier observations (Walton et al., 1989; Carter et al., 1991; Huang and Spector, 1991; Wang et al., 1991; Xing, 1993; Xing et al., 1993; Jiménez-García and Spector, 1993).

Total Uridine Incorporation: Transcription of What?

The tenet or expectation that prominent SC-35 domains would not be associated with short-lived pre-mRNA or mRNA stems largely from uridine incorporation studies, which clearly show that general transcription occurs primarily in the interdomain (or inter-IGC) space (Fakan and Puvin, 1980; Warsink et al., 1993). However, a key question is, what classes of RNA does this transcription represent (Moen et al., 1995)? Uridine labeling has in some cases been equated with pre-mRNA for protein coding genes (Mattaj, 1994). However, there are many caveats to this assumption, as considered previously (Carter et al., 1993; Visa et al., 1993; Moen et al., 1995). For instance, most uridine label (possibly as much as 75–90%) will not be in mRNA precursors, but in other classes of RNA, especially noncoding hnRNAs of unknown function, and in intron sequences that may be rapidly spliced and dispersed from the pre-mRNA molecule (for example, Salditt-Georgieff et al., 1981; Lewin, 1975; Brawerman, 1981; Beyer and Osheim, 1988; Lewin, 1990; Xing et al., 1993). Some labeling of IGC and SC-35 domains does occur (Fakan and Bernhard, 1971; Raska et al., 1991; Jackson et al., 1993; Warsink et al., 1993), and studies using Br-UTP are impacted by the fact that the Br-UTP inhibits splicing (Warsink, 1994). Most importantly, despite the impression of random transcription, these studies cannot address the question of whether transcription and splicing of specific genes is randomly localized. Our results provide direct evidence that it is not.

Two exceptions to the generalization that poly(A)+ RNAs encode mRNAs have recently been described; Omega-n in Drosophila (Hogan et al., 1994) and XIST RNA, which associates with the inactive X chromosome of mammals (Brown et al., 1992; Clemson et al., in press). XIST RNA represents a precedent for an apparently long-lived structural RNA that functions in the nucleus; other nuclear RNAs likely contribute significantly to the label observed with uridine incorporation. Although most are not polyadenylated (Salditt-Georgieff et al., 1981; reviewed in Lewin, 1990), some long-lived nuclear RNAs could potentially contribute to the poly(A)+ RNA distributions observed.

SC-35 Domains Are Not Just Storage Sites: Inhibition of What?

Differing effects of transcriptional inhibition have been observed which result in complex and variable rearrangements or loss of nuclear poly(A)+ RNA (Lawrence et al., 1993; Moen et al., 1995). Contradictory conclusions have been reported which support that poly(A)+ RNA in discrete domains (or IGC) is short-lived pre-mRNA (Visa et al., 1993) or, alternatively, argue that it is long-lived structural RNA (Huang et al., 1994). In many cases, domains break down or dissipate. Even where there clearly is a retention and redistribution of nuclear poly(A)+ RNA after inhibition (Lawrence et al., 1993; Huang et al., 1994; Moen et al., 1995), interpretation of such results is compromised by the secondary effects on global nuclear function (reviewed in Brasch, 1990), such as inhibition of RNA trans-
port and message-specific increases in RNA stability and poly(A)⁺ tail length (for example see Herman and Pen-
man, 1977; Herman et al., 1976, Brawerman, 1987; Hogan et al., 1994). The recent observation that XIST RNA is
largely retained after inhibition whereas collagen RNA is not lends some credence to the idea that two classes of
poly(A)⁺ RNA with different stabilities may be detected (Xing, 1993; Clemson et al., in press). However XIST
RNA does not localize to poly(A)⁺ RNA-rich domains (Clemson et al., in press). It is an intriguing possibility that
there may be some structural poly(A)⁺ RNAs in domains and potentially elsewhere in the nucleus (Lawrence et al.,
1993; Huang et al., 1994), but this would by no means pre-
clude the presence of pre-mRNA in those same regions.

An important result presented here clearly indicates that the subset of domains which remain after transcriptional
inhibition cannot simply be interpreted solely as storage sites or as not directly involved in pre-mRNA metabolism,
as has been suggested (Huang et al., 1994; O’Keefe et al.,
1994; Mattaj 1994). We demonstrate here that domains re-
main in actinomycin inhibited cells correspond to the
transcription and splicing site of one of the cell’s most
highly active genes.

A Model for Domain Substructure: Transcription at
SC-35 Domain Borders

A key point of our results is that transcription occurs at or
near the border of SC-35 domains, as suggested by studies
quantitatively demonstrating the peripheral position of the
fibronectin gene/RNA (Xing et al., 1993). Subcom-
partmentalization of light microscopic domains was sug-
gested by Carter et al. (1993) (Fig. 2 l). We propose a
model showing a high-level of transcription and splicing of
some specific genes in the outer rim of the light micro-
scopic domain, abutting the SC-35 core (Fig. 2 l). Al-
though indirect, several results support that this SC-35
core corresponds to IGC; particularly our finding that the
gene remains associated with a domain in inhibited cells,
since such domains have been directly correlated to IGC
(Davis et al., 1993; Visa et al., 1993; Huang et al., 1994).

The subcompartmentalization model (Fig. 2 l) potentially
reconciles results from our lab and others with earlier elec-
tron microscopic reports, some of which showed heavy ur-
dine labeling at the borders of IGC, as well as the per-
eipheral association of perichromatin fibrils (Fakan and Puvion,

Several points indicate that SC-35 domains in general
cannot simply be explained by the presence of splicing
components wherever there are individual intron-containing
RNAs. (a) Some specific RNAs (e.g., β-actin and fi-
bronectin) clearly do not overlap most of the domain, but
are smaller and peripheral to it. (b) While the association
of the collagen gene likely depends upon its transcrip-
tional history within the cell, the large SC-35 domain with
which the gene associates remains after collagen RNA
synthesis is inhibited and the RNA accumulation is essen-
tially undetectable. (c) We have found large accumu-
lations of specific intron-containing RNAs that do not asso-
ciate with concentrated accumulations of SC-35, hence
RNAs are likely processed by more diffusely distributed
splicing factors (Clemson et al., in press; Coleman, J. R.,
P. T. Moen, and J. B. Lawrence, manuscript in prepara-
tion). (d) The number of discrete domains of SC-35 (up to
~50) is far fewer than the number of active genes. We
note that points 3 and 4 rely on an assumption that essen-
tially the same splicing factors are involved in processing
all pre-mRNAs, which may not be the case (Fu, 1993).

It will be important to determine directly whether multi-
ple pre-mRNAs associate with an individual domain. Rec-
ognizing that not all prominent domains (or IGC) may be
the same functional entities, we suggest two general ideas
for what some or all domains might be if they are indeed
more than the accumulation of splicing components asso-
ciated with transcripts of a single gene: (a) Domains may
represent the association of some transcriptionally active
genes with centers involved in maturation and transport of
RNA and/or RNA metabolic components. (b) Domains
may reflect the accumulation of splicing components on
pre-mRNA from multiple specific genes, which associate
due to either the mutual affinity of their RNAs for splicing
components or their genomic organization, or both. Either
possibility could facilitate the expression of an mRNA by
raising the concentration of RNA metabolic components
in the immediate vicinity and would involve nonrandom
compartmentalization of specific genes relative to regions
of enhanced RNA metabolic activity, as previously sug-
gested (Carter et al., 1991; discussed in Lawrence et al.,
1993).

Sequence-specific Structural Arrangements
within the Nucleoplasm

Results presented here are indicative of a structured nu-
clear interior. While splicing factor/poly(A)⁺ RNA rich
domains reside in regions of little to no DNA (Spector,
1990; Carter et al., 1991; Cremer et al., 1993; Visa et al.,
1993; Zirbel et al., 1993), we view it unlikely that localiza-
tions observed merely reflect diffusion of splicing compo-
ments or RNAs within inter-chromosome space (see for
example Zachar et al., 1993, Cremer et al., 1993). The nu-
clear splicing factor/poly(A)⁺ RNA-rich compartment has
itself a specific, reproducible 3-D topology that can vary
with cell type (Carter et al., 1993; Lawrence et al., 1993).
Collagen RNA tracks are highly resistant to triton extrac-
tion and remain after nuclear matrix factionation (Clem-
sen et al., in press). There may well be DNA and RNA or
splicing components which randomly distribute and freely
diffuse within the nucleus. However, the results presented
here show nonrandom distributions of specific genes, a po-
lar configuration of DNA and RNA in RNA tracks, and
specific spatial relationships of different pre-mRNAs to
domains, all suggest structural relationships with a sub-
stantial degree of order.

We appreciate the contributions of clones and antibodies from several sci-
etists which greatly facilitated this work, including X. D. Fu and T. Mani-
atis, L. Leinwand, B. Strauss, D. Rowe, J. Stein, and P. Dobner. We thank
K. Smith for critical comment on the manuscript. We thank H.-L. Haley
and P. Sculthorpe for help with manuscript preparation.

This work was supported by National Institutes of Health GM 49254,
HG00251, and an RCDA to J. B. Lawrence.

Received for publication 14 August 1995 and in revised form 22 Septem-
ber 1995.

Xing et al. Nonrandom Association of Specific Genes to SC-35 Domains

1645


