Dynamic Reconfiguration of Long Human Genes during One Transcription Cycle

Joshua D. Larkin, Peter R. Cook and Argyris Papantonis

10.1128/MCB.00179-12.
Published Ahead of Print 14 May 2012.

Updated information and services can be found at:
http://mcb.asm.org/content/32/14/2738

These include:

REFERENCES
This article cites 50 articles, 17 of which can be accessed free at:
http://mcb.asm.org/content/32/14/2738#ref-list-1

CONTENT ALERTS
Receive: RSS Feeds, eTOCs, free email alerts (when new articles cite this article),
more»

Information about commercial reprint orders: http://journals.asm.org/site/misc/reprints.xhtml
To subscribe to another ASM Journal go to: http://journals.asm.org/site/subscriptions/
Dynamic Reconfiguration of Long Human Genes during One Transcription Cycle

Joshua D. Larkin, Peter R. Cook, and Argyris Papantonis

Sir William Dunn School of Pathology, University of Oxford, Oxford, United Kingdom

We analyzed three human genes that were >200 kbp in length as they are switched on rapidly and synchronously by tumor necrosis factor alpha and obtained new insights into the transcription cycle that are difficult to obtain using continuously active, short, genes. First, a preexisting “whole-gene” loop in one gene disappears on stimulation; it is stabilized by CCCTC-binding factor and TFIIH and poises the gene for a prompt response. Second, “subgene” loops (detected using chromosome conformation capture) develop and enlarge, a result that is simply explained if elongating polymerases become immobilized in transcription factories, where they reel in their templates. Third, high-resolution localization confirms that relevant nascent transcripts (detected using RNA fluorescence in situ hybridization) lie close enough to be present on the surface of one factory. These dynamics underscore the complex transitions between the poised, initiating, and elongating transcriptional states.

MATERIALS AND METHODS

Cell culture. HUVECs from pooled donors (Lonza) were grown to 80 to 90% confluency in endothelial basal medium 2-MV with supplements (EBM; Lonza) and 5% fetal bovine serum (FBS), regrown (“starved”) for 16 to 18 h in EBM plus 0.5% FBS, treated with TNF-α (10 ng/ml; Peprotech), and harvested at different times after stimulation. In some cases, 50 μM 5,6-dichloro-1-b-β-ribofuranosyl-benzimidazole (DRB; Sigma-Aldrich) was added 25 min before harvesting.

Oligonucleotides. PCR primers were designed using Primer3Plus (version 3.0; Whitehead Institute for Biomedical Research [http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi]) to have an optimal length of 20 to 22 nucleotides, to have a melting temperature of 62°C, and to yield amplimers of 125 to 225 bp. For quantitative PCR (qPCR), they were designed after activating qPCR settings. All primer sequences are available on request.

3C. Chromosome conformation capture (3C) was performed as described previously (33). In brief, 10^6 cells were fixed (10 min at 20°C) in 1% paraformaldehyde (Electron Microscopy Sciences), aliquots of 10^6 cells in 0.125 M glycine in phosphate-buffered saline (PBS) were centrifuged, and the cells were resuspended in the appropriate restriction enzyme buffer and lysed (16 h at 37°C) in 0.3% sodium dodecyl sulfate (SDS). After shearing SDS by adding 1.8% Triton X-100 (1.5 h at 37°C), the cells were treated overnight with HindIII or SacI (a total of 800 U added in four sequential steps; New England Biolabs), the enzyme was heat inactivated (25 min at 65°C), and the digestion efficiency was determined by qPCR. Samples with digestion efficiencies of >75% were ligated using T4 ligase (6,000 U [New England Biolabs]), the enzyme was heat inactivated (25 min at 65°C), and the digestion efficiency was determined by qPCR. Samples with digestion efficiencies of >75% were ligated using T4 ligase (6,000 U [New England Biolabs]); DNA concentration, <0.5 ng/μl; 3 to 5 days at 4°C to minimize unwanted ligations), and cross-links were reversed (16 h at 65°C) in proteinase K (10 μg/ml; New England Biolabs) before the DNA was purified using an EZNA MicroElute DNA Kit (OmegaBiotech).
DNA cleanup kit (Omega BioTek) and a PCR purification kit (Qiagen). To control for amplification of randomly ligated segments, nondigested/ligated and digested/nonligated 3C templates were also prepared. For 3C-PCRs, amplification efficiency controls using bacterial artificial chromosomess were as described previously (33). PCR products were conducted using 1.75 mM MgCl₂, 1% dimethyl sulfoxide, 10 pmol of each primer, and GoTaq polymerase (Promega) per reaction (95°C for 2 min, followed by 31 cycles of 95°C for 55 s, 59°C for 45 s, and 72°C for 20 s, followed in turn by 1 cycle at 72°C for 2 min). Amplifiers were resolved in 2.5% agarose gels, stained, and imaged on an FLA-5000 scanner (Fuji). The identities of 3C-PCR products were verified by sequencing (Geneservice, Oxford, United Kingdom) and/or restriction digestions. The results shown were reproduced using at least two independently obtained templates.

**Chromatin immunoprecipitation (ChIP).** Approximately 10⁷ HUVECs were cross-linked (10 min at 20°C) in 0.8% paraformaldehyde at the appropriate times after TNF-α induction. Chromatin was prepared, fragmented, washed, and eluted using a ChIP-It-Express kit according to the instructions for enzymatic shearing (Active Motif). Immunoprecipitations were performed using a rabbit polyclonal against the N terminus of the RNA polymerase II large subunit (sc-889X; Santa Cruz Biotechnology), TFIIH (sc-274X; Santa Cruz Biotechnology), and CTCF (ab70303; Abcam). DNA was purified using a MicroElute Cycle-Pure kit (Omega BioTek) prior to qPCR analysis. The region of GAPDH containing the TATA box was used as a positive control for RNA polymerase II binding, and a portion of the 3⁰ TATA box was used as a positive control for RNA polymerase II binding, respectively. The qPCR analysis was performed using at least two independently obtained templates. For native ChIP (data not shown), the cross-linking step was omitted, and a rabbit polyclonal recognizing histone H3 (sc-10809X; Santa Cruz Biotechnology) was used. The ReChIP-It kit (Active Motif) was used for sequential ChIP (ReChIP).

**Coimmunoprecipitation.** Approximately 5 × 10⁷ HUVECs, grown and induced with TNF-α as described above, were lysed (20 min at 4°C) in 5 ml of lysis buffer (20 mM Tris-HCl [pH 7.5], 150 mM NaCl, 5 mM MgCl₂, 2 mM EDTA, 0.3 mM sucrose, 1% NP-40, and 1 mM dithiothreitol) per ml of lysis buffer (20 mM Tris-HCl [pH 7.5], 150 mM NaCl, 5 mM MgCl₂, 2 mM EDTA, 0.3 mM sucrose, 1% NP-40, and 1 mM dithiothreitol) complemented with protein inhibitor cocktail (PIC; Roche) and spun at 10,000 × g (10 min; 4°C), and the supernatant was precleared (1 h; 4°C) with 5 µg of rabbit IgG (Upstate) in 50 µl of protein A-coupled agarose beads (Pierce). Protein complexes were then pulled down (16 h at 4°C) using anti-CTCF (2 µg) or anti-TFIIH (3 µg) in 100 µl of protein A-coupled agarose beads in 125 or 250 mM NaCl. The complexes were washed 10 times (for 10 min at 4°C each time) in wash buffer (50 mM Tris-HCl [pH 7.5], 0.02% Tween 20, 0.05% Triton X-100, 250 mM NaCl) plus PIC, eluted by boiling in 1× SDS loading buffer (50 µl/ml of lysate), separated in 10% acrylamide gels (Bio-Rad), and transferred onto nitrocellulose using an iBlot transfer system (Invitrogen), and CTCF or TFIIH were detected by immunoblotting (1:5,000 dilution) and visualized by chemiluminescence (SuperSignal West Pico; Pierce) with a ChemiDoc XRS imager (Bio-Rad).

**qPCR.** For 3C and ChIP, quantitative real-time PCR (qPCR) was performed using a Rotor-Gene 3000 cycler (Corbett) and Platinum SYBR green qPCR SuperMix-UDG (Invitrogen). After incubation at 50°C for 2 min to activate the mix and treatment at 95°C for 5 min to denature the templates, reactions were carried out for 40 cycles at 95°C for 15 s and 60°C for 30 s. The presence of single amplifiers was confirmed by melting-curve analysis and gel electrophoresis, and the data were analyzed as described by Nelson et al. (29) for ChIP (to obtain enrichments relative to input) and by Hagège et al. (18) for 3C (to obtain relative interaction frequencies relative to the “loading” and “intra-3C” control).

**RNA FISH.** RNA FISH was performed as described previously (44), using four sets of five 50-mers (Gene Design, Japan). Each set targeted region b, c, d, or e of SAMD4A intron 1. In each 50-mer, roughly every tenth thymine residue was substituted by an amino modifier C6-dT coupled to the Alexa Fluor 488 or Alexa Fluor 555 reactive dyes (Invitrogen), and the five 50-mers in a set targeted <450 contiguous nucleotides. Then, after hybridization of the five 50-mers to one (fully stretched) target RNA, the distance between the outermost fluoros is ~150 nm. Since this distance is below the resolution limit of the microscope, the ~25 fluoros in each probe set appear as one diffraction-limited spot. (Of course, nascent transcripts are unlikely to be fully stretched.) Probes were purified using G-50 columns (GE Healthcare), ethanol precipitated twice, and concentrated using a Microcon-30 column (Millipore), and labeling efficiencies were calculated using the BaseDye ratio calculator (Invitrogen). ~8 fluoros/100 nucleotides. For each experiment, HUVECs on coverslips were grown to ~80% confluence, treated with TNF-α, fixed (17 min at room temperature) in 4% paraformaldehyde–0.05% acetic acid–0.15 M NaCl, washed three times in PBS, permeabilized (5 min at 37°C) in 0.1% PBS (pH 2.0), rinsed in water treated with diethyl-pyrocatecholate, postfixed (5 min at 20°C) in 4% paraformaldehyde-PBS, and stored (overnight at ~20°C) in 70% ethanol. Coverslips were dehydrated in 70, 80, 90, and 100% ethanol and hybridized (overnight at 37°C in a moist chamber) with 25 ng of labeled probes in medium containing 25% deionized formamide, 2× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate), 250 ng of sheared salmon sperm DNA/ml, 5× Denhardt’s solution, 30 mM phosphate buffer (pH 7.0), and 1 mM EDTA. Next, the cells were washed once in 4× SSC (15 min at 37°C) and three times in 2× SSC (10 min at 37°C) and then mounted in Vectashield (Vector Laboratories) supplemented with 1 µg of DAPI (4',6'-diamidino-2-phenylindole; Sigma)/ml.

The analysis depicted in Fig. 5 relies on the targets of the two probes uses lying on different RNA molecules (i.e., in the traveling and standing waves, respectively), even though they both lie within SAMD4A intron 1. (Note that exonic regions can be retained by “exon tethering” [12] as intronic RNA is degraded cotranscriptionally before the polymerase reaches the next exon by any number of mechanisms, including “recursive splicing” [6].) Various types of evidence indicate that the two targets are unlikely to lie on the same molecule. (i) The presence of a “trough” in the tiling microarray data summarized in Fig. 1A indicates that intronic RNA is degraded before the pioneering polymerase reaches exon 2 (if this intronic RNA persisted, there would be no trough). Similar troughs are seen in all five long genes analyzed in reference 44. (ii) The signal intensities in this microarray data indicate there are (at the very least) 3-fold more molecules of region b RNA compared to region c RNA after stimulation (with b:d and b:e ratios being even higher). This is because many polymerases generating the standing wave abort soon after transcribing region b (before they reach c, d, or e). Even if no degradation of intron 1 occurs before the polymerase reaches exon 2 (to generate one RNA molecule containing both targets), this still means that probe b would usually hybridize at the same allele to a different RNA molecule than probe c (or d or e). (iii) The half-lives of regions c, d, and e in intron 1 are between 3 and 6 min (measured using DRB and qRT-PCR [unpublished data]), which is comparable to those seen with intronic RNAs of other mammalian genes (19, 39). Since the pioneering polymerase takes ~10 such half-lives to transcribe intron 1, much of this intron will inevitably be degraded well before the polymerase reaches exon 2.

**Image analysis and high-resolution localization.** Images were collected using an Axioplan 2 inverted microscope (Zeiss) with a Cool-SNAPμs camera (Photometrics) via Meta Morph 7.1 (Molecular Devices). Yellow foci such as those in Fig. 5A were selected for analysis, and the distance between the red and green peaks constituting the focus determined (with 27-nm precision) after localizing each peak (with 16-nm precision). Thus, a 2D Gaussian intensity profile was statistically fit to a peak using regression analysis to minimize least-squares distances between intensity values (23, 43, 49). The pixel shift between fluorescence channels was assessed using 0.1-µm TetraSpeck beads (Molecular Probes) fluorescing at relevant wavelengths. Residual differences in alignment were accounted for, along with spot intensity, shape, and the signal-to-noise ratio to calculate distance uncertainty. All calculations were performed in MATLAB (MathWorks) using custom software routines.
According to the manufacturer’s instructions. The knockdown efficiency was assessed using Western blotting and ChIP; experiments were performed twice.

Statistical analysis. P values (two-tailed) from unpaired Student t tests and Fisher exact tests were calculated using GraphPad; results were considered significant when $P < 0.01$.

RESULTS

Strategy. When TNF-α is added to serum-starved HUVECs, SAMD4A is one of the first genes to respond. It is 221-kbp long and encodes a regulator of the inflammatory response. Transcription begins within 10 min, and then pioneering polymerases transcribe steadily at ~3 kbp/min to reach the terminus after ~85 min. This transcription cycle has been analyzed in detail (33, 44) using, among others, tiling microarrays (where the total RNA isolated every 7.5 min over a period of 3 h was hybridized to the array [44]); the diagram in Fig. 1A summarizes these results. Since exons represent such a small fraction of the gene (e.g., exon 1 is just 193 bp, compared to intron 1 of 134,000 bp), most signal detected by the array results from intronic (nascent) RNA. Within 10 min, additional polymerases continuously reinitiate but terminate prematurely (circular dotted arrow) to create a “standing” wave (green). Thus, the gene is transcribed by two polymerases that generate “traveling” and “standing” waves. (B) Working 3D model. Within 10 min of adding TNF-α, segment b has collided with a factory (sphere), and initiated; the polymerase in the factory is now elongating productively by reeling in b as it extrudes a transcript (red). By 30 min, this pioneering polymerase is transcribing c, and the promoter is now tethered close to the factory and thus likely to collide with another polymerase in the factory (arrow). By 31 min, the pioneer continues to reel in its template as it elongates (to generate the “traveling” wave), and the promoter has initiated again. However, this second polymerase will soon abort. Successive cycles of promoter attachment and detachment—and initiation and abortion (circular dotted arrow)—now follow (to generate the “standing” wave). While b and c are attached, a small subgene loop exists. By 60 min, the pioneer is transcribing e, as another polymerase transcribes b; the enlarged subgene loop now stretches from b to e.

soon abort within 10 kbp of the transcription start site (TSS). Successive cycles of initiation and abortion then continue to generate (intronic) transcripts within the first 10 kbp; these are depicted as the green “standing” wave. After ~15 min, a “trough” develops between the two waves; this can only result if the intronic RNA is removed and degraded cotranscriptionally and rapidly and if the following polymerases soon abort (otherwise signal would fill the trough). The presence of polymerases and nascent (intronic) RNA at the points indicated in the diagram has been confirmed using ChIP both conventionally (see below) and coupled to high-throughput sequencing, “ChIP-on-chip,” RT-PCR, and RNA FISH (33, 44). Similar patterns are also seen on two other long, responsive genes: 312-kbp ZFPM2 and 458-kbp ZFPM2 during a transcription cycle. HUVECs were treated with TNF-α for 0 to 85 min, and ChIP followed by qPCR were applied. In some cases, the transcriptional inhibitor, DRB (50 μM), was added 25 min before harvesting. Diagrams on the left illustrate regions targeted by primer pairs used and their distance (in kbp) from the TSS. Bound RNA polymerase was detected using an antibody recognizing the largest subunit of the enzyme. The results (percent enrichments relative to the input ± the standard deviation [SD]; $n = 6$) are superimposed on data obtained previously using microarrays (44). Peak enrichments at different times coincide with the traveling (yellow) and standing waves (green). * DRB significantly reduces enrichments seen ($P < 0.01$; unpaired two-tailed Student t test).

The model in Fig. 1B illustrates how the conformation of SAMD4A might change during a transcription cycle, assuming that elongating polymerases are immoveable molecular machines.
housed in factories (33). By 31 min, two polymerases may be transcribing the long gene—the pioneer and a follower—which tether a subgene loop to the factory. After 60 min, this subgene loop has enlarged.

**A subgene loop grows between transcribed regions.** We first validated that polymerases were bound to the appropriate segments of all three long genes at the relevant times using ChiP coupled to qPCR (ChiP-qPCR) and antibodies recognizing the N terminus of the major subunit of RNA polymerase II (RNAPII). As expected, binding reflects the presence of the standing and traveling waves, and treatment with a transcriptional inhibitor, DRB, abolishes the traveling wave (Fig. 2). When we used antibodies recognizing different phosphorylated isoforms of the polymerase (9, 44), similar data were obtained (not shown).

We next applied 3C to determine whether a subgene loop appears in SAMD4A and then enlarges using the TSS (segment b) as a reference point (Fig. 3A). At no time does the TSS contact points 10 kbp upstream or downstream of the gene (i.e., no band is seen when primer b is paired with primer a or i in the 3C-PCR).

Before stimulation, the TSS is hardly transcribed and contacts no part of SAMD4A, except for region h at the terminus. This indicates that the TSS contacts the terminus to generate a “whole-gene” loop. (This loop will be analyzed in detail later.) After 10 min, the sole band is an emerging one seen with region c. Since region c is not maximally transcribed until 30 min after stimulation, we attribute this to a subset of polymerases initiating well before the majority in the population (since synchrony is not perfect). After 30 min, the contact with region c becomes prominent, and another with region d develops. After 60 min, contacts spread further 3'; after 85 min, they reach the end of the gene (so the subgene loop has grown to encompass the whole gene). A similar wave of evolving contacts is seen when HindIII replaces SacI during preparation of the 3C template, using 3C coupled to either conventional PCR (Fig. 3B) or qPCR (Fig. 3C). These interactions depend on transcription, since they are abolished by DRB (Fig. 3, gray panels).

The two other long, responsive genes yield similar 3C contacts that spread down the gene with time and which are indicative of an enlarging subgene loop (Fig. 4). However, in both cases no whole-gene loop equivalent to that in SAMD4A was seen at 0 min (Fig. 4). All of these results support the model in Fig. 1B.

**Localization of single nascent RNA transcripts in the standing and traveling waves.** 3C shows that the two transcribed templates of one long gene are together, so we used an independent method—RNA FISH—to confirm that the nascent transcripts encoded by the contacting sequences also lie together. Here we use probe pairs (labeled with red or green fluors) targeting a short intronic RNA segment within each wave (Fig. 5A). As we have seen, the two targets lie on different RNA molecules generated by different polymerases. (See Materials and Methods for additional evidence that the two targets are unlikely to lie on one RNA molecule.) Previous work (33, 44) has shown that (i) the oligonucleotide probes used can detect a single intronic RNA efficiently to yield a diffraction-limited spot, (ii) HUVECs contain only two SAMD4A alleles (they are diploid and synchronized in the G0 phase of the cell cycle), (iii) ~30% alleles in the population are being transcribed by a pioneering polymerase at any one time after stimulation, and (iv) a yellow spot indicative of colocalizing targets can only result from RNA copied from the same allele (since the spot area is so small compared to nuclear area that a green
focus will only overlap by chance a red focus copied from a different allele in \(<1\) nucleus in a thousand, assuming random distributions).

Targets for probe pairs b plus c and b plus d (at 30 and 45 min after stimulation, respectively) are copied from two DNA regions lying \(~33\) and \(~44\) kbp apart and yield yellow foci in \(20\%\) of cells (Fig. 5A). Since electron microscopy shows that nascent transcripts typically lie on the surface of a factory with an \(~90\)-nm protein-rich core (13,14), we wanted to see whether these colocalizing transcripts lay this close together. Unfortunately, the conventional fluorescence microscope has a resolution of \(~250\) nm, at best. Therefore, we used an approach (33) that allows resolution beyond the diffraction limit; we assume the red and green signals that yield a yellow focus mark two subdiffraction spots (one red, the other green), fit Gaussian curves to their intensities, and measure the 2D distance between peaks. There were few interpeak distances of \(\leq25\) nm (the mean values were 64 and 60 nm for b plus c and b plus d, respectively), and the maximum was \(~160\) nm (Fig. 5B). In other words, short interpeak distances are rare. Such a distribution is not that expected of red and green foci randomly distributed in a small (spherical) volume of analogous dimensions, where most interpeak distances lie close to zero (since so many red foci lie immediately above or below a green focus to give a short 2D distance). We then compared these experimental results to those obtained from a computer simulation of transcripts on the surface of a factory: red and green points were randomly distributed in a (variably sized) shell around a (variably sized) sphere, and then the 2D distance between points was measured. The best fit to the experimental data was given by points randomly distributed in a 35-nm shell around a 90-nm core (Fig. 5B, orange line), which is consistent with the known dimensions and location of transcripts determined by electron microscopy (13,14). (Similar results are obtained with transcripts encoded by two co-associating genes on different chromosomes [33].) As a control, we imaged multicolor fluorescent beads that should colocalize “perfectly.” As expected, measured red-green distances were \(\leq25\) nm (Fig. 5B, gray panel), the precision level of the method. As another control, overlapping red and green signals emitted by a mixture of red and green probes targeting just region c also had a significantly smaller mean separation than those given by b plus c or b plus d (data not shown).

We now argue that the separations seen above are inconsistent with a model involving the generation of two transcripts by two tracking polymerases. If such a model applied, and as the two
adopted a self-avoiding walk. (i.e., fixed on the surface of one factory), with the intervening DNA forming an extending subgene loop. Note also that these results are also inconsistent with targets being encoded by a template that (i) follows a self-avoiding random walk (Fig. 5C, blue curve) or (ii) is tightly packed into a sphere between targets (if packed at the highest concentration ever seen in vivo [46], targets would still lie twice as far apart as observed [data not shown]).

Polymerases are poised on the promoters of three long genes before stimulation. It is attractive to suppose that polymerases might be “poised” on the promoter of these long genes ready to facilitate a rapid response to stimulation (see also reference 15). ChIP with an antibody targeting the N terminus of the largest subunit of the enzyme shows it is indeed bound to the 5' end of each gene before induction (Fig. 6) and that binding increases in synchrony with p65 binding to its cognate elements (data not shown). Since polymerases bound at “poised” promoters fire continually and unproductively to generate short sense and antisense transcripts (11, 36, 38), we used qRT-PCR to determine whether this was true here. Before stimulation, sense and antisense transcripts copied from one allele. Bar, 2 μm. (B) Single-molecule localizations. Gaussian curves were fitted to red and green foci, giving yellow foci such as those in panel B, and separations between peaks were determined with 27-nm accuracy. Few separations were <25 nm, and their distribution is consistent with foci being randomly distributed in a 35-nm shell around one 90-nm factory (diagram; orange curve from reference 33). A control for “perfect” colocalization is provided by 110-nm red/green fluorescent beads, n, the number of yellow foci analyzed. (C) Relative separations (± the SD; green line) between red and green peaks in yellow foci given by the probe pairs b plus c, b plus d, and b plus e (normalized to the mean value obtained with b plus c). The dotted black line shows what is expected if separations increased in proportion to the number of base pairs between b, c, d, and e. (The contour lengths of B-DNA between targets b plus c, b plus d, and b plus e are 11,120, 14,620, and 43,180 nm, respectively.) The blue line shows what is expected if the template adopted a self-avoiding walk.

nascent RNAs lie near their immediate templates, we would expect the separations seen with probe pairs b plus c, b plus d, and b plus e to increase in proportion to the number of base pairs between the DNA encoding the targets (Fig. 5C, black line). However, the observed separations (normalized relative to that seen with b plus c) remain essentially constant (Fig. 5C, green line). This result is consistent with the two targets being generated by polymerases lying a constant distance apart (i.e., fixed on the surface of one factory), with the intervening DNA forming an extending subgene loop. Note also that these results are also inconsistent with targets being encoded by a template that (i) follows a self-avoiding random walk (Fig. 5C, blue curve) or (ii) is tightly packed into a sphere between targets (if packed at the highest concentration ever seen in vivo [46], targets would still lie twice as far apart as observed [data not shown]).

The SAMD4A whole-gene loop seen before stimulation is CTCF/TFIIB-dependent. Loops formed by juxtaposition of the 5' and 3' ends of various genes have been observed (see, for example, references 30, 31, 40, and 42), and such a configuration is seen in SAMD4A before stimulation with TNF-α (but not in EXT1 or ZFPM2) (Fig. 3 and 4). This whole-gene loop is lost on stimulation (shown using 3C applied both conventionally and coupled to qPCR on templates generated with two different restriction enzymes in Fig. 3). One difference between SAMD4A and EXT1 or ZFPM2 lies in the nucleosome occupancy of their 5' proximal regions; SAMD4A possesses such a nucleosome-free region both before and after stimulation, but one only appears in the two other long genes after TNF-α treatment (Fig. 6).

CCCTC-binding factor (CTCF) has been implicated in stabilizing chromatin loops (see, for example, references 26 and 28), so we examined whether it was involved. An in silico search (using the algorithm available from the School of Biological Sciences, University of Essex, Essex, United Kingdom [http://www.essex.ac.uk/bs/molonc/binfo/ctctbind.htm]) uncovered potential CTCF-binding sites at both ends of SAMD4A, but only at the 5' end of EXT1 and ZFPM2. Binding was confirmed by ChIP-qPCR (Fig. 6) (see also reference 44). Therefore, binding to both SAMD4A ends correlates with the formation of the whole-gene loop. (Note that CTCF monomers—in contrast to dimers—cannot mediate looping in vitro [25].) However, CTCF remains bound to these cognate sites throughout the 85 min we follow (Fig. 6), whereas the whole-gene loop disappears on stimulation (Fig. 3).
The general transcription factor, TFIIB, has also been implicated in stabilizing whole-gene loops in yeast (40). Like CTCF, it is bound before stimulation to both ends of \textit{SAMD4A}, with significant enrichment at the 3' end; again, this is not true of \textit{EXT1} or \textit{ZFPM2} (Fig. 6). Binding to the 3' end of \textit{SAMD4A} falls significantly upon stimulation (Fig. 6), correlating with the disappearance of the whole-gene loop. This suggests that CTCF and TFIIB might interact \textit{in vivo}; both sequential ChIP (ReChIP) and coimmunoprecipitation experiments support this idea (Fig. 7B and C).

Knocking down CTCF using siRNAs (Fig. 8A) reduces binding to both ends of \textit{SAMD4A} (as expected) and also reduces TFIIB binding (Fig. 8B). The CTCF knockdown also eliminates the 3C contact indicative of the whole-gene loop (Fig. 8C, compare m to KD). The presence of this loop depends on continued transcription (Fig. 8C, compare m to +DRB) and does not result from serum starvation (Fig. 8C, compare NS to m). Taken together, these results indicate that the polymerase, CTCF, and TFIIB, jointly stabilize the exceptional whole-gene loop in \textit{SAMD4A}.

We also monitored the effects of knocking down CTCF on transcription 30 min after stimulation (using RNA FISH with probes targeting region c of \textit{SAMD4A} and the equivalent region of \textit{EXT1}). In mock-treated cells, 26% and 21% of the nuclei possessed at least one focus indicative of an active \textit{SAMD4A} or \textit{EXT1} allele, respectively (Fig. 8D). After knockdown, substantially fewer nuclei (i.e., 15%) contained active \textit{SAMD4A} alleles, while \textit{EXT1} activity remained unaffected (at 22%) (Fig. 8D). Finally, we checked (by qRT-PCR) whether knocking down CTCF reduced the sense and antisense transcripts at the \textit{SAMD4A} promoter. Previously, we saw there was a strong bias toward sense transcription at or near the TSS. Knockdown reduced this strong bias (e.g., before

FIG 6 Changes induced by TNF-\textalpha on the binding of RNA polymerase II, CTCF, TFIIB, and nucleosomes to each end of \textit{SAMD4A}, \textit{EXT1}, and \textit{ZFPM2}. HUVECs were treated with TNF-\textalpha for the times indicated, and ChIP was performed with antibodies against the largest subunit of RNA polymerase II, CTCF, TFIIB, and histone H3 (percent enrichments + the SD; n = 6). The diagrams on the left illustrate targets of the primer pairs in promoters (S1, E1, and Z1), the \textit{SAMD4A} gene body (S2), and 3' untranslated regions (UTRs; S3, S4, E2, and Z2). The polymerase is found at all promoters, both before and after stimulation. It is also seen at the 3' UTR of \textit{SAMD4A} after 85 min (as expected), but not at the 3' UTR of \textit{EXT1} or \textit{ZFPM2}, since pioneering polymerases only reach the end of these two long genes after the end of the experiment. CTCF binds to promoters throughout the time course with comparable enrichments. It also binds to S3 and S4 but not to S2, E2 or Z2. TFIIB binding to all promoters peaks after 10 to 30 min. Some TFIIB binds to the 3' UTR of \textit{SAMD4A} but not to the middle of \textit{SAMD4A} (as expected) or the 3' UTRs of \textit{EXT1} or \textit{ZFPM2}. When 50 \textmu M DRB is added to cells 25 min before harvesting (gray panel), less TFIIB binds to the \textit{SAMD4A} 3' UTR. H3 levels are generally low, except for E1 and Z1 at 0 min, which changes on stimulation (perhaps as the nucleosomes reposition). *, difference significant (P < 0.01, two-tailed unpaired Student t test).
FIG 7 CTCF promotes sense transcription at the SAMD4A promoter and interacts with TFIIB. HUVECs were treated with TNF-α for 0 to 60 min. (A) Levels of sense and antisense transcripts at the SAMD4A promoter determined by qRT-PCR. HUVECs were treated with siRNAs that knock down CTCF and TFIIB (KD) or mock treated (mock) prior to stimulation with TNF-α (44). Within 10 min, pioneering polymerases initiate and then transcribe steadily to terminate $>1$ h later. Meanwhile, other polymerases initiate, but these soon abort. As a result, each of these long genes becomes transcribed by two polymerases, a pioneer and a follower, and these generate traveling and standing waves of nascent RNA (Fig. 1A). This system allows us to obtain new insights into the transcription cycle that are difficult, if not impossible, to obtain using continuously active, short genes.

We first examined the formation of subgene loops. If both elongating and initiating polymerases are transiently immobilized in one transcription factory (8, 32, 33), stimulation should induce a subgene loop to form and then enlarge (Fig. 1B). 3C analyses show this to be so (Fig. 3 and 4), and RNA FISH allied to subdiffraction localizations confirms that nascent RNAs copied from the contacting sequences lie sufficiently close together to be on the surface of one 90-nm factory (Fig. 5). Equivalent subgene loops are likely to be found in many human genes simply because $\sim 42\%$ are longer than 30 kbp (4), the length of the smallest subgene loop we see. Since all parts of the transcription unit contact the 5′ end in the course of one transcription cycle, such subgene loops are likely to remain undetected by current approaches applied to unsynchronized genes. This model shares with others involving “transcriptional compartments” (48) or “active chromatin hubs” (3, 34) the idea that elongating polymerases interact (directly or indirectly) with the promoters where they initiated (see also reference 22).

We also investigated the whole-gene loop seen in SAMD4A before stimulation. CTCF, TFIIB, and ongoing transcription all seem to stabilize this loop. Knocking down CTCF reduces binding of both to both ends (Fig. 8B), eliminates the loop (Fig. 8C), and reduces productive elongation (Fig. 8D). It also reduces "noisy" transcription upstream of the promoter and relieves the bias toward sense promoter transcripts (Fig. 7A), a switch anticipated to remain undetected by current approaches applied to unsynchronized genes. This model shares with others involving “transcriptional compartments” (48) or “active chromatin hubs” (3, 34) the idea that elongating polymerases interact (directly or indirectly) with the promoters where they initiated (see also reference 22).

DISCUSSION

We analyzed the changing conformations of three long genes (221-kbp SAMD4A, 312-kbp EXT1, and 458-kbp ZFPM2) which respond to TNF-α (44). Within 10 min, pioneering polymerases initiate and then transcribe steadily to terminate $>1$ h later. Meanwhile, other polymerases initiate, but these soon abort. As a result, each of these long genes becomes transcribed by two polymerases, a pioneer and a follower, and these generate traveling and standing waves of nascent RNA (Fig. 1A). This system allows us to obtain new insights into the transcription cycle that are difficult, if not impossible, to obtain using continuously active, short genes.

We first examined the formation of subgene loops. If both elongating and initiating polymerases are transiently immobilized in one transcription factory (8, 32, 33), stimulation should induce a subgene loop to form and then enlarge (Fig. 1B). 3C analyses show this to be so (Fig. 3 and 4), and RNA FISH allied to subdiffraction localizations confirms that nascent RNAs copied from the contacting sequences lie sufficiently close together to be on the surface of one 90-nm factory (Fig. 5). Equivalent subgene loops are likely to be found in many human genes simply because $\sim 42\%$ are longer than 30 kbp (4), the length of the smallest subgene loop we see. Since all parts of the transcription unit contact the 5′ end in the course of one transcription cycle, such subgene loops are likely to remain undetected by current approaches applied to unsynchronized genes. This model shares with others involving “transcriptional compartments” (48) or “active chromatin hubs” (3, 34) the idea that elongating polymerases interact (directly or indirectly) with the promoters where they initiated (see also reference 22).

We also investigated the whole-gene loop seen in SAMD4A before stimulation. CTCF, TFIIB, and ongoing transcription all seem to stabilize this loop. Knocking down CTCF reduces binding of both to both ends (Fig. 8B), eliminates the loop (Fig. 8C), and reduces productive elongation (Fig. 8D). It also reduces "noisy" transcription upstream of the promoter and relieves the bias toward sense promoter transcripts (Fig. 7A), a switch anticipated to reduce productive elongation (16). Since DRB treatment also eliminates the whole-gene loop (Fig. 8B), its existence depends on continuing transcription (either in the sense or antisense direction). However, here the polymerase seems to be bound only to
the 5′ contacting partner and not to the 3′ one. It is then attractive to suppose that this whole-gene loop "poises" the gene so it can respond rapidly to TNF-α. An analogous role in increasing the efficiency of mRNA production has been suggested for such loops in yeast, the HIV-1 provirus, and other mammalian genes, either directly through close association with transcription/processing factors or indirectly through contacts with the nuclear pore and the mRNA export machinery (1, 20, 31, 35, 40–42, 50). Again, the use of a switchable and synchronized system allows us to monitor the appearance and (critically) the disappearance of this whole-gene loop.

In conclusion, all of these results are consistent with a model for transcription in which active polymerases are immobile molecular machines housed in factories, where components of the transcription machinery act as the critical molecular ties that loop the chromatin fiber (10, 20). In the case of the subgene loops in SAMD4A, EXT1, and ZFP2M, polymerases are probably the sole ties; in the whole-gene loop in SAMD4A, the polymerase (and perhaps CCTF and TFIIB) probably constitutes the tie at the promoter, while CTCF and TFIIB act at the terminus. Our data also highlight the cross talk between the poised, initiating (whether productively or unproductively), and elongating transcriptional machinery and the immediate and changing effects this has on the 3D structure of chromatin.

ACKNOWLEDGMENTS

We thank Tatsuhiko Kodama for RNA FISH probes, Davide Marenduzzo for discussions on the self-avoiding fiber calculations, and Shona Murphy and Clelia Laitem for siRNA reagents and advice.

This study was supported by the Biotechnology and Biological Sciences Research Council via the ERA-SysBio+ FfP7 initiative (A.P.) and Wellcome Trust (A.P. and J.D.L.). A.P. is the Kemp Junior Research Fellow of Lincoln College; P.R.C. holds the EP Abraham Chair of Cell Biology and a Professorial Fellowship at Lincoln College.

REFERENCES


